I Don't Need an Expert! Making URL Phishing Features Human Comprehensible

Kholoud Althobaiti
University of Edinburgh
Taif University
Edinburgh, UK -Taif, KSA
kholod.k@tu.edu.sa

Nicole Meng
University of Edinburgh
Edinburgh, UK
nicole.meng@ed.ac.uk

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

ABSTRACT

Judging the safety of a URL is something that even security experts struggle to do accurately without additional information. In this work, we aim to make experts' tools accessible to non-experts and assist general users in judging the safety of URLs by providing them with a usable report based on the information professionals use. We designed the report by iterating with 8 focus groups made up of end users, HCI experts, and security experts to ensure that the report was usable as well as accurately interpreted the information. We also conducted an online evaluation with 153 participants to compare different report-length options. We find that the longer comprehensive report allows users to accurately judge URL safety (93% accurate) and that summaries still provide benefit (83% accurate) compared to domain highlighting (65% accurate).

CCS CONCEPTS

• **Security and privacy → Phishing**.

KEYWORDS

Phishing; URL reading; phishing awareness; usable privacy and security; real-time learning; security education, decision support

1 INTRODUCTION

Determining if a link (URL) in a communication is malicious phishing designed to trick users or not is something that even security experts struggle to do without the aid of tools and additional information. While looking at the URL text is a good first step, fully reading a URL and determining its actual destination is surprisingly complex and often requires the help of third-party services that provide information like how long ago the domain was registered or if the URL redirects anywhere. Yet, despite these complexities,

URLs remain one of the stronger indicators of malicious communication [52, 62], particularly if a communication claims to come from an organization but the URLs lead to other destinations. For example, an email claiming to be from PayPal but containing links to `PayPal-com-security-website.org` is quite likely a malicious email. While the example is simple, it brings up several key issues with detecting malicious links. First off, it requires that the person making the judgment knows PayPal's correct URL and is also able to compare it to the one in the email. The "correct" URL for a website is not necessarily obvious; for example, which of the following is the correct website for the New York Times newspaper: `nyt.com` or `newyorktimes.com`? The answer is that both URLs redirect to the real URL `www.newyorktimes.com`. Comparing URLs is also not necessarily easy. End users often confuse elements of a URL, such as the domain and subdomain [2] making comparing URLs error-prone. Experts handle these complexities using a range of tools and information sources that help them make decisions, but end users are often only provided with training on lexical reading [43, 74, 82] and possibly a tool checks if the URL has been confirmed as malicious. In this work, we aim to change this situation by making the types of approaches used by experts more accessible to end users.

Obviously users are not the best first option to detect phishing. Automated phishing filters are far less expensive and also relatively accurate [40]. They can quickly compare a URL to lists of known phishing or break the URL into features used to classify it as phishing or not. Most organizations already use automated approaches to protect users on their networks with great success. However, this usage means that any phishing communications a user sees has likely already been through an automated filter and therefore has already been scanned against common computer-friendly features. Assisting users in making these judgments on their own is necessary because automated approaches are not yet 100% accurate [61] and experts are also typically not available to consult on every potential phishing communication in a timely manner. Phishing communication also often uses tactics to pressure the user into responding quickly such as threatening to shut down their account, charge them money, upset their boss, or lose out on a limited time offer [55, 86]. These time pressures make it emotionally hard for users to report the email and then wait for an official response, leading them to use their own or their peers' judgment [56]. The effects are readily apparent in public phishing reports. Phishing is regularly listed as one of the top causes of data breaches (93%) [80] and the most frequent Internet crimes complaint to the FBI [21]. Financial losses from phishing can also be expensive, exceeding $29 million in 2017 [32] and $1.7 billion in 2019 [21]. Tools supporting users in accurately making such decisions on their own are

therefore needed. So when providing users with support we need to think beyond simply telling them if something is phishing or not and instead focus on helping them leverage their contextual knowledge of the situation in conjunction with available data to reach a decision.

When judging the safety of a URL, experts generally have more experience and data sources to draw from but at the end they look for discrepancies in the data and their expectations [84]. They can collect the data using tools like WHOIS (ICANN's domain lookup) to learn about the registered domain owner or understand the implications of a links up-time and popularity. However, using such tools requires an impressive amount of both access to information and knowledge about how to interpret them. Each URL case also requires a slightly different set of information sources and knowledge, making training users to make such judgments on their own overly burdensome.

Our goal is to support end users, so that they can engage in some of the informed reasoning experts currently use when they want to decide on a URL's safety. More specifically, we want to take existing information sources along with knowledge about how to interpret those sources and use it to help end user decision making. To do so, we started with a grid-based report structure inspired by the Privacy Nutrition Labels work by Kelly et al. [38]. The grid presents the user with information about the URL, drawn from existing research on the URL features that are likely to be the most useful to humans [5], and is annotated with explanations aimed at helping users interpret the information. We then iterated on its design with the assistance of 8 focus groups consisting of end users, security experts, and design experts to simplify the interface and improve the explanation of features. After we created a stable design, we analyzed what it would look like on 4640 URLs from two phishing datasets and two safe URL datasets. The goal was to determine if there was any redundant information on the report and also if the features we chose do generally align with the safety state of the URLs. Finally, we ran a user study with 153 Prolific users to determine if they could correctly understand the report contents and make accurate safety judgments.

We found that focus group participants saw how such a report could be useful in cases where they were unsure about a communication. The later focus groups also found the report design useful and informative. Final versions of the report featured only showing relevant information, and colors to help users know where to focus. In our analysis of how the report would look with real URLs we found that for most URLs the report only needed to show about 7 of the 23 possible information rows, greatly limiting the user's reading burden. The colors also tended to align with the URL being phishing or not. The online participants and focus group participants exhibited similar interpretations of the various report elements, suggesting that our richer focus group data was a good representation of what Prolific users also thought.

2 DECIDING IF A URL GOES WHERE THE USER THINKS IT GOES

Phishing communication often works by convincing the user that the message they received is from a legitimate group they want to interact with. Examples include: their bank, their IT department,

their email provider, their gaming account, or a lottery site they just won money on. Attackers are also often interested in account credentials, so it is useful for them to mimic existing services to trick users into entering their credentials or providing other sensitive data that the user might normally only give to a trusted party. One side effect of this approach is that the user has strong expectations of where they think they are going when they click on any links. The other side effect is that, except in very rare situations, the attacker does not have access to the company's real URL and instead must setup a fake one, so the URL the user clicks on is owned by someone other than the group the user thinks they are interacting with.

In a review of URL phishing features used by humans and by automated systems, Althobaiti et al. [5] observed that the domain part of the URL is the most used feature in human-based detection because they can compare it against their expectations. It is less useful for computers because the computer has to guess if the URL matches the content of the communication. The problem is that while theoretically combining the domain with contextual knowledge should make phishing easy to detect by humans, in practice, people struggle to accurately parse URLs [2], making comparison extra challenging.

2.1 Mouse over the link and look at the URL

One common piece of advice users are given is to mouse over links in communications and look at where they go [33, 54, 68]. This is good advice, especially in cases where the URL is very different from expectations such as an email, supposedly from PayPal, containing a `moonstone235432.net` link. But the advice gets harder to follow if the attacker uses any of a wide range of tricks [22, 26, 49, 89].
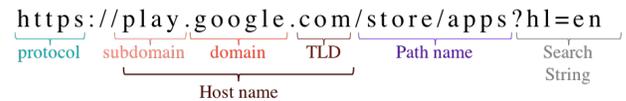

Figure 1: Example URL along with its structure.

As Figure 1 shows, a URL is made up of many elements which impact its destination and can be easy for end users to confuse. For phishing, the most important element to look at is the host name, particularly the domain [12, 20, 39, 43, 45, 74, 90]. This part of the URL controls what server will be contacted to fetch the page, essentially, who controls the page. In order to divert the user to a page they control, the attacker must specify their own domain and use tricks to make it look legitimate. We detail some of the tricks here and refer the reader for a more comprehensive overview [6, 22, 26, 64].

The simplest and oldest approach is using a complicated looking domain name like the raw IP address, hex or decimals characters instead of the real one [26, 47]. A slightly more advanced approach is to pick a domain that looks visibly similar to the real one, but is actually different [26, 49, 78, 89]. Even skilled security experts have difficulties with this kind of deception [17, 24]. For example, in so-called homograph attacks English characters are substituted with identical looking UTF8-encoded characters from different alphabets such as `páypal.com` and `paypal.com` [22, 25, 69]. Another example of

a look alike attack is misspelling (typosquatting). A classic example is substituting characters like 'vv' for 'w' or capital 'I' for lowercase 'L' which look identical with a sans-serif font [69, 76]. These two types of look alike attacks, while dangerous, are very popular and hard to detect by current industry anti-phishing tools [65, 77].

Another trick is to leverage users' inability to differentiate between URL components [22]. For example, Albakry et al. [2] found that users cannot differentiate between a company name in the subdomain vs. the domain of a URL. Similarly, Reynolds et al. [69] found that users struggle to correctly parse URLs, but have high self-confidence in their ability to interpret URLs. A dangerous combination that helps attackers. A common trick involves putting a brand name into an incorrect position, such as in the subdomain (e.g. `amazon.evil.com`), path (e.g. `evil.com/amazon`), search string (e.g. `evil.com?amazon`), or even username (e.g. `amazon@evil.com`). A similar trick is to swap out the top-level domain (TLD) such as `amazon.evil` instead of `amazon.com` [71] or put a fake TLD into a subdomain (`amazon.com.evil.com`).

2.2 What is the "correct" URL anyway?

The ability to compare a URL in a communication with the "correct" domain for that organization is a major factor in determining if a URL is legitimate or not. Unfortunately, it is surprisingly hard to determine which domains are associated with a given organization.

Large companies will typically use a domain name that matches their brand name; e.g, CNN uses `cnn.com` and Chase Bank uses `chase.com`. But some organizations select domains that are not necessarily obvious; for example, Fifth Third Bank has a domain of `53.com` which relates to its brand, but is not immediately obviously the correct URL. Often, companies also have multiple domains associated with their brand such as Mirosoft which owns `microsoft.com`, `live.com`, and `xbox.com` some of which are not obviously Microsoft domains from their text. There are several good reasons why a company might have multiple domains such as having multiple product lines, or registering "defensive domains" to protect their customers from both typing errors and attackers.

Organizations can also have similar names to other organizations, which makes it challenging for users to know what domain name to compare against. For example, many banks share the name "Citizen's Bank" resulting in a confusing array of both bank names and URLs. Citizen's Bank (`citizens-bank.org`) and Citizen's Bank (`citizenbank.bank`) are two different banks which are not to be confused with Charter One Bank (`citizensbankonline.com`) which is owned by another Citizen's Bank (`citizensbank.com`). The point here is that it is not trivial to just look at a domain name and associate it with a particular organization.

Finally, websites that host others' content can make the situation even more confusing. For example, `windows.net` is owned by Microsoft but is a content hosting site. That is, people pay Microsoft for web space and then can create websites like `evil.windows.net` which are then hosted from a Microsoft-owned domain [15]. The result is a phishing site that is linked to a real Microsoft domain but where the content of the page is actually controlled by attackers.

2.3 Redirects and Short URLs

While the domain shown in a clicked URL is often the same as the final destination URL, that is not always true. Organizations commonly do minor redirects such as adding 'www'. Some may also redirect to their preferred brand such as `nyt.com` redirecting to `www.nytimes.com`. More challenging are URLs that obscure the real URL completely making the URL's destination impossible to predict without assistance [13, 28]. Examples include URL shortening services (e.g. `bit.ly`) [8], QR codes, and URL-rewriting by email servers (e.g. `safelinks.protection.outlook.com`). Thankfully, users do seem to be aware that they cannot predict the destination of shortened URLs [2].

3 RELATED WORK

Research on preventing phishing attacks has adopted three complementary approaches: automating phishing detection, educating users about phishing, and supporting users' decisions with security indicators. A full review of automated phishing detection is outside the scope of this paper, though we review some of the features in Section 5. We present a brief overview of the other areas.

