Assessing URL Literacy in an Enterprise Setting

Gokul Chettoor Jayakrishnan[1 [0000−0001−7505−7624]], Ajit Dhebe[1 [0009−0005−5203−2148]], Chinmay Mulay[1 [0009−0000−5031−1155]], Vijayanand Banahatti[1 [0000−0003−0321−498X]], and Sachin Lodha[1 [0000−0001−5771−4977]]

TCS Research, Tata Consultancy Services, Pune, India
{gokul.cj,ajit.dhebe,chinmay.mulay,
vijayanand.banahatti,sachin.lodha}@tcs.com

Abstract. Human-targeted cyber threats such as phishing often exploit users' limited understanding of URL structures when assessing website legitimacy. This study investigates URL literacy among enterprise users and evaluates the effectiveness of a lightweight email-based intervention for improving phishing resilience. A study involving 40 enterprise participants revealed only moderate baseline URL literacy, with challenges in identifying specific URL components. Following the intervention, participants demonstrated significant improvements in both URL component recognition and the ability to distinguish legitimate from malicious URLs, with gains retained in a delayed post-test. The study further examines the relationship between URL literacy and legitimacy judgement, suggesting that improved understanding of URL structures supports more accurate phishing detection. Finally, a detailed analysis of URL components and phishing attack patterns highlights areas where users remain susceptible to deception. These findings contribute insights into URL literacy, URL legitimacy judgement, and user susceptibility to different URL-based phishing attacks in enterprise environments.

Keywords: Security · URL literacy · URL legitimacy · Enterprise study · Phishing awareness.

1 Introduction

Phishing continues to pose a significant security risk, with attackers frequently using fake website links to lure users onto fraudulent sites. Proofpoint's Human Factor 2025 report notes that malicious URLs are used four times more often than attachments in phishing emails [34]. While anti-phishing tools with automated filters catch many threats, some phishing attacks still manage to slip through, highlighting the importance of user training as an extra layer of defense [3]. Research shows that attackers exploit the common difficulty users have in accurately reading or understanding URLs [13]. For example, one study by Albakry et al. [1] found that people often struggle to tell if a company name is in the subdomain or the main domain of a web address. A recent study by Sarno et al. [39] indicates that digital literacy and cognitive reflectiveness serve as significant predictors of individuals' ability to discern deception. Research

indicates that users frequently misunderstand URLs, often overestimating their ability to interpret them accurately [36], which presents opportunities for malicious actors. Attackers may manipulate various URL components, such as the domain, subdomain, or protocol, in ways that are not immediately apparent to users. As a result, individuals may inadvertently click on altered links, exposing themselves and their affiliated organizations to cyberattacks that can lead to significant financial and reputational damage. To counter deceptive tactics, building URL literacy (understanding URL components such as domains, subdomains, and protocols) is crucial for protecting users from phishing. This foundational skill also aids in phishing detection. Although research shows users often struggle to interpret URLs [3, 23], few interventions specifically teach URL components, with most existing methods focusing on interactive or just-in-time tools like URL inspection aids. This exploratory study investigates the enterprise participants' URL literacy and whether email nudges can enhance it. We address the following research questions:

- **RQ1**: What is the baseline level of URL literacy among enterprise participants?
- **RQ2**: Does an email-based nudge improve the users' URL literacy?
- **RQ3**: Does higher URL literacy lead to more accurate identification of phishing and legitimate URLs? (exploration)
- **RQ4**: Which URL components (domain, subdomain, TLD, and protocol) and attack types (typosquatting, combosquatting/subdomain spoofing, and protocol-related URLs) are most and least effectively recognised by users?

This paper introduces our initial method for improving URL literacy, summarizes quantitative analysis from 40 enterprise users, and shares key trends, findings, and recommendations. This paper also makes four key contributions: (1) assessing baseline URL literacy among users in an enterprise, (2) evaluating the effectiveness of an email-based nudge for improving URL literacy, (3) examining the relationship between URL literacy and URL legitimacy judgement, and (4) analysing user performance across URL components and phishing attack types.

2 Background and Related Work

2.1 Structure of a URL

Uniform Resource Locators (URLs) specify the structured details required to locate and access resources on the Internet [4]. URLs serve to identify resources by offering an abstract representation of their location. A URL can be represented in the following format:
`https://www.example.com/software/index.html?q=aBcDeF#section1`

The hypothetical URL presented above consists of the following components:

- Scheme or **Protocol**: `https`
  This element defines the protocol used to access resources online. Protocols include HTTP or HTTPS, with the latter providing Secure Sockets Layer (SSL) encryption.
- **Subdomain**: `www`
- **Domain name**: `example.com`
- **Top-level domain** (TLD): `.com`
  The website host is delineated by periods (".") into subdomain, domain, and TLD, formatted as `subdomain.domain.top-leveldomain`. An effective top-level domain (eTLD) is a public suffix recorded in the Public Suffix List. Examples include conventional TLDs such as com and multi-label public suffixes such as co.uk and github.io [26].
- **Path**: `/software/index.html`
  Indicates the directory location of the file (index.html).
- **Query string**: `q=aBcDeF`
  Specifies the search parameters used to assign values to particular keys, commonly represented as key-value pairs.
- **Fragment**: `#section1`
  Identifies a specific section within a webpage for linking and highlighting purposes. The information above is sourced from various references, including Web.dev, Berners-Lee (1994), Github, Sharma, P., & Nagpal, B. (2015) [4, 11, 41, 45].

2.2 Manipulation of URLs

Cyberattackers can exploit different parts of a URL to trick web servers into sending harmful or fake webpages to users. These attacks may lead to stolen data, unauthorized system access, the delivery of dangerous files, and theft of financial information [41]. Common deceptive tactics include positioning a brand name in misleading locations within a URL, such as the subdomain (e.g., google.phish.com), the path (e.g., phish.com/google), or the query string (e.g., phish.com?google). Frequently observed domain-related attacks consist of typosquatting (altering the spelling of domains (e.g., facebook to facedook)) and combosquatting, which involves combining a trusted brand or domain name with additional attacker-controlled words, strings, or identifiers within a URL or hostname (e.g., facebook to facebook-text) [22]. Typosquatting and combosquatting account for 53% and 30%, respectively, of squatting technique prevalence [22]. Attackers may change the top-level domain (TLD), such as using google.phish or inserting a fake TLD in a subdomain (e.g., google.com.phish.com) [23, 38]. Directory traversal attacks modify URL paths to access unauthorized site areas and can include redirections to fraudulent sites. For instance, consider the URL: `http://some_site.com.br/somepage?page=http://other-site.com.br/other-page.html/malicious-code.php`
In this example, the user is redirected to a "malicious code" page instead of the intended "some_site" [32]. With various URL manipulation techniques and

AI-driven strategies [43], phishing attacks are increasingly common in enterprises. Effective mitigation, warnings, and regular user training and awareness are critical to reduce susceptibility to such threats [37] .

2.3 Training and Mitigation

A significant challenge in phishing is that many users do not consistently check or accurately interpret browser address bars, URL cues, or use hover functions when deciding whom to trust. Dhamija et al. [10] found most participants ignored security indicators, leading to poor judgements. The study suggests developing alternative methods to help users understand URLs and distinguish legitimate from fraudulent security indicators. Several studies have focused on teaching users to assess URL legitimacy. Anti-phishing Phil [42] trains users to spot visual cues in URLs, while Phishy [6], a serious game, helps users recognise phishing URLs and has improved participants' understanding.

Recent research by Lain et al. [23] introduced micro-tasks to improve users' understanding of URLs by focusing on identifying key components, such as retyping domains. Lin et al. [24] found that highlighting URL domains can aid users but isn't sufficient alone. Similarly, Xiong et al. [46] reported that drawing attention to highlighted elements does not offer effective protection, as many users lack awareness of security cues, suggesting training may help. Albakry [1] found that most participants believed a URL containing a brand name would navigate to that organization's site. As highlighted in a recent study by Jabir et al. [20], the detection of malicious URLs encompasses techniques such as extracting lexical features [14, 33], domain-based features [25], and machine learning-based features capable of identifying rendered website elements by analyzing HTML DOM structures and images [44]. Nonetheless, phishing attacks utilizing URLs have evolved, enabling adversaries to evade numerous detection strategies. Contemporary research indicates that attackers increasingly exploit compromised and legitimate websites, meticulously crafting domain names to circumvent URL detection mechanisms [9, 31].