3.1 User Training

Humans are the last point of defense for organizations as detecting phishing emails requires humans awareness of the context in which they received the phishing message, such as who they expect to receive the message from and which website they expect to visit [84]. Experts, for example, identify phishing emails by hovering over links, looking at sender emails address, and other technical information of the email; they typically learn to look at these features from training materials [84]. Training average users to identify phishing messages is a common approach which is often combined with automatic detection [9, 12, 43, 74]. On average, organizations spend about $290k every year on training [73], which is often either done upfront [12, 74] or embedded in daily work [45].

Upfront training explains concepts in a dedicated training session. Its effectiveness depends on the user's ability to recall and apply these learned lessons in later situations, which can be challenging since people tend to forget unused information [18, 64], necessitating periodical training [34]. Even if they can remember, attackers also continuously adjust their tactics over time, invalidating some of the learned information [5, 22].

Embedded training is designed to be integrated into users' daily routines, such as receiving training if they fall for phishing attacks. While effective [75], embedded training is costly with the need for a human administrator's time to craft simulated phishing communications [63] that must also be realistic and up to date [31, 40].

Even if training could improve users' ability to detect phishing websites, URLs can be manipulated in ways that are not easily discernible by the human eye. No amount of training will solve these issues because of humans' physical limitations [12, 17]. Consequently, while phishing education does offer skills people need to improve their phishing detection abilities, they are unlikely to be able to discern challenging URLs without the assistance of tools which are not currently easy to find or use [5].

3.2 Phishing detection support

In phishing detection support, a computer assists the user by providing extra information or comparing the URL to known labeled ones. These support systems can take several forms, e.g., browser warnings, chatbots, and toolbars. This collaborative approach is suggested to be complimentary by Park et al. [61] who argue that through utilizing the complementary strengths of a human and an agent, we can achieve the results we desire.

Several existing tools provide phishing-detection support for users. Netcraft's browser plugin [48] warns users about blacklisted webpages once they visited them; as well as clearly displaying the website's country, site rank, hostname, and other facts to help users identify fraudulent URLs. SpoofStick presents the domain name in the browser toolbar to highlight cases where there is a legitimate-looking domain name in a wrong position [88]. Yang et al. also designed security warnings based on website traffic ranks [91]. The Faheem chat bot [6] provides basic facts about any given URL; including the existence of misspellings, non-ASCII characters and redirection. Users can also ask the bot to elaborate on any term and receive a longer explanation. TORPEDO [82], a Thunderbird add-on, presents and highlights the domain of a URL linked in hypertext on a email. The add-on will disable the links for 3 seconds so users stop and think about the URL safety.

The above security indicators take a similar approach to our proposed report. We are presenting the user with information about the URL prior to visiting it under the assumption that with support, they will have the ability to identify unexpected aspects of the link. Our work differs from existing solutions in that it focuses on how to express potentially complex URL and web hosting concepts to users in an easy-to-comprehend way. Existing solutions either focus on providing support to more technical users who may already have a strong lexicon of internet terms like "host", "domain" and "hosting provider" or providing basic support that does not add much to the upfront user training. Our work is aimed at bringing this type of information to a broader audience.

4 DESIGN GOALS

From the above related work we can see that there are three large problems that need to be solved: 1) human judgment is needed to determine if a URL is safe because the human has contextual knowledge that is not available to the computer, 2) URLs are made up of a large number of components that are hard to parse correctly and contain information like certificates and redirects that require computer assistance to read, and finally, 3) there are many disparate data repositories that contain data pertinent to URL trustworthiness, e.g. DNS records of registration dates and phishing feed lists of known malicious URLs, which have a wide range of interfaces and locations making them non-trivial to use.

Therefore, as mentioned in Section 2, to judge a URL correctly a large range of URL features is required. For humans, the most indicative feature is the domain since they understand the context in which they see the URL and understand which organization's domain they would expect [5]. But predicting the destination of URLs is non-trivial. So, to best assist users in this task, we drew inspiration from the privacy policy nutrition label work by Kelly et al. [38] where a large number of privacy policy elements were put

into a food nutrition label like format. We thought that a similar approach might show important URL features to users in a consistent format that might allow them to learn over time. Thus, our goal is to develop a "URL nutrition label", including framing URL information in a way that assists users in leveraging their contextual knowledge and expectations to judge if a given URL belongs to the organization they expect. We call our design a URL feature report. Our report aims to address the following key design goals:

A. **Comprehensive.** The report should include enough information to help users make an informed decision about the safety of almost all URLs, including the ones in Section 2. To avoid overloading users, the interface should also present only necessary information [60].

B. **Support knowledge acquisition.** Each phishing indicator needs to have an explanation that helps non-experts understand the information as well as support higher level reasoning about it [16].

C. **Promote confidence.** Users need to have confidence in their final decision in order for the report to have its intended impact. Therefore, the report should support users in confidently making decisions on their own rather than blindly trusting recommendations. We aim to support users' confidence by providing conceptual and procedural knowledge (know-how) when explaining the phishing indicators [7, 53].

D. **Inspire Trust.** The report should inspire users to trust it by regularly providing accurate information and explaining its recommendations in a way that a user can verify themselves. Building trust with users when they need help will also improve their acceptance to taking the help [41, 67].

E. **Support comparisons.** The report should allow users to compare the aspects of the report to their own understanding and, potentially, against reports of other URLs. Supporting comparisons makes it easier for people to use the report for their tasks, the consistent positioning of information also allows them to learn the location of data for faster future access [38]. For example, a user may bank with Skrill and sees on the report that the domain is registered to an address on the Isle of Man, unsure if that is correct or not, they also ask for a report on Skrill's main website to see if it is also registered to the Isle of Man.

5 DESIGNING THE INITIAL REPORT

Reading a URL and making an accurate judgment requires accessing a wide variety of URL facts as well as understanding what those facts mean. These facts are consistent between URLs. As part of our design goals, we focus on selecting features that will help people most in making informed decisions about URL safety and how to present them to users. For an initial list, we started with the findings of Althobaiti et al. who reviewed phishing features used in human-training and automated detection research [5]. We then narrowed the list down to features that had been shown to be robust and had the potential to be human-friendly. We also excluded features that were highly technical and could not be combined with contextual knowledge to make informed decisions, for example, the DNS-based features [5].

I Don't Need an Expert! Making URL Phishing Features Human Comprehensible

CHI '21, May 8–13, 2021, Yokohama, Japan

Table 1: The threshold for the features used in the URL report. '-' means that the row will not be shown in that situation. Features were also added (★) and removed (★★) from the report due to design iteration changes.

	Feature	Red	Yellow	Green
Facts (mostly neutral)	Domain	-	-	-
	Category ★	Malicious	Web-host	-
	Registrar Location ★	-	-	-
Facts	Domain Popularity	-	< 300K	< 150k
	PageRank	-	0-3	4-10
	Domain Age	< 3 M	< 6 M	≥ 6 M
	In Search Engine	No match	Partial match	Match
	Encryption ★★	-	Unencrypted	-
Tricks	No. of External Domains	> 4	2-4	-
	No. of Short URLs in Chain	-	> 1	-
	Blacklisted in Chain	>0	-	-
	IP Address	1	-	-
	Non-standard Port	-	1	-
	No. of Subdomains	> 4	3-4	-
	Credential in Host	1	-	-
	Has Unicode '%'	1	-	-
	Hex Code in Host	1	-	-
	Non-ASCII	Mixed lang.	Non-ASCII	-
	Out-of-position TLD	A token	-	-
	Out-of-position Protocol	A token	-	-
	Out-of-position 'www'	A token	A sub-token	-
	Top Targeted in subdomain	A token	A sub-token	-
	Similarity to Top Targeted	-	1	-
	Similarity to Alexa Top 10k	-	1	-

To present the features in a comparable, well-arranged format as required by our design goals A and E, we split our initial design into four sections (See Figure 2).

5.1 Notice and reminder

On the top of the report we show the URL that was asked about for reference. For URLs that redirect, we display both the requested URL and the one that would be redirected to if they clicked the link. We also check the URL against known malicious URLs and clearly state if it is already known to be safe or malicious.

PhishTank and Google Safe Browsing both provide lists of reported malicious URLs, approved by security communities, which could be used to automatically alert a user [30, 79].

On the other hand, the Extended Validation Certificate (EV certificate) is used to mark a URL as safe because it indicates that a site's ownership has been verified by a certificate authority, which is sufficient evidence [51]. For other URLs, neither blacklisted nor with a verified owner, the safety is unknown. Thus, we avoid false positives and inspire trust in the report's safety information (Goal D).

5.2 Facts

In this section, we provide more details about the website's URL features to help users decide whether this domain indeed belongs to the expected institution or not. Each fact is presented with a fact name in bold on the left followed by a short description and the value on the right (Goal B). This section has the most consistent structure; however, we only show relevant features to achieve goal A of being comprehensive without overloading the users. Red

text is used to both highlight potential issues as well as provide guidance as to what the problem might be. For example, in the initial design, a Google PageRank of 0 is low suggesting that the page is probably not `apple.com`.

The first and foremost indicative feature is the domain itself. If the user is able to detect that the domain is not what they expected, they are likely to succeed in avoiding the attack. We adopted the common advice to search for the company's name in Google and look at the top few results. This works well because most modern search engines use popularity as an ordering metric [74].

Two more revealing components are the relative popularity of a website, which we determine using Alexa's most popular domains [4, 26, 50, 91], and the PageRank of a webpage [4, 26]. These two popularity scales both imply how popular a website is, but we present both to users because they do not always agree, most commonly in web hosting situations. If the domain is a web host, the Alexa popularity is the same for all pages and subdomains under that domain, whereas the PageRank may differ between pages under one domain. Finally, we use the domain age from Whois records in our report since users can efficiently compare it to the expected duration of the organization's online presence.

Encryption is another hint for safety, indicating if the connection with the server will be encrypted or not (https vs http). HTTPS adds encryption so users' information remains protected from unauthorized access in transit. Unfortunately, encryption is not a highly reliable indication of phishing websites [57], especially, since the introduction of LetsEncrypt [19] which gives free encryption certificates to anyone. It is, however, a useful security aspect.

5.3 Tricks

There are many ways to manipulate a URL to look legitimate. In this section we aim at identifying and pointing out these malicious tricks to users. Since the existence of tricks is very indicative of phishing, we check the URL for about 16 different tricks, many of them lexical. For example, we examine the URL for the existence of misspelling by comparing the domain with top targeted domains on PhishTank and Alexa's top 10,000 domains [77]. Each identified trick is then shown to the user as a row under the tricks section along with both a explanation of the trick and evidence (Goal C). To limit the length of the report, only identified tricks are shown, and if a URL has no tricks we simply state that no tricks were found.

In addition to misspellings, we also look at mixed language use, i.e. the existence of characters from conflicting alphabets. While no longer popular [22], the existence of IP address, hex or decimal characters often indicate phishing URLs. We also reverse IP addresses to the human-readable domain when possible.

Other tricks used by attackers to mislead users are also specified in our report such as the number of subdomains, '@' in the hostname, and out of position 'http', 'https', TLD, and 'www'. We also use the PhishTank's top targeted brands to identify if a targeted brand name is in the subdomain [10]. Additionally, we consider redirections, including multiple chained redirection. We determine the number of external domains [50], number of shortened URLs [29] and blacklisted URLs in that chain [46] and flag them as suspicious if they exceed a threshold. The full list of tricks is shown in Table 1.

Figure 2: Our initial design of the report which was shown to the first focus group (left) and the final design of the report (right).

5.4 Severity colors

We use a traffic light system with symbols to draw attention to important information. A colored circle on the far left indicates how problematic the value is, ranging from green (no issue) to red (known issue) whereas no symbol is provided for neutral information such as the domain. Red indicators are restricted for reliable features with few false positives such as an IP address in the hostname [6, 74, 82]. At the bottom of the report is a legend explaining the symbols and colors.

The color thresholds for each feature are adopted from automated detection research with more restriction for red color as presented in Table 1. The first block in the table includes context-related elements such as the domain name, which are mostly presented without color indicators. The second block shows facts that use different colors while the last block lists the tricks that are displayed when applicable in red or yellow.

We aim to support user's decision confidence through the use of color (Goal C). If a user sees many severe (red) color indicators in a potentially suspicious URL, then it should reinforce their confidence in their decision; conversely, many green rows may help them be confident of a URL's authenticity.

6 ITERATING WITH FOCUS GROUPS

We conducted a set of eight focus group sessions with Human-Computer Interaction (HCI) experts, security experts, and students from a UK university with a non-technology major. In the first focus group we took a co-design approach, but learned that users rationally just want the system to tell them if it is safe or not, which cannot be accurately done. So for later groups we focused on discussion and feedback. After each focus group, we used the feedback to iterate on the design.

Consequently, each group saw a slightly different version of the interface starting with the left version in Figure 2 for G1 and ending with the right image after G8. As the FG sessions progressed, we saw less new suggestions and more discussion of the content and phishing itself with G7 providing only minimal improvements, suggesting that we were reaching saturation or at least had created a reasonably understandable design. The study complied with our university's ethics procedure.

6.1 Participants

Our first three focus groups consisted of experts in HCI (G1) and security (G2, G3) (See Table 2). The purpose of these groups was to provide expert-level advice and to ensure that our design both matches strong HCI standards and is accurate in terms of security.

I Don't Need an Expert! Making URL Phishing Features Human Comprehensible

CHI '21, May 8–13, 2021, Yokohama, Japan

Table 2: Focus groups including their participants' expertise and group size.

Group	Type	Size	Gender
G1	HCI	3	2F, 1M
G2	Security	2	1F, 1M
G3	Security	4	4M
G4	Non-technical	4	4F
G5	Non-technical	4	3F, 1M
G6	Non-technical	5	4F, 1M
G7	Non-technical	5	5F
G8	Non-technical	5	3F, 2M

G1 and G2 were recruited from our University community. G3 was recruited from a local security workshop and contained security experts from industry. All three groups were unpaid and participated primarily out of interest in the project topic.

We recruited non-expert participants from The University of Edinburgh using various email lists including students from art, psychology, and physics while computer science and informatics were excluded. We chose this group because students are known for falling for these types of attacks meaning that they represent the type of people our report should support. They also rely heavily on the Internet for their studies [36] making them vulnerable to malicious links [81]. They were compensated £10 for 90 minutes.

6.2 Procedure

We first provided a consent form and collected demographics via a paper survey. In expert focus groups (G1-G3), we gave a 10 minute presentation on phishing, our motivation for the project, and common URL manipulation tricks. The presentation was provided to ensure all the experts are aware of the context which allowed us to best leverage their expertise. For average users, we excluded the URL manipulation tricks part of the presentation because we wanted their normal reaction to the reports without a prior knowledge of the tricks. As a warm up for all groups, we asked participants to share a recent experience with phishing communications including how they discovered that it was phishing. Doing so helped the participants better conceptualize what "phishing" meant, while also providing a set of concrete examples which were often referenced in later discussions.

After the initial discussion, we handed out two sheets of paper: an email containing a URL and the report about the URL. The email was provided so that participants would have the contextual information necessary to use the report. We used real non-malicious emails previously sent to the researchers as a start-point and replaced some of the existing URLs with malicious ones. Participants were told to imagine that they had received the email but were worried about it so they entered the URL into an online report generator and got the provided report. They were first asked to use the report to decide on their own if the message was real or phishing. Meanwhile, they were encouraged to mark elements of the interface that they found helpful or confusing with provided colored pens. Participants also had access to a range of co-design style materials including blank paper, stickers, colored pens, sticky notes, and scissors. After

everyone finished, the researcher moderated a discussion about the report. This process was repeated with another 2-4 email and report combinations depending on time.

6.3 Outcomes

6.3.1 **Overall Impressions**. Participants generally liked the report, both content and design, and found themselves well supported making a decision. Initially, they wished for a clear statement whether the URL is safe or not. After we explained that most URLs cannot be definitively classified that way, they tended to understand, but the concept did not come naturally. A G6 member, for example, started with the strong view that safe/unsafe presentation would be best, but after being presented with a URL from a real phishing email sent to most of the University population, he immediately identified it as phishing and recalled that his own anti-phishing tools had failed to identify it at the time. Our report showed him that the URL lead to an organization located in South Africa, which is not an expected location for a Microsoft URL.

Users had mixed opinions about the interface and its long-term usability. Early groups found the interface overwhelming and very long, but the perception improved through iteration on the content and presentation. The last few groups found the interface appealing and were even interested in using it either in their daily lives or as a tool when uncertain about a phishing message. They described the report as a useful tool to make a confident decision about the safety of a URL. A member of G5 for example explained its usefulness as "Alongside with intuition, there is relevant support and information for me to make decision on whether to trust the website subsequently". Similarly, a member of G7 explained: "I think for most people this would provide enough information to make informed decisions with a high level of confidence. Very interesting".

They had varied feelings about the trustworthiness of auto-detection tools in general. A G6 participant stated: "I trust the machine a lot, but I will trust myself more. This interface will help me to educate myself". This attitude is not only in line with the goal to support a user's decision, but also typical of phishing training which teaches users to not completely rely on the severity level of the indicators but also encourages them to consider their expectations. Similarly, a participant of G7 said: "It would work for helping classify URLs to safe and not safe. It is important to educate users and not just trust the software of taking decisions". Participants were also able to learn from the report itself as a participant in G7 described: "I learned to prioritize the results".

6.3.2 **Visual Appearance and Interaction**.

Symbols and Colors. Over the course of the focus groups, we adjusted the use and prominence of symbols and colors according to feedback. The first group, G1, found the colored symbols in the left column too small to read and were concerned that they would not be sufficiently obvious to readers and have unclear meaning. We therefore added descriptions such as "known issue" on a solid background color to make the meaning clear (See Figure 3). In the final design we removed these aids all together and added a legend to just below the summary so that it would be visible when needed.

With the new design, G2 was concerned that the colors might be inappropriate for color-blind users and G4 mentioned that different cultures might interpret red and green differently. In Chinese culture, for example, red is considered a happy color. To handle both issues, we converted to a color-blind friendly pallet and watermarked severity symbols to clarify the meaning [35]. The final report was tested on the iOS grayscale display which produces a colorless version of the report and allows to evaluate how the choice of colors would be for a colorblind person. Later focus groups had mixed opinions about the water-marked severity indicators. Some agreed that the symbols enhanced the meaning while others found them distracting. Therefore, we removed them from the final version and kept only the symbols in the legend.
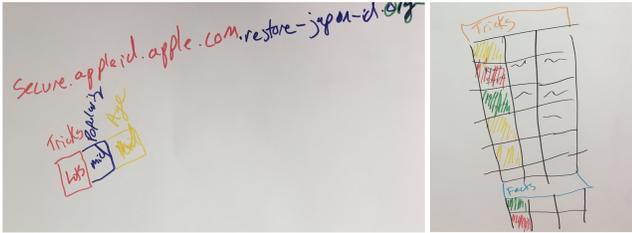
Figure 3: Both images were created by a participant in G1. The left image shows a URL with the domain, subdomain, and additional URL components highlighted. Below is the proposed summary with *Tricks, Popularity,* and *Age.* In the right image, the suggested new report structure continues with a list of tricks followed by URL facts.

Facts Order. Focus groups had several suggestions about how to adjust the presentation order of the rows. Members of G1 suggested we weight the presented features and present the most reliable features at the top. Given that we wanted to present features in a consistent order (Goal E), we instead located the strongest features at the top and put the tricks section above the facts.

G1 also suggested that ordering the facts by color indicator, beginning with the most severe (red) at the top. Another suggestion was ordering them based on each feature priority. We decided against these approaches because the relative value of facts depends on the information only the user knows. For example, popularity is a valuable feature if its value is unexpected, such as if the user thinks the URL is an Apple domain, but the popularity is low.

Another suggestion by G6 was to remove non-critical (green) facts so as to not overwhelm the user as we hide green rows for tricks. However, a G5 participant felt that green facts were easy to ignore if not needed. We decided against hiding green facts because doing so would make it harder to compare reports. It might also incorrectly make all URL reports look overly red and negative, leading to users incorrectly rejecting good URLs.

6.3.3 **Report Content***.* Additionally to the report interface, we iterated over the report components and wording. After each group, we incorporated suggestions to make the report better accessible and understandable.

Domain and Hostname Highlighting. At the top of the report, we highlight the URL domain to provide the domain information together with summarizing facts. Our initial design did not include domain highlighting since we list the domain as the first fact. However, after moving the manipulation tricks further to the top and adding a summary, the domain is not obvious enough. Therefore, we added domain highlighting at the top similar to how industry tools use it. Using highlighting similar to web browsers also keeps it familiar for non-technical users as member of G1 suggested that the concepts of domains and subdomains is too technical and lay users are unlikely to understand them. So the highlight in the domain row will help the user learn about the domain aspects.

Report Summary. As mentioned before, participants of almost all groups suggested that we provide a clear binary answer of whether the URL is safe or not.

When G1 understood that a binary answer is not possible for all URLs, they instead suggested an overall score or severity bar for URLs that could be easily used for judgment. Similarly, a participant from G3 wanted some sort of classification such as maliciousness percentages. We felt that a single overall score would mislead users and not encourage them to read and learn from the presented information. Instead, we tried to use a combination of clearly visible colors and added a summary highlighting key issues to the top of the report to help users get the requested high-level sense safety.

Security groups G2 and G3 saw two versions of the report, one with and one without a summary. Both thought the summary was a good idea and debated about which topics should be included. They liked that the summary told them which features to focus on first. Since both expert groups considered it beneficial, we added a summary section at the top of the report for the remaining groups and continued to iterate on its presentation.

Several iterations later, we settled on four summary boxes: used manipulation tricks, search result, domain age, and domain popularity. *Manipulation tricks* was chosen because their presence is a strong indicator for phishing [70] and the meaning was clear to most focus groups. The *search results* box indicates whether the URL appears in Google top results when searching for it. Both *domain age* and *popularity* are common features that made sense to users and were generally well understood in focus groups. Groups G2 and later considered the summary to be quite useful. Initially, they felt that such a summary definitely required the rest of the report for explanation. However, after reading the report, they quickly understood the meaning and had no difficulty using it when reading future reports.

Tricks. G1 found that the tricks section is very useful, especially the clarification of what is wrong, but the facts section did not adequately explain the meaning of the information. A G1 participant said: "If I show the tricks to my gran, she will say yeah cool, but if show her the facts she wouldn't know what is going on". The tricks are indeed a stronger indication of a malicious link, potentially eliminating the need to look further [27]. Thus, as suggested by G1 in Figure 3, they should appear at the top. Especially since for later groups key facts such as age already appeared in the summary.

One of the comments from G6 recommended removing the tricks we found in a verified (safe) URLs because it will distract the users; however, we decided that displaying tricks even for safe URLs will

I Don't Need an Expert! Making URL Phishing Features Human Comprehensible

CHI '21, May 8–13, 2021, Yokohama, Japan

help users to learn to judge the features' importance for making informed judgments. For example, non-ASCII characters can occur even in safe URLs after the evolution of Internationalized Domain Names; thus, based on the context, users can use the feature to judge if a URL is safe or not. This decision supports our design goal B.

Location and Category. We added a location field to better support users in identifying any inconsistency between the domain location and the expected location. Usually, the stated physical location of malicious domain registrars differs from legitimate ones [70]. However, understanding the meaning of the location was challenging for focus group members with some users interpreting location based on the trustworthiness of the country. For example, in G4 one of the participants stated: "Apple in Japan, so what? Japan is not questionable". She was confused and thought that the location referred to the server location rather than the location of the organization. We thus adjusted the description to clarify that it was the location of the domain owner.