These findings highlight the need to educate users about URL components and how attackers exploit them. The recent study by Sarno et al. [39] also recommends training to enhance digital literacy as a safeguard against online deception. Given that the majority of phishing attacks exploit vulnerabilities through manipulated URLs, it is essential for research to examine whether training users solely on identifying phishing elements is sufficient. To our knowledge, there are no comprehensive studies assessing enterprise users' proficiency with URL components or their understanding of general URL structures. This study aims to evaluate the necessity for improving URL literacy, as it represents a foundational step in empowering users to recognize suspicious components within a URL.

3 Our Approach

Our methodology assessed whether enterprise participants understand the components of a URL, as outlined in Section 2.1. To help with this, we provided

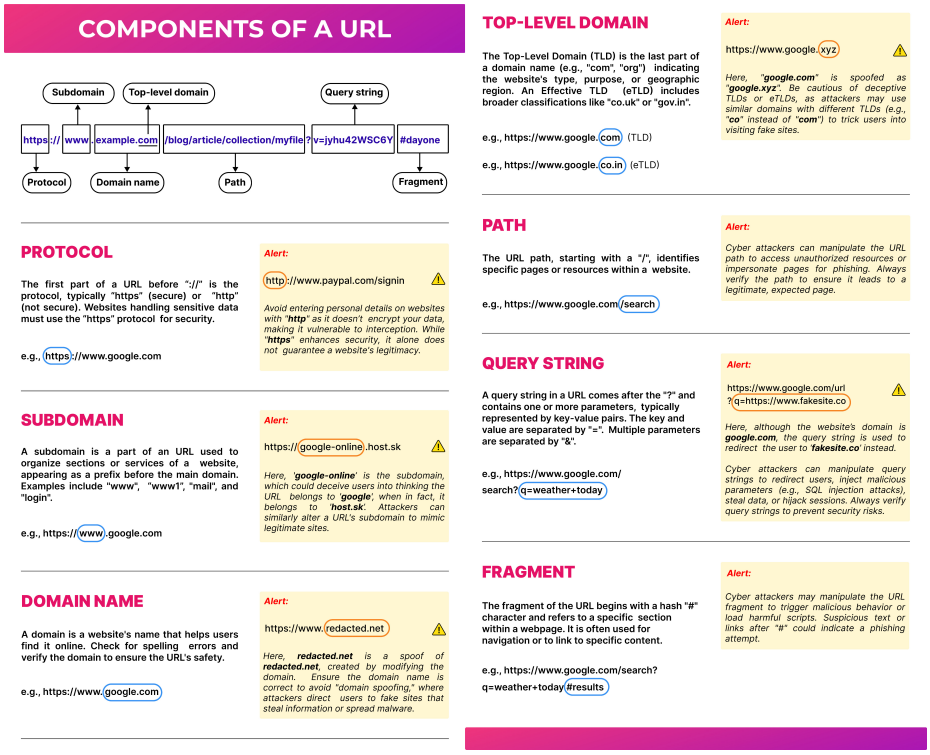

Fig. 1. The Infographic Nudge sent over email.

an infographic summarizing the seven main parts of a URL: protocol, domain, subdomain, TLD, path, query string, and fragment. The next section reviews this infographic in detail.

3.1 Cybersecurity Bytes: The Nudge

We used a brief security information nudge by embedding an infographic in an email sent to participants through their work addresses. The document, titled "Cybersecurity Bytes: Components of a URL," was created to visually explain the different parts of a URL. It highlights important signs to spot each component and gives brief tips on checking whether a URL is genuine or potentially risky. The material was distributed using company email, purposefully designed to match the organization's regular practice of sharing educational content in this way, maintaining consistency with usual communication styles. Fig. 1 displays the infographic nudge sent over email to the participants. It was challenging to find consistent information on URL components, so we consulted several credible sources and created a clear list for easy reference. The list comprises the following

URL components: protocol, subdomain, domain name, top-level domain, path, query string, and fragment. The content pertaining to these components was developed using several reputable sources, including Mozilla documentation [29], IBM documentation [18], GeeksforGeeks documentation [15], HubSpot blog [17], Google documentation [16], ICANN documentation [19], Mozilla resources on TLDs and fragments [27, 28], Claravine blog [7], and the public suffix list by Mozilla [26]. The synthesized information from these references was utilized to develop comprehensive learning material suitable for email distribution.