A security expert from G3 commented that one of his common approaches to detecting phishing websites is to look up the URL's category on FortiGuard which categorizes URLs into groups such as shopping or governmental organizations. This feature can be used to check whether a suspicious page has a similar category as the expected one [72]. Additionally, FortiGuard categorizes the full host name of provided URLs in case a domain includes different subdomains such as WordPress.

Web Hosting. As mentioned previously, some popular domains host content for others. As a result, it is possible for the domain to be popular and registered a long time ago, but the specific page or subdomain is malicious. Discrepancy between Domain Popularity and PageRank should highlight this situation to users, but focus group participants found the discrepancy confusing rather than helpful. So mid-way through the focus groups we started experimenting with wordings suggested by the participants to directly explain the issue. We tried several approaches, including dividing the facts into page and domain facts or creating a large warning on the top of facts. In the final design, to determine which domains automatically offer web hosting services, we used the FortiGuard website categorization service [23]. Then, we hide the domain-only facts (location, age, and popularity) and in their place we state that: "This domain hosts multiple sites, some are good and some may be problematic. Usually only small companies and personal websites are hosted by other domains." In general, the focus group participants liked this warning and felt that it was very important and useful for their decision making. G7 felt that they lacked direction on where to look after seeing it. They understood that there might be a problem but they were unsure how to distinguish between safe and malicious hosted sites. Conversely, a G6 participant commented how the warning helped him to be confident visiting personal pages since he expected them to be hosted on other sites.

Domain Popularity and PageRank. The domain popularity is drawn from Alexa and is an indicator of how often people visit the domain, with the most visited domain being ranked 1. The PageRank roughly indicates how often other pages link to this page [11]. Here, the most linked to pages have a rank of 10. The two measures are naturally easy to confuse as they both deal with popularity. They also have inverted scales with 1 being good for domain popularity and bad for PageRank. In the initial design, we tried to explain the difference in words. However, G1 suggested a visual range for the numbers instead to indicate clearly which value is problematic. To further emphasize the scale, we added colors to the range. G2 and G3 saw colored bars with raw values below and showed no difficulties reading it. G5, however, commented that the numbers looked like they were written in error due to the opposite directions of the popularity scale numbers. After iterating several approaches on later groups, we settled on removing the numbers entirely and simply showing "popular" and "not popular" as the ends of the ranges.

Differentiating between the domain popularity and PageRank was challenging for all groups except the security ones (G2, G3). Domain popularity made the most sense, likely because it is roughly based on the number of people visiting the site, which is easy to explain and understand. The concept of PageRank, however, was much harder to grasp for participants even when verbally explained. Also, the "domain" versus "page" difference was subtle leading to difficulties articulating why a page might have a different popularity from the domain.

Eliminating one or the other was also not an option as our security groups explicitly mentioned how useful it was to include both since they are fundamentally different measures. For example, sites like WordPress have high domain popularity even if a hosted page has a low page rank. So it is possible for a very popular site to be hosting an unpopular malicious page. To reduce confusion, we iterated on the wording to improve section explanations. G8 in particular was shown several wording options and provided extensive feedback on how to express the concepts more clearly. However, even in the final design, the difference between domain popularity and PageRank is still hard to grasp quickly.

Encryption. Initially we thought that encryption would be useful information in the report. In the encryption component we stated if the connection was encrypted or not. But we were also concerned that user would equate encryption with owner validity, so we added, "This URL is encrypted but we couldn't verify the owner". G1 understood this concept and explained: "If this is an Apple URL, you would expect to have a verified owner".

However, we found that showing encryption information may mislead users. For example, a G2 participant marked a legitimate URL as phishing because she did not think a reputable company would use a HTTP connection. When we tried incorporating ownership information as well to provide a more complete view of encryption, participants just became more confused. A participant from G4 asked "why is it a good sign if you cannot verify the owner of the organization?" after seeing that the connection was encrypted (green) but the website owner could not be verified due to the information not being in the SSL/TLS certificate. Showing information that could mislead the user violates our design goals D of inspiring trust and C of confidence by showing correct information that will lead to the right decision. Therefore, in line with our design goal *A* to avoid overload, we removed encryption entirely from the report. The only exception are Extended Validation (EV) Certificates. These are TLS/SSL certificates which have gone

through an extensive ownership verification process. When we encounter a EV we put a green check in the summary and a statement like: "MoneyGram International Inc. verified its ownership of this domain." The focus groups liked this feature and found it helpful. A member of G6 stated "that's all I need".

6.3.4 **Ideas for Workflow Integration**. Participants were enthusiastic about having easy access to the report and explained how they would integrate it in their work flows. A G2 participant suggested developing a browser plugin or email client, which shows an option to "Look up this URL" when a user right-clicks on a URL and would show the report in a new tab. They explained that a plugin would be best for them as they thought people are "too lazy" to visit a website. A G5 participant suggested providing the summary for each link on a page automatically.
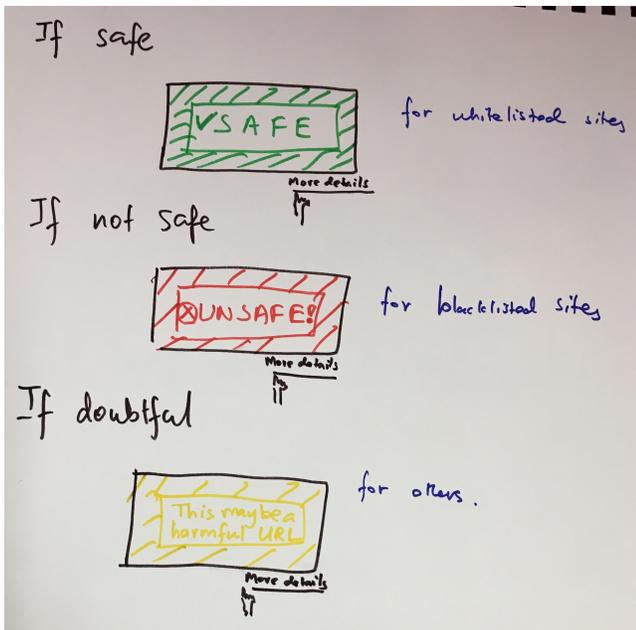
Figure 4: A G6 participant began designing a report showing "safe", "unsafe", "doubtful" and adding an option for "more details".

Both G3 and G6 suggested a tiered design where initially only limited data like a score or the summary is shown with an option to view the full report (See Figure 5); possibly limiting the type and detail of data depending if the user is novice or advanced, e.g. an IT helpdesk employee. In G6, a participant suggested that they would like to see a high level safety flag before seeing any reports (See Figure 4): "Give me three flags (green-safe, red-blacklisted, yellow-unsure), then give me the ability to drill down the summary and if I want more details, give me a link to the webpage". Adding the high level estimate will not "overwhelm [them] with the details at a starting point" when using everyday. This suggestion contributed to the "report summary only" design we evaluate in later sections.
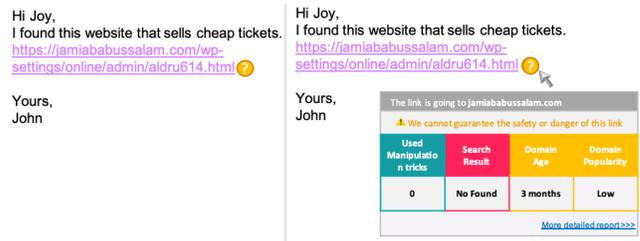
Figure 5: An example of how the report could be integrated into a user's workflow as suggested by G3 and G6. The image shows a small circle flagging the URL and a version of the report summary when hovering over the symbol with the option of opening the full report.

6.4 Expert Interview

We showed an early version of the report to two experts who are designing anti-phishing training for the University and responding to requests regarding phishing from the front-line helpdesk. Overall, they were very positive about the report design. Although they felt that the report is too complex for average users, they thought that it might be of great use to the help desk staff who has to handle reports of potential phishing URLs.

We also asked them if there was any part of the report they felt was unnecessary. They commented that the encryption component, which we later removed, is not needed to judge if a URL was phishing and will likely confuse users. Otherwise, they felt that all other parts of the report were necessary to make an informed decision.

7 FEATURES VALIDATION

While the features selected for the report are known to be strong phishing features, we still wanted to test the visual appearance of the report on known phishing and safe URLs. We analyzed 6877 URLs from four data sets, two phishing (PhishTank [58], Open-Phish [59]), and two safe (DMOZ [14], ParaCrawl [42]), to explore what the report could realistically look like for users. The sets of safe URLs contained 2615 URLs, of which 592 (23%) were excluded due to '4xx' and '5xx' response codes. For phishing URLs, we collected a total of 4262 URLs, of which 1645 (39%) were excluded due to unsuccessful response codes. Phishing URLs from OpenPhish and PhishTank were processed every two hours to extract features while the pages were still live.

To reduce load on the reader, reports only include data relevant to the particular URL. For example, tricks do not appear in every URL, so we hide them by default and only show them if they are present such as the presence of non-ASCII characters. So while there are 23 possible rows, only 6.7 rows were shown on average (*Min* = 4, *Max* = 10), with phishing (*Mean* = 6.8) and safe (*Mean* = 6.5) having similar row counts.

Tricks were rare for safe URLs with only 77 (3.8%) showing one trick and the remainder having no tricks. Phishing URLs more commonly had tricks with 868 (33.2%) containing between 1–3 tricks. The primary cause of tricks for safe URLs was the similarity between the domain and one of top 10,000 popular domains (30 URLs), a phenomenon already seen in previous research [78]. Thus,

we only show a yellow indicator for this feature to avoid false positives and hand the task of comparing and deciding whether this is expected in their context to the user.

Looking at the red color in the reports, 30.2% of safe URLs and 88.0% of phishing URLs had at least one red row. Of the safe URLs with a red row, 99% did not appear in Google's top 10 search results, causing the red. Search result may therefore vary a good bit, making it a difficult feature to interpret. However, it is still a good indicator for the illegitimacy of a page, which is why we decided to keep it. We also found that 23.4% of phishing URL reports had only green rows. Further examination showed that only three of them were compromised websites, while the rest were URLs that redirected to safe URLs; thus, the features referred to safe URLs which were indeed safe. This redirection tactic is used by attackers to serve advertisements and then send the user to the expected safe website [77]. Therefore, it is important for the report to urge users to visit the shown final link instead of the original one. We also considered using the color frequency in each report to predict whether a URL is safe or not. Applying linear regression, we found that the frequency of each color in the report significantly predicts whether a URL is safe or not ($R2 = 0.44$, $F(4635) = 1238$, $p < .001$) with Red ($\beta = -0.48$, $p < .001$), Yellow ($\beta = -0.17, p < .001$), and Green ($\beta = 0.11, p < .001$).

Finally, we measured the features' redundancy to ensure we are not showing unnecessary features. We computed pair-wise correlations between features using their severity color as presented to users as the feature's value and found no correlation between any of them.

8 ONLINE STUDY

Focus groups are an excellent way to get rich feedback but a poor way to get a truly wide range of participants. To address that gap, we decided to use an online survey to test the clarity of report content as well as its ability to support users in making accurate safety judgments about a URL. We used a between-subjects experimental design where each participant saw one of: the full report, just the summary, and just showing the URL with domain highlighting.

8.1 Questionnaire Instrument

For all conditions, the survey started with informed consent. Participants were then asked how familiar they were with 13 website terms and 6 companies followed by study instructions to not visit any of the links and only read them. A question then tested whether they had read the instructions and terminated the study if they failed it twice. To test their existing URL-reading skills, participants were then given three URLs and asked to choose which company those URLs lead to. The first URL has Google in the pathname, the second has Facebook in the subdomain part, and the third is for New York Times which uses an abbreviation of the brand name.

Participants were then shown 6 URLs. For each URL, they were told to imagine that they wanted to visit a particular company, given a brief description of that company, for example, "eBay, an auction and consumer to consumer sales website", and then asked if the URL "leads to a page owned by the above company or is it a malicious URL". In the domain highlighting condition the URL was domain highlighted in the question, for the other conditions

a report was provided and participants were encouraged to use it when answering. Participants were then asked how confident they were in their decision followed by a question about what most influenced their decision. After answering questions about all 6 URLs, all groups were asked a set of comprehension questions to make sure they did, or could, understand the content of the report or the report summary, the comprehension questions were multiple choice and asked the participant what they thought the different parts of the report meant. Full and control were asked about the full report elements, and the report summary group was asked about the report summary elements. The answer options were drawn from common misunderstandings observed in the focus groups. The survey ended by collecting background information. We used a phishing susceptibility scale from Wright & Marett [87] with 5 subscales to test participants' computer self-efficacy, web experience, trust, risk beliefs, and suspicion of humanity. Also, we included basic demographics questions on age, gender, and highest degree obtained.

8.1.1 Study conditions. In the following, we describe the conditions and the questions that differed between them.

Domain highlighting: In this condition, we showed the full URL with the domain highlighted and asked participants if the URL leads to the given company name. Existing research already shows that users cannot read URLs unaided [2, 4, 6, 69]. Domain highlighting has already been adopted by several browsers, e.g. Safari, making it a state-of-the-art approach that has been shown to help users decide on URL safety [88]. Thus, we chose domain highlighting as the control condition. To determine what most influenced participants' safety judgments, they were asked to select up to three of: the domain, the protocol (https), the URL path and query strings, their prior knowledge, and their familiarity of the company's URL.

Full report: This condition is the longest. Before showing the 6 URLs to participants, we first showed them a fictitious report with obviously fake values and asked 10 questions in a random order about the features, i.e., "How old is this website?". Doing so gave participants some basic practice with the report and allowed us to test for any serious misunderstandings about how to use it.

To determine what most influenced the participants' safety judgments, they were shown a list of the different report elements along with "my own prior experience reading URLs" and asked to select up to three that most influenced their decision.

Finally, we asked a set of 7-point Likert assertion questions to measure the report usefulness and satisfaction, loosely based on the SUS, such as "I can learn a lot about phishing using this report" and others drawn from focus group participants' opinions about the report. We ended with an optional free text comment section.

Report summary: Many participants in our focus groups suggested to show the report summary when a user hovers over a link. The idea has merit, so we evaluate it here as a middle option between domain highlighting and showing the full report. Participants in this condition saw only the summary part of the report, with no option to see the full report.

To determine what influenced the participants' safety judgments most, they were shown a list of answers including the summary report boxes, elements of the URL, their own prior experience, and

the colors. After answering questions for all 6 URLs, they were asked about the meaning of each of the summary report elements, with multiple choice answer options derived from common focus group misconceptions and an other option. Finally, they were asked the same 7-point Likert questions about the report usability as the Full Report condition.

We categorized URLs into 3 reading difficulties levels: (1) Parse and Match: any URL knowledgeable person can find the domain and compare it to brand name, (2) Domain Knowledge: a URL knowledgeable person has to know which organization a domain belongs to before judging the URL, and (3) Misleading Flags: URLs have information that may mislead participants to misjudge them. For each category we have two organizations, one popular and one not popular based on the top targeted domains on PhishTank, and for each organization, we have a phishing and a safe URL (see Table 3). With 6 organizations, we ended up with 12 URLs in total. For each condition, participants were divided into two groups with every group being shown one link of each organization at random, six URLs in total.

The presented URLs are real-life URLs with the phishing taken from our analyzed data set in Section 7. We made minor manipulations to control some variables. They have an approximately similar length, https protocol, as well as path and query strings. To reduce a bias in the selected URLs, we ensured that the color indicators were in-line with real observed color combinations from the data set. As we had abnormal false positive search results, we included one safe URL with a red search result. For participants' safety, we selected phishing URLs that were no longer active. Additionally, in case they clicked on the links, we added a hyperlink which leads to a page belonging to the research group about the danger of clicking on these links.

8.2 Survey Results

Participants. We recruited participants from Prolific for a 30 minute study on phishing. The time estimate was based on a short pilot study. We limited participants to those with approval rates above 90% and Native English speakers to avoid language issues. We then excluded those who did not answer the attention check questions accurately.

We had a total of 153 participants (*domain highlighting* = 51, *report summary* = 50, *full report* = 52), 63.4% were female. Participants had an average age of 31.89 years ($\sigma = 9.9$). Compensation was £3.5. The average time required to complete the survey was 18.26 minutes. For the prior URL reading skill, on average 1.6 of the questions were answered correctly with only 14 answering all questions correctly (9%) and (15%) not answering any URL correctly.

Accuracy of safety judgment. We found that participants in general were able to accurately judge URLs' safety. The average accuracy was highest for the full report (5.5/6, SD = .28), with the report summary also doing well (4.96/6, SD = .38) and the domain-highlight doing the worst (3.88/6, SD = .48). The false positive (FPR) and false negative (FNR) rates are also encouraging with all participants more likely to incorrectly mark safe URLs as phishing: *full report* (FPR = .12, FNR = .05), *report-summary* (FPR = .23, FNR = .11), and *domain-highlight* (FPR = .44, FNR = .26)).

We used an ANOVA followed by applying Cohen's F for the effect size to test if the three conditions (domain-highlight, full report, and report summary) impacted the accuracy of participants' judgments and we found a statistically significant impact of the condition on the judgment accuracy ($\alpha = .01$, $p < .001$, $r = 0.29$). We then computed follow-up t-tests and found a significant difference between all three pairs of conditions: domain-highlight and full report ($p < 0.0001$, $d = 0.7$), domain-highlight and report-summary ($p < 0.0001$, $d = 0.4$), and report-summary and full report ($p < 0.001$, $d = 0.3$).

We separately tested if any other variables impacted accuracy using ANOVA as well. These variables are the time spent on the question, the condition, and the level of difficulty of each URL, the actual safety (malicious/trustworthy), participants' confidence in their answer, their familiarity with the company, the company the URL leads to, their prior knowledge of URL reading, and their phishing susceptibility factors. We found that the accuracy of users' judgments is significantly impacted by the condition, the URL safety, and the URL hardness level ($\alpha = .01$, $p < .001$), with a large effect size for the condition ($r = 0.31$) and small for the other two (0.16 and 0.13). The remaining variables had no significant impact on the judgment accuracy.

We tested URLs associated with 6 organizations as shown in Table 3 where each organization had a phishing and safe URL associated with it, resulting in 12 URLs tested. For all URLs, the full report has higher accuracy than the domain highlighting. The summary report is slightly mixed, mostly sitting between the domain highlighting and the full report, but occasionally showing more accuracy than the full report. The four "parse and match" URLs are theoretically the easiest to determine from only reading the URL string, which is mostly born out with the high accuracy for even the domain highlighting condition. The exception, `bestchange.ru`, was incorrectly marked as phishing by the majority of participants in the domain-highlighting condition. For the "domain knowing" URLs, a user has to know the correct domain of the organization to be able to accurately judge safety if unaided. Here the `email.microsoftonline.com` URL was the most challenging for all conditions. The `bìttrêx.com` URL was also challenging for the domain highlighting group, possibly because they were unsure if the non-ASCII character (Vietnamese) should be there or not. In the misleading URLs, Tripod was a confusing case where the safe URL positions the brand name in the subdomain while the phishing URL includes the brand name in the domain but actually is a hosting service for other websites. Similarly, the `fb.me` legitimate short URL confused many of the domain-highlight participants.

Comprehension of the report elements. Full report participants were able to provide a correct answer for 7.73 out of 10 report comprehension questions on average.

The most common error was in regards to the location feature where 57.7% (30/52) of participants indicated that the location means the physical location of the server they were contacting rather than the self-reported location of the organization that registered the domain.

PageRank continued to be a source of confusion with 34.62% (18/52) of participants providing an incorrect answer. They commonly confused PageRank with domain popularity indicating that

I Don't Need an Expert! Making URL Phishing Features Human Comprehensible

CHI '21, May 8–13, 2021, Yokohama, Japan

Table 3: The URLs used in the online study to judge URLs. Each condition was divided to two groups and see only one URL for each company.

URL	Hardness	Popularity	Safety	Group		% of participants who accurately judged safety		
				G1	G2	Highlight	Full report	Summary
https://resolutioncenter.ebay.com/policies/?id=123	Parse and match	Popular	Safe		X	81	100	81
https://itmurl.com/www.ebay.co.uk/item=30327559652			Phish	X		96	96	88
https://www.bestchange.ru/exchangers/mkt=en&id=234		Unpopular	Safe	X		44	88	83
https://www.bestcnange.ru/exchangers/mkt=en&id=234			Phish		X	92	100	92
https://email.microsoftonline.com/login/?mkt=en-GB	Domain knowing	Popular	Safe	X		64	73	50
https://www.365onmicrosoft.com/login/?langua=en-GB			Phish		X	73	100	96
https://international.bittrex.com/account/?id=2423		Unpopular	Safe		X	73	85	65
https://international.bittrêx.com/account/?id=2423			Phish	X		56	96	92
https://fb.me/messages/t/788720331154519	Misleading flags	Popular	Safe		X	15	92	85
https://l.facebook.com/l.php?u=http%3A%2F%2F67.23.238.165			Phish	X		60	100	83
https://www.tripod.lycos.com/pricing/?plan=free-ad		Unpopular	Safe	X		56	92	96
https://webmasterq.tripod.com/pricing?plan=free-ad			Phish		X	65	77	81

the value meant how popular the site was rather than the individual page. One option we are considering for future work is to hide PageRank when it is in alignment with the popularity (both high or both low) and only show it when it is different with direct explanations of how the miss-alignment could be problematic.

All questions in the section had an option to indicate that the description was confusing. The web hosting element confused participants most with 12% indicating that the description is unclear and 73% (38/52) of them answering it correctly. The result suggests that the new wording is mostly working though there is some room for improvement.

In the summary group, participants were able to provide a correct answer for an average of 4.69 out of 6 questions. The most common error concerned the web hosting feature with 57.7% (38/52) answering correctly.

Report usefulness and satisfaction. We asked participants to pick the report elements that they found most helpful after deciding about each URL to get a sense of if they were relying on a small set of features or using the whole report. Participants chose different information for different URLs. "Domain age" was the most influential feature for 4/12 (eBay and Bittrex Safe URLs and eBay and Microsoft Phish URLs), "Manipulation tricks" for 3/12 (Facebook, Bittrex, and BestChange phishing URLs), "Domain popularity" 4/12 (BestChange, Microsoft, Tripod and Facebook safe URLs), and "Search result" for 1/12 (Tripod Phishing URL). For the safe Microsoft URL, we displayed a warning that 'microsoftonline' is similar to 'Microsoft' and it does not match Google's top search results, however, it was still not the most helpful feature. The fact that participants were looking at different elements for different URLs shows that they were making use of the full report and balancing and weighing features, instead of just sticking to the one aspect that made the most sense to them. Participants indicated that they used prior knowledge but it was not in the top three features for any of the URLs.

The self-reported answers for satisfaction indicate that the full report (*Mean* = 5.78, *Median* = 6, *SD* = 0.72) was more preferable than the summary-report (*Mean* = 5.36, *Median* = 5.57, *SD* = 1.13). For the full report, participants found that the survey taught them about phishing, using the report would help them, they understand

the report content, and did not need to learn new skills to use it. However, in the summary group, users felt they needed to learn a lot of things before using the report.

9   LIMITATIONS

Our report aims to support users in deciding if potential phishing URLs are or are not safe to click on. Therefore, our work is limited to the types of information available to a user in advance of loading the page itself and does not include solutions that look at the safety of the resulting page such as identifying compromised code or layouts that are visually similar to frequently targeted sites.

We endeavored to put together focus groups looking at HCI, Security, and non-technical students to get a range of opinions and experience. We also conducted multiple focus groups to offset some of their known issues, such as participants getting distracted by irrelevant topics, or being influenced by a dominant peers' opinions. We also ensured that a moderater was present to keep the groups focused and on topic. Finally, we also used an online survey to further verify our focus group findings on a large scale.