4 Study

Our study was exploratory in nature and involved a sample of enterprise participants. We assessed participants' URL literacy using a within-subject design. Participants completed a pre-test, received an information nudge via email, and subsequently participated in post-test and delayed post-test evaluations, with results compared across these stages. As recognition for their involvement, participants were awarded points from the organization's in-house rewards system, which could be exchanged for purchase vouchers.

Study Participants: We performed an enterprise case study with $n = 40$ participants (Gender: Male: 70%, Female: 30%; Age group: 18-30: 47.5%, 31-40:35%, 41-50:17.5%; Educational background: Bachelor's: 45%, Master's: 52.5%, Doctorate: 2.5%) who were employees belonging to an information technology company. Participants were recruited via email communication. Employees who expressed interest and were available for the duration of the study were selected regardless of their location or job function.

Ethics Process: All participants gave their consent before the study began by completing an online "Information and Consent Form" approved by the organization's Ethics Committee to ensure compliance with ethical standards. Participation was entirely voluntary. After giving consent, participants registered for the study and chose to provide their demographic details.

4.1 Study Procedure

In the study, we referred to the pre-test, post-test, and delayed post-test as Survey A, Survey B and Survey C respectively for participants. Each survey included 15 questions, each featuring one URL.

- **Questions 1 to 7**: These questions assessed participants' understanding of the URL components, explicitly mentioned in the nudge. Example: "Identify the domain name of this URL": `https://www.linkedin.com/linkedin/answer/[redacted]/linkedin\-groups-member-join?lang=en`
  Options: (1) www, (2) https, (3) linkedin.com, (4) com, (5) I don't know.

– **Questions 8 to 15**: As an exploration, we included the task of URL classification in all tests, where participants judged whether URLs were legitimate or suspicious. Example: "Based on your understanding, answer the question about the following URL": `https://www.amazon.in/amazonprime?ref_=nav_cs_primelink_non\member`
Options: This URL is (1) Legitimate, (2) Suspicious.

The study had the following experimental steps:

– **Pre-test**: Participants completed Survey A at the study's outset (see Appendix).
– **Nudge**: Two days later, participants received an informational nudge by email (Fig. 1).
– **Post-test**: After another two days, they took Survey B (see Appendix).
– **Delayed post-test**: Two weeks after the post-test, retention was measured with Survey C (see Appendix), following standard research practices [35].

To minimize practice effects, a two-day gap was kept before and after training. All surveys were created in Microsoft Forms and administered via one-on-one Microsoft Teams calls (voice only). The participants were instructed to complete them without consulting external materials. A fixed script highlighted the need for honest responses and was used for each session. Participants were not instructed to share their screens, to avoid any additional pressure. Participants were recruited and managed by three researchers across different locations. As the participants did not interact with one another during the study, the risk of information sharing or discussion of study materials was considered negligible.

4.2 Selection of URLs for the Study

URL Selection and Categorisation: A total of 45 URLs (15 per assessment) were used across the pre-, post-, and delayed post-tests. Seven questions evaluated participants' ability to identify URL components (domain, subdomain, protocol, TLD, fragment, path, and query string), while the remaining eight assessed URL legitimacy, with two questions each focusing on domain, subdomain, TLD, and protocol cues. An equal number of legitimate and suspicious URLs were included in each assessment. The suspicious URLs represented three attack categories: typosquatting, combosquatting/subdomain spoofing, and unsafe protocol usage. All URLs were selected to ensure comparable difficulty (based on their structure and the types of services they represent) and contextual relevance (sourced based on reputable institutions within the country).

URL Source and Distribution: The most frequently accessed websites were identified using the Semrush website ranking list [40] across various categories, specifically banking, insurance, e-commerce, and online services (such as social media and general web pages, e.g., Wikipedia and Google). These categories and URLs were chosen because they represent the websites users most frequently interact with daily, making them realistic targets for attacker manipulation and

relevant for assessing legitimacy judgement in practical, high-exposure contexts. The questions were evenly distributed between the pre-test and post-tests (including delayed post-test), with each containing an equal number of items from each category. Additionally, the lengths of the URLs within similar categories were standardised to minimise potential biases. The URLs were also balanced by component, legitimacy label, and attack type to maintain similar difficulty levels.