Prolific, similar to other online micro work sites, is known to have users who are more computer literate than the average internet user, they also tend to be more privacy aware [37] which may impact their knowledge of URL reading. Though recent studies of online workers suggest that online workers, including Prolific workers, still struggle with predicting where a URL will go [2, 69]. This type of user is also the type of person that less skilled people may go to for assistance [56], so supporting them well is likely to have a broader positive impact.

The phishing feature list we used is also not exhaustive. There are a wide variety of features used to detect phishing URLs, and many of them appear in only one or two papers. To mitigate the issues, we made use of existing reviews of the range and accuracy of features [5]. However, features that have been mostly tested on automated systems are not necessarily the best features possible for people. In this work we have started with known good features and narrowed in on those that best support people, but it is possible that other features exist that support people better but did not show up in our review.

10 DISCUSSION

URLs are known to be complicated for users to read unaided making it challenging for them to accurately judge the safety of a URL, even when they are aware of context like what website they expect to visit. Our report design is intended to support users in making informed decisions about URLs that they are concerned about. While many users have access to some form of expert advice, either through their employer's help desk or through the help mechanisms of the targeted company, that expert advice takes effort to engage with and the response will likely be slow if it comes at all. Our report is intended to help users help themselves when they encounter a URL they consider suspicious by allowing them to make use of many of the same data sources used by experts.

Our focus group participants were able to use the report by utilizing their own contextual knowledge and their expectations of the organization the URL represents. For every URL, they picked a different feature that influenced their answers, giving them the flexibility to decide what phishing feature is the right clue for each case. Our online study showed similar findings with both the full report and summary report conditions able to more accurately identify phishing URLs than users who only had domain highlighting to help them. Online participants also made use of many elements of the report to make their decisions, supporting the view of our expert focus group participants that a large number of features really is needed to accurately judge URL safety.

User Empowerment. In security work it is easy to take a paternalistic approach with end users where the security expert knows best, gives users minimally explained rules to follow, then gets upset or blames them when those rules are not followed as expected. Part of the goal of this work is to shift that interaction to one where the users are given more knowledge about the specific situation as well as more control while also being asked to remember less facts and rules. Ideally the report structure will also support user confidence and fast feedback where if they think a URL is potentially unsafe, they can quickly gain more information about it. While tools that support users in making decisions about safety do exist, there are few of them and they are mainly aimed at expert users who already know terms like "domain", most of the other tools instead focus on providing users with binary decisions with minimal to no reasoning provided [1]. While the binary advice is what users would like, having too many false positives also leads to users no longer trusting tools [3], so such tools have to be careful about which URLs they mark as unsafe.

Empowered users are also important because the context each person works in is different making it nearly impossible to provide one set of comprehensive rules that work well for everyone. The behaviors of phishers also adapt and change over time [22]. Asking users to keep all this information in their head is impossible [41], but with the support of a tool, users can always be basing their decision on up-to-date information and guidance. Over time they may also start learning the tricks and indicators that are most useful for the type of content they see.

Training. While the primary purpose of the report is decision support, it is also has a potential for education. Existing education approaches tend to focus on training the user either through a dedicated up-front training [53, 85] or through smaller training embedded in existing work practices. In both cases, a security professional decides what is most important for the user to learn and bases their training around that. The timing of the training is also outside the user's control, either dictated by the organization or appearing in their normal work unasked for. The design of our report is intended to fill a gap where instead of telling users what and when they should learn about phishing, we instead wait till they have a specific case that they would like to learn more about. Similar to earlier work by Kumaraguru et al. [44], this type of training is timed at a *teachable moment.* But where earlier work has focused on the moment after the user fell for a phishing attack, our report instead focuses on a time point when the user is curious and seeking advice.

Use Cases. The report is meant to support a user who has already identified a potentially fraudulent URL, but is uncertain about it and either does not have access to experts or does not have time to wait for expert feedback. Our focus groups, however, also suggested other uses cases, such as using it to help explain a safe/unsafe decision to someone else. People often reach out to peers when uncertain about potentially malicious communications [56, 66], having a report like ours would enable people to provide not only a recommendation but also a reason behind that recommendation. Similarly, participants thought that the report might be useful for help desk workers who may not be security experts, but are regularly asked about potentially fraudulent communications. Such a report might improve their ability to respond to requests accurately as well as provide useful feedback to users. Finally, they suggested adding this type of report to automated systems as a better way to persuade users to adhere to the warning [67]. Several of these use cases would be interesting to study in future work.

Focused Attention. Initially we set out to create a short and simple report that users could easily use at a glance to understand the safety of URLs. One hard lesson we have learned over the course of this project is that URL safety is a complex topic. Many of the basic concepts needed to understand the evidence require explanation for an end user to be able to understand them and use the knowledge correctly. Because the URLs being looked at are expected to have already gone through a phishing filter, what the user needs the most help with is comparing their own contextual knowledge, which was not available to the automatic filter, to the information about the URL. This type of comparison is necessary at this stage in the process and requires focused attention from the user to accomplish the task. Our report is designed to support users in this process by clearly explaining the different elements, supporting comparisons, and enabling users to more efficiently use the report on subsequent accesses. However, it is important to recognize that the report is best used in cases where the user is proactively seeking more knowledge about a URL rather than pushing the report into their workflow unasked for.

Deployment Potential. While the main contribution is the report design itself, we also consider the practicalities of potentially implementing and deploying it. In order to test the report content with real data (See Section 7), we programmatically sourced the feature data and computed what would be displayed in the report.

I Don't Need an Expert! Making URL Phishing Features Human Comprehensible

CHI '21, May 8–13, 2021, Yokohama, Japan

Our code automatically queries several third party APIs, such as PhishTank and Google Safe Browsing, to retrieve features not possible to extract from the URL itself. Using such APIs in a deployed system would be practical since they are continuously updated and, thus, require minimal to no ongoing maintenance for the report accuracy. Other report elements like the tricks and the explanations may require more expert involvement to maintain over time, but the effort of doing so is not large. To further explore the potential of deployment, a masters student build a prototype of the report as a Chrome browser plugin as a thesis project [83]. Their system allowed a user to request a variation of our report for any URL they saw inside the browser. As the prototype was a proof-of-concept, it is not suitable for real-world testing. But it did demonstrate the feasibility of integrating such a report into common user tools, like browsers.

11 CONCLUSION

We have presented the design for a new URL feature report which assists users in deciding whether a URL is malicious or not. The reports are intended for users who are trying to judge a URL's safety as part of a primary task. To refine the report's design, we conducted 8 focus groups with experts in HCI, experts in security, and average users. Finally, we conducted a survey to measure the readability and effectiveness of the report. We found that participants could generally read the reports, understand phishing features, and use them to successfully decide if a URL is malicious or safe. However, some participants still had difficulty understanding the more complex concepts such as PageRank and location.

REFERENCES

[1] Sara Albakry and Kami Vaniea. 2018. Automatic Phishing Detection vesus User Training, Is there a Middle Ground Using XAI?. In *Proceedings of the SICSA Workshop on Reasoning, Learning and Explainability (CEUR Workshop Proceedings)*, Kyle Martin, Nirmalie Wiratunga, and Leslie S. Smith (Eds.), Vol. 2151. CEUR-WS.org, Aberdeen, Scotland, UK, 1–2. http://ceur-ws.org/Vol-2151/Paper_P2.pdf

[2] Sara Albakry, Kami Vaniea, and Maria K. Wolters. 2020. What is this URL's Destination? Empirical Evaluation of Users' URL Reading. In *CHI '20: CHI Conference on Human Factors in Computing Systems*, Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy Cockburn, Ignacio Avellino, Alix Goguey, Pernille Bjøn, Shengdong Zhao, Briane Paul Samson, and Rafal Kocielnik (Eds.). ACM, Honolulu, HI, USA, 1–12. https://doi.org/10.1145/3313831.3376168

[3] Hazim Almuhimedi, Adrienne Porter Felt, Robert W. Reeder, and Sunny Consolvo. 2014. Your Reputation Precedes You: History, Reputation, and the Chrome Malware Warning. In *Tenth Symposium on Usable Privacy and Security, SOUPS*, Lorrie Faith Cranor, Lujo Bauer, and Robert Biddle (Eds.). USENIX Association, Menlo Park, CA, USA, 113–128.

[4] Mohamed Alsharnouby, Furkan Alaca, and Sonia Chiasson. 2015. Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human-Computer Studies* 82 (2015), 69–82. https://doi.org/10.1016/j.ijhcs.2015.05.005

[5] Kholoud Althobaiti, Ghaidaa Rummani, and Kami Vaniea. 2019. A Review of Human- and Computer-Facing URL Phishing Features. In *European Symposium on Security and Privacy Workshops, EuroS&P Workshops*. IEEE, Stockholm, Sweden, 182–191. https://doi.org/10.1109/EuroSPW.2019.00027

[6] Kholoud Althobaiti, Kami Vaniea, and Serena Zheng. 2018. Faheem: Explaining URLs to people using a Slack bot. In *2018 Symposium on Digital Behaviour Intervention for Cyber Security (AISB 2018), April 5 2018*. University of Liverpool, Liverpool, UK, 1–8. http://aisb2018.csc.liv.ac.uk/PROCEEDINGS%20AISB2018/Digital%20Behaviour%20Interventions%20for%20CyberSecurity%20-%20AISB2018.pdf

[7] Nalin Asanka Gamagedara Arachchilage and Steve Love. 2014. Security awareness of computer users: A phishing threat avoidance perspective. *Comput. Hum. Behav.* 38 (2014), 304–312. https://doi.org/10.1016/j.chb.2014.05.046

[8] Krishna Bhargrava, Douglas Brewer, and Kang Li. 2009. A study of URL redirection indicating spam. In *Sixth conference on e-mail and anti-spam CEAS*. Steve ShengâĂŽs Publications, California, USA, 1–4. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.536.2821

[9] Jim Blythe, L. Jean Camp, and Vaibhav Garg. 2011. Targeted risk communication for computer security. In *Proceedings of the 16th International Conference on Intelligent User Interfaces, IUI*. ACM, Palo Alto, CA, USA, 295–298. https://doi.org/10.1145/1943403.1943449

[10] Giovanni Bottazzi, Emiliano Casalicchio, Davide Cingolani, Fabio Marturana, and Marco Piu. 2015. MP-Shield: A Framework for Phishing Detection in Mobile Devices. In *15th International Conference on Computer and Information Technology, CIT; 14th International Conference on Ubiquitous Computing and Communications, IUCC; 13th International Conference on Dependable, Autonomic and Secure Computing, DASC; 13th International Conference on Pervasive Intelligence and Computing, PICom*, Yulei Wu, Geyong Min, Nektarios Georgalas, Jia Hu, Luigi Atzori, Xiaolong Jin, Stephen A. Jarvis, Lei (Chris) Liu, and Ramón Agüero Calvo (Eds.). IEEE, Liverpool, United Kingdom, 1977–1983. https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.293

[11] Sergey Brin and Lawrence Page. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks* 30, 1-7 (1998), 107–117. https://doi.org/10.1016/s0169-7552(98)00110-x

[12] Gamze Canova, Melanie Volkamer, Clemens Bergmann, and Benjamin Reinheimer. 2015. NoPhish App Evaluation: Lab and Retention Study. In *Internet Society, 8 February 2015 (Usec '15)*, Vol. 453. The Internet Society, San Diego, CA, USA, 1–10. http://dx.doi.org/10.14722/usec.2015.23009

[13] Sidharth Chhabra, Anupama Aggarwal, Fabrício Benevenuto, and Ponnurangam Kumaraguru. 2011. Phi.sh/$oCiaL: the phishing landscape through short URLs. In *The 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference, CEAS*. ACM, Perth, Australia, 92–101. https://doi.org/10.1145/2030376.2030387

[14] CMBuild. 2013. Archive of dmoz.org. (2013). https://dmoz-odp.org/Reference/ Accessed Dec. 2020.

[15] Lucian Constantin. 2019. Attackers Host Phishing Pages on Azure. (Mar. 2019). https://securityboulevard.com/2019/03/attackers-host-phishing-pages-on-azure/ Accessed Jun. 2019.

[16] Lorrie Faith Cranor. 2008. A Framework for Reasoning About the Human in the Loop. In *Usability, Psychology, and Security, UPSEC'08*, Elizabeth F. Churchill and Rachna Dhamija (Eds.). USENIX Association, San Francisco, CA, USA, 1–15. http://www.usenix.org/events/upsec08/tech/full%5Fpapers/cranor/cranor.pdf

[17] Rachna Dhamija, J. D. Tygar, and Marti A. Hearst. 2006. Why phishing works. In *Proceedings of the 2006 Conference on Human Factors in Computing Systems, CHI*, Rebecca E. Grinter, Tom Rodden, Paul M. Aoki, Edward Cutrell, Robin Jeffries, and Gary M. Olson (Eds.). ACM, Montréal, Québec, Canada, 581–590. https://doi.org/10.1145/1124772.1124861

[18] Hermann Ebbinghaus. 2013. Memory: a contribution to experimental psychology. *Annals of neurosciences* 20, 4 (Oct. 2013), 155–156. https://doi.org/10.5214/ans.0972.7531.200408

[19] Let's Encrypt. 2019. Free SSL/TLS Certificates. (2019). https://letsencrypt.org/ Accessed Dec. 2020.

[20] J Erkkila. 2011. Why we fall for phishing. In *Proceedings of the 2011 CHI Conference on Human Factors in Computing Systems (Chi '11)*. ACM, ancouver, BC, Canada, 1–8. https://juerkkil.iki.fi/files/writings/phishing

[21] FBI. 2020. *2019 Internet Crime Report, Data Reflects an Evolving Threat and the Importance of Reporting*. Technical Report. The Federal Bureau of Investigation, Internet Crime Complaint Center. https://www.fbi.gov/news/stories/2019-internet-crime-report-released-021120 Accessed Aug. 2020.

[22] Matheesha Fernando and Nalin Asanka Gamagedara Arachchilage. 2020. Why Johnny can't rely on anti-phishing educational interventions to protect himself against contemporary phishing attacks? *CoRR* abs/2004.13262 (2020), 1–12. arXiv:cs.CR/2004.13262 https://arxiv.org/abs/2004.13262

[23] Fortinet. 2021. Web Filter Categories. (Jan. 9 2021). https://www.fortiguard.com/webfilter/categories Accessed Aug. 2020.

[24] Lorenzo Franceschi-Bicchierai. 2016. How Hackers Broke Into John Podesta and Colin Powell's Gmail Accounts. (2016). https://motherboard.vice.com/en%5Fus/article/mg7xjb/how-hackers-broke-into-john-podesta-and-colin-powells-gmail-accounts Accessed Aug. 2020.

[25] Evgeniy Gabrilovich and Alex Gontmakher. 2002. The homograph attack. *Commun. ACM* 45, 2 (2002), 128. https://doi.org/10.1145/503124.503156

[26] Sujata Garera, Niels Provos, Monica Chew, and Aviel D. Rubin. 2007. A Framework for Detection and Measurement of Phishing Attacks. In *Proceedings of the 2007 ACM Workshop on Recurring Malcode (Worm '07)*. Association for Computing Machinery, New York, NY, USA, 1âĂŞ8. https://doi.org/10.1145/1314389.1314391

[27] Dan J. Graham, Jacob L. Orquin, and Vivianne H.M. Visschers. 2012. Eye tracking and nutrition label use: A review of the literature and recommendations for label enhancement. *Food Policy* 37, 4 (2012), 378–382. https://doi.org/10.1016/j.foodpol.2012.03.004

[28] Chris Grier, Kurt Thomas, Vern Paxson, and Chao Michael Zhang. 2010. spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM Conference on Computer and Communications Security, CCS 2010, October 4-8, 2010*. ACM, Chicago, Illinois, USA, 27–37. https://doi.org/10.1145/1866307.1866311

[29] Neha Gupta, Anupama Aggarwal, and Ponnurangam Kumaraguru. 2014. bit.ly/malicious: Deep dive into short URL based e-crime detection. In *APWG Symposium on Electronic Crime Research, eCrime*. IEEE, Birmingham, AL, USA, 14–24. https://doi.org/10.1109/ecrime.2014.6963161

[30] Srishti Gupta and Ponnurangam Kumaraguru. 2014. Emerging phishing trends and effectiveness of the anti-phishing landing page. In *2014 APWG Symposium on Electronic Crime Research, eCrime*. IEEE, Birmingham, AL, USA, 36–47. https://doi.org/10.1109/ecrime.2014.6963163

[31] Masayuki Higashino. 2019. A Design of an Anti-Phishing Training System Collaborated with Multiple Organizations. In *Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services, iiWAS 2019, December 2-4, 2019*. ACM, Munich, Germany, 589–592. https://doi.org/10.1145/3366030.3366086

[32] FBIâĂŹs Internet Crime Complaint Center (IC3). 2017. *2017 Internet Crime Report*. Technical Report. The Federal Bureau of Investigation (FBI), Internet Crime Complaint Center. https://pdf.ic3.gov/2017%5FIC3Report.pdf Accessed Aug. 2020.

[33] Iulia Ion, Rob Reeder, and Sunny Consolvo. 2015. "...No one Can Hack My Mind": Comparing Expert and Non-Expert Security Practices. In *Eleventh Symposium On Usable Privacy and Security, SOUPS*, Lorrie Faith Cranor, Robert Biddle, and Sunny Consolvo (Eds.). USENIX Association, Ottawa, Canada, 327–346. https://www.usenix.org/conference/soups2015/proceedings/presentation/ion

[34] Daniel Jampen, Gürkan Gür, Thomas Sutter, and Bernhard Tellenbach. 2020. Don't click: towards an effective anti-phishing training. A comparative literature review. *Human-centric Computing and Information Sciences* 10 (2020), 33. https://doi.org/10.1186/s13673-020-00237-7

[35] Bernhard Jenny and Nathaniel Vaughn Kelso. 2007. Color Design for the Color Vision Impaired. *Cartographic Perspectives* 58 (2007), 61–67. https://doi.org/10.14714/CP58.270

[36] Joseph Johnson. 2019. UK: number of internet users who are students 2011-2019. (May. 2019). https://www.statista.com/statistics/940040/number-of-student-internet-users-in-the-uk/

[37] Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. 2014. Privacy Attitudes of Mechanical Turk Workers and the U.S. Public. In *10th Symposium on USAble Privacy and Security, SOUPS*, Lorrie Faith Cranor, Lujo Bauer, and Robert Biddle (Eds.). USENIX Association, Menlo Park, CA, USA, 37–49. https://www.usenix.org/conference/soups2014/proceedings/presentation/kang

[38] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on USAble Privacy and Security, SOUPS*. ACM, Mountain View, California, USA, 1–a12. https://doi.org/10.1145/1572532.1572538

[39] Timothy Kelley and Bennett I. Bertenthal. 2016. Attention and past behavior, not security knowledge, modulate users' decisions to login to insecure websites. *Inf. Computer Security* 24, 2 (2016), 164–176. https://doi.org/10.1108/ics-01-2016-0002 arXiv:https://doi.org/10.1108/ICS-01-2016-0002

[40] Mahmoud Khonji, Youssef Iraqi, and Andrew Jones. 2013. Phishing Detection: A Literature Survey. *IEEE Communications Surveys Tutorials* 15, 4 (2013), 2091–2121. https://doi.org/10.1109/surv.2013.032213.00009

[41] Iacovos Kirlappos and Martina Angela Sasse. 2012. Security Education against Phishing: A Modest Proposal for a Major Rethink. *IEEE Security and Privacy* 10, 2 (2012), 24–32. https://doi.org/10.1109/MSP.2011.179

[42] Philipp Koehn, Huda Khayrallah, Kenneth Heafield, and Mikel L. Forcada. 2018. Findings of the WMT 2018 Shared Task on Parallel Corpus Filtering. In *Proceedings of the Third Conference on Machine Translation: Shared Task Papers, WMT 2018, October 31 - November 1, 2018*. Association for Computational Linguistics, Belgium, Brussels, 726–739. https://doi.org/10.18653/v1/w18-6453

[43] Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Theodore Pham. 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, Article 3, 12 pages. https://doi.org/10.1145/1572532.1572536

[44] Ponnurangam Kumaraguru, Yong Rhee, Alessandro Acquisti, Lorrie Faith Cranor, Jason I. Hong, and Elizabeth Nunge. 2007. Protecting people from phishing: the design and evaluation of an embedded training email system. In *Proceedings of the 2007 Conference on Human Factors in Computing Systems, CHI*, Mary Beth Rosson and David J. Gilmore (Eds.). ACM, San Jose, California, USA, 905–914. https://doi.org/10.1145/1240624.1240760

[45] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason I. Hong. 2010. Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology* 10, 2 (2010), 7:1–7:31. https://doi.org/10.1145/1754393.1754396

[46] Sangho Lee and Jong Kim. 2013. WarningBird: A Near Real-Time Detection System for Suspicious URLs in Twitter Stream. *IEEE Transactions on Dependable and Secure Computing* 10, 3 (2013), 183–195. https://doi.org/10.1109/tdsc.2013.3

[47] Chunlin Liu, Lidong Wang, Bo Lang, and Yuan Zhou. 2018. Finding Effective Classifier for Malicious URL Detection. In *Proceedings of the 2nd International Conference on Management Engineering, Software Engineering and Service Sciences (Icmss 2018)*. Association for Computing Machinery, New York, NY, USA, 240–244. https://doi.org/10.1145/3180374.3181352

[48] Netcraft Ltd. 2019. Internet Security and Data Mining. (2019). https://www.netcraft.com/ Accessed Jun. 2020.

[49] Justin Ma, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. 2009. Identifying suspicious URLs: an application of large-scale online learning. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML 2009, June 14-18, 2009 (ACM International Conference Proceeding Series)*, Andrea Pohoreckyj Danyluk, Léon Bottou, and Michael L. Littman (Eds.), Vol. 382. ACM, Montreal, Quebec, Canada, 681–688. https://doi.org/10.1145/1553374.1553462

[50] Samuel Marchal, Kalle Saari, Nidhi Singh, and N. Asokan. 2016. Know Your Phish: Novel Techniques for Detecting Phishing Sites and Their Targets. In *36th International Conference on Distributed Computing Systems, ICDCS*. IEEE, Nara, Japan, 323–333. https://doi.org/10.1109/icdcs.2016.10

[51] Ulrike Meyer and Vincent Drury. 2019. Certified Phishing: Taking a Look at Public Key Certificates of Phishing Websites. In *Fifteenth Symposium on USAble Privacy and Security, SOUPS*. USENIX Association, Santa Clara, CA, USA, 210–223. https://www.usenix.org/conference/soups2019/presentation/drury

[52] Microsoft. 2018. *Microsoft Security Intelligence Report, Volumne 23*. Technical Report. Microsoft. https://www.microsoft.com/en-us/security/intelligence-report Accessed Aug. 2018.

[53] Gaurav Misra, Nalin Asanka Gamagedara Arachchilage, and Shlomo Berkovsky. 2017. Phish Phinder: A Game Design Approach to Enhance User Confidence in Mitigating Phishing Attacks. In *Eleventh International Symposium on Human Aspects of Information Security & Assurance, HAISA, Proceedings*, Steven Furnell and Nathan L. Clarke (Eds.). University of Plymouth, Adelaide, Australia, 41–51. http://www.cscan.org/openaccess/?paperid=349

[54] Mattia Mossano, Kami Vaniea, Lukas Aldag, Reyhan Düzgün, Peter Mayer, and Melanie Volkamer. 2020. Analysis of publicly available anti-phishing webpages: contradicting information, lack of concrete advice and very narrow attack vector. In *European Symposium on Security and Privacy Workshops, EuroS&P Workshops*. IEEE, Genoa, Italy, 130–139. https://doi.org/10.1109/EuroSPW51379.2020.00026

[55] Rennie Naidoo. 2015. Analysing Urgency and Trust Cues Exploited in Phishing Scam Designs. In *10th International Conference on Cyber Warfare and Security, ICCWS*. Academic Conferences International Limited, The University of Venda and The Council for Scientific and Industrial Research, South Africa, 216–222. search.proquest.com/conference-papers-proceedings/analysing-urgency-trust-cues-exploited-phishing/docview/1781336050/se-2?accountid=10673

[56] James Nicholson, Lynne M. Coventry, and Pam Briggs. 2018. Introducing the Cybersurvival Task: Assessing and Addressing Staff Beliefs about Effective Cyber Protection. In *Fourteenth Symposium on USAble Privacy and Security, SOUPS, August 12-14, 2018*. USENIX Association, Baltimore, MD, USA, 443–457. https://www.usenix.org/conference/soups2018/presentation/nicholson

[57] Adam Oest, Yeganeh Safaei, Adam Doupé, Gail-Joon Ahn, Brad Wardman, and Gary Warner. 2018. Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis. In *2018 APWG Symposium on Electronic Crime Research, eCrime 2018, May 15-17, 2018*. IEEE, San Diego, CA, USA, 1–12. https://doi.org/10.1109/ecrime.2018.8376206

[58] LLC OpenDNS. 2019. PhishTank: Join the fight against phishing. (2019). https://www.phishtank.com/ Accessed Dec. 2020.

[59] OpenPhish. 2019. OpenPhish: Phishing Intelligence. (2019). https://openphish.com Accessed Dec. 2020.

[60] Charles A. O'Reilly. 1980. Individuals and Information Overload in Organizations: Is More Necessarily Better? *The Academy of Management Journal* 23, 4 (1980), 684–696. http://www.jstor.org/stable/255556

[61] Gilchan Park, Lauren M. Stuart, Julia M. Taylor, and Victor Raskin. 2014. Comparing machine and human ability to detect phishing emails. In *2014 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2014, October 5-8, 2014*. IEEE, San Diego, CA, USA, 2322–2327. https://doi.org/10.1109/smc.2014.6974273

[62] Cofense PhishMe. 2017. *Enterprise Phishing Resiliency and Defense Report*. Technical Report. PhishMe, Inc. https://cofense.com/wp-content/uploads/2017/11/Enterprise-Phishing-Resiliency-and-Defense-Report-2017.pdf Accessed Aug. 2020.

[63] Swapan Purkait. 2012. Phishing counter measures and their effectiveness - literature review. *Information Management & Computer Security* 20, 5 (2012), 382–420. https://doi.org/10.1108/09685221211286548

I Don't Need an Expert! Making URL Phishing Features Human Comprehensible

CHI '21, May 8–13, 2021, Yokohama, Japan

[64] Issa Qabajeh, Fadi A. Thabtah, and Francisco Chiclana. 2018. A recent review of conventional vs. automated cybersecurity anti-phishing techniques. *Computer Science Review* 29 (2018), 44–55. https://doi.org/10.1016/j.cosrev.2018.05.003

[65] Florian Quinkert, Tobias Lauinger, William K. Robertson, Engin Kirda, and Thorsten Holz. 2019. It's Not what It Looks Like: Measuring Attacks and Defensive Registrations of Homograph Domains. In *7th Conference on Communications and Network Security, CNS 2019, June 10-12, 2019*. IEEE, Washington, DC, USA, 259–267. https://doi.org/10.1109/cns.2019.8802671

[66] Elissa M. Redmiles, Amelia R. Malone, and Michelle L. Mazurek. 2016. I Think They're Trying to Tell Me Something: Advice Sources and Selection for Digital Security. In *IEEE Symposium on Security and Privacy, SP*. IEEE Computer Society, San Jose, CA, USA, 272–288. https://doi.org/10.1109/SP.2016.24

[67] Robert W. Reeder, Adrienne Porter Felt, Sunny Consolvo, Nathan Malkin, Christopher Thompson, and Serge Egelman. 2018. An Experience Sampling Study of User Reactions to Browser Warnings in the Field. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI*, Regan L. Mandryk, Mark Hancock, Mark Perry, and Anna L. Cox (Eds.). ACM, Montreal, QC, Canada, 512. https://doi.org/10.1145/3173574.3174086

[68] Robert W. Reeder, Iulia Ion, and Sunny Consolvo. 2017. 152 Simple Steps to Stay Safe Online: Security Advice for Non-Tech-Savvy Users. *IEEE Security & Privacy* 15, 5 (2017), 55–64. https://doi.org/10.1109/msp.2017.3681050

[69] Joshua Reynolds, Deepak KuMar., Zane Ma, Rohan Subramanian, Meishan Wu, Martin Shelton, Joshua Mason, Emily Stark, and Michael Bailey. 2020. Measuring Identity Confusion with Uniform Resource Locators. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, Honolulu, HI, USA, 1–12. https://doi.org/10.1145/3313831.3376298

[70] Doyen Sahoo, Chenghao Liu, and Steven C. H. Hoi. 2019. Malicious URL Detection using Machine Learning: A Survey. (2019). arXiv:cs.LG/1701.07179 http://arxiv.org/abs/1701.07179

[71] Maria Sameen, Kyunghyun Han, and Seong Oun Hwang. 2020. PhishHaven - An Efficient Real-Time AI Phishing URLs Detection System. *IEEE Access* 8 (2020), 83425–83443. https://doi.org/10.1109/ACCESS.2020.2991403

[72] Nuttapong Sanglerdsinlapachai and Arnon Rungsawang. 2010. Using Domain Top-page Similarity Feature in Machine Learning-Based Web Phishing Detection. In *Third International Conference on Knowledge Discovery and Data Mining, WKDD*. IEEE, Phuket, Thailand, 187–190. https://doi.org/10.1109/wkdd.2010.108

[73] Tara Seals. 2017. ost of user security training tops $290K per year. (2017). https://www.infosecurity-magazine.com/news/cost-of-user-security-training Accessed Nov. 2020.

[74] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason I. Hong, and Elizabeth Nunge. 2007. 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 2007, July 18-20, 2007 (ACM International Conference Proceeding Series)*, Lorrie Faith Cranor (Ed.), Vol. 229. ACM, Pittsburgh, Pennsylvania, USA, 88–99. https://doi.org/10.1145/1280680.1280692

[75] Hossein Siadati, Sean Palka, Avi Siegel, and Damon McCoy. 2017. Measuring the Effectiveness of Embedded Phishing Exercises. In *10th USENIX Workshop on Cyber Security Experimentation and Test, CSET 2017, August 14, 2017*. USENIX Association, Vancouver, BC, Canada, 8. https://www.usenix.org/conference/cset17/workshop-program/presentation/siadatii

[76] Gabor Szathmari. 2020. Why Outdated Anti-Phishing Advice Leaves You Exposed (Part 2). (Jul. 2020). https://blog.ironbastion.com.au/why-outdated-anti-phishing-advice-leaves-you-exposed-part-2/

[77] Janos Szurdi, Balazs Kocso, Gabor Cseh, Jonathan Spring, Márk Félegyházi, and Chris Kanich. 2014. The Long "Taile" of Typosquatting Domain Names. In *Proceedings of the 23rd USENIX Security Symposium*. USENIX Association, San Diego, CA, USA, 191–206. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/szurdi

[78] Rashid Tahir, Ali Raza, Faizan Ahmad, Jehangir Kazi, Fareed Zaffar, Chris Kanich, and Matthew Caesar. 2018. It's All in the Name: Why Some URLs are More Vulnerable to Typosquatting. In *Conference on Computer Communications, INFOCOM 2018, April 16-19, 2018*. IEEE, Honolulu, HI, USA, 2618–2626. https://doi.org/10.1109/infocom.2018.8486271

[79] Nikolaos Tsalis, Nikos Virvilis, Alexios Mylonas, Theodore K. Apostolopoulos, and Dimitris Gritzalis. 2014. Browser Blacklists: The Utopia of Phishing Protection. In *E-Business and Telecommunications - 11th International Joint Conference, ICETE, Revised Selected Papers (Communications in Computer and Information Science)*, Mohammad S. Obaidat, Andreas Holzinger, and Joaquim Filipe (Eds.), Vol. 554. Springer, Vienna, Austria, 278–293. https://doi.org/10.1007/978-3-319-25915-4_15

[80] Verizon. 2017. *2017 Data Breach Investigations Report*. Technical Report. Verizon. https://www.verizonenterprise.com/resources/reports/rp%5FDBIR%5F2018%5FReport%5Fexecsummary%5Fen%5Fxg.pdf Accessed Jun. 2018.

[81] Verizon. 2019. *2019 DataEnterprise Phishing Resiliency and Defense Repor Breach Investigations Report*. Technical Report. Verizon. https://enterprise.verizon.com/resources/reports/2019-data-breach-investigations-report.pdf Accessed Jun. 2020.

[82] Melanie Volkamer, Karen Renaud, Benjamin Reinheimer, and Alexandra Kunz. 2017. User experiences of TORPEDO: TOoltip-poweRed Phishing Email DetectiOn. *Computer Security* 71 (2017), 100–113. https://doi.org/10.1016/j.cose.2017.02.004

[83] Stephen Waddell. 2020. *CatchPhish: A URL and Anti-Phishing Research Platform*. Master's thesis. University of Edinburgh. https://groups.inf.ed.ac.uk/tulips/projects/19-20/waddell-2020.pdf

[84] Rick Wash. 2020. How Experts Detect Phishing Scam Emails. *Proc. ACM Human Computer Interaction* 4, CSCW2 (2020), 160:1–160:28. https://doi.org/10.1145/3415231

[85] Patrickson Weanquoi, Jaris Johnson, and Jinghua Zhang. 2017. Using a Game to Teach About Phishing. In *Proceedings of the 18th Annual Conference on Information Technology Education and the 6th Annual Conference on Research in Information Technology*, Stephen J. Zilora, Tom Ayers, and Daniel S. Bogaard (Eds.). ACM, Rochester, New York, USA, 75. https://doi.org/10.1145/3125659.3125669

[86] Emma J. Williams, Joanne Hinds, and Adam N. Joinson. 2018. Exploring susceptibility to phishing in the workplace. *International Journal of Human Computer Studies* 120 (2018), 1–13. https://doi.org/10.1016/j.ijhcs.2018.06.004

[87] Ryan T. Wright and Kent Marett. 2010. The Influence of Experiential and Dispositional Factors in Phishing: An Empirical Investigation of the Deceived. *Journal of Management Information Systems* 27, 1 (2010), 273–303. http://www.jmis-web.org/articles/1038

[88] Min Wu, Robert C. Miller, and Simson L. Garfinkel. 2006. Do security toolbars actually prevent phishing attacks?. In *Proceedings of the 2006 Conference on Human Factors in Computing Systems, CHI 2006, April 22-27, 2006*. ACM, Montréal, Québec, Canada, 601–610. https://doi.org/10.1145/1124772.1124863

[89] Guang Xiang, Jason I. Hong, Carolyn Penstein Rosé, and Lorrie Faith Cranor. 2011. CANTINA+: A Feature-Rich Machine Learning Framework for Detecting Phishing Web Sites. *ACM Trans. Inf. Syst. Secur.* 14, 2 (2011), 21:1–21:28. https://doi.org/10.1145/2019599.2019606

[90] Aiping Xiong, Robert W. Proctor, Weining Yang, and Ninghui Li. 2017. Is Domain Highlighting Actually Helpful in Identifying Phishing Web Pages? *Hum. Factors* 59, 4 (2017), 640–660. https://doi.org/10.1177/0018720816684064

[91] Jun Yang, Pengpeng Yang, Xiaohui Jin, and Qian Ma. 2017. Multi-Classification for Malicious URL Based on Improved Semi-Supervised Algorithm. In *IEEE International Conference on Computational Science and Engineering, CSE 2017, and IEEE International Conference on Embedded and Ubiquitous Computing, EUC, Volume 1*. IEEE Computer Society, Guangzhou, China, 143–150. https://doi.org/10.1109/CSE-EUC.2017.34