Suspicious URL Inclusion: Suspicious URLs included in the survey were obtained from the PhishTank database [5] (while following contextual relevance and familiarity; e.g., widely used within the country). We selected those URLs that have slight modifications to the domain, subdomain, protocol, and TLDs from the original URLs. To ensure participant safety, only phishing URLs previously confirmed as inactive in the database were selected, and these were displayed to participants in image format to prevent accidental clicks and copy-pasting. For suspicious URLs where HTTP was the primary identifying cue, the linked website involved user data entry, which generally requires encryption. HTTP usage was therefore treated as an indicator of a suspicious URL rather than direct evidence of phishing.

4.3 Data Collection

We collected survey responses from participants, along with their voluntarily provided demographic details such as gender, age group, and educational background. To address RQ1 and RQ2, we assessed the accuracy of answers for the first seven questions on identifying URL components. For answering our exploratory RQ3, we evaluated the correctness of responses to the subsequent eight questions. To evaluate RQ4, we measured responses of questions belonging to certain attack types and URL components and measured the improvement in post-test and retention in delayed test.

5 Results

5.1 RQ1: On Baseline URL Literacy of Participants

When we evaluated how accurately participants answered questions on identifying each of the seven URL components, we found that their average score in the pre-test was only 64.6% (with a standard deviation (SD) of 21.9%). This relatively low result for an enterprise audience indicates there is room for improvement in this area. The average pre-test correctness was 68.4% for male participants and 56% for female participants. The 18-30 and 31-40 age groups had about 63% correctness, while the 41-50 group reached 71.4%.

5.2 RQ2: URL Literacy Improvement Using a Nudge

Following the email-based nudge and participants' engagement with the learning content, their mean post-test correctness rose to 82.1% ($SD = 23.0\%$), further

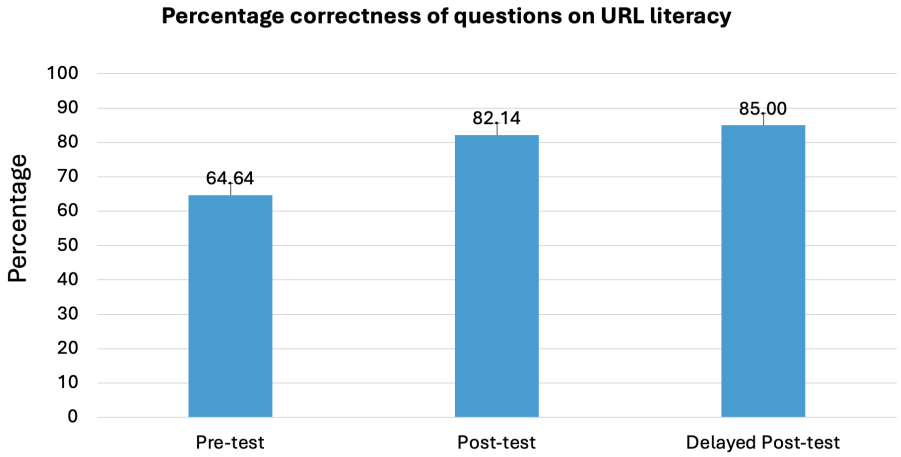

Fig. 2. Mean correctness of questions on identifying URL components.

increasing to 85% ($SD = 26.3\%$) on the delayed post-test (Fig. 2). Results from a paired-samples t-test indicated that post-test scores were significantly higher than pre-test scores, $t(39) = 5.0, p < .001$, with a large effect size (Cohen's $d = 0.8$).

We observed that the accuracy on URL component-based questions improved following the informational nudge. All demographic groups improved by 26–28%, and hence there were not much difference between the different genders and age groups. The component-wise analysis revealed substantial improvements in the recognition of several URL components following the intervention (Table 1). The largest gains were observed for TLDs and fragments (+0.35 each), followed by subdomains (+0.30), indicating that participants initially struggled with these components but showed considerable learning after the email nudge. Recognition of query strings also improved (+0.20), while protocols showed a modest gain (+0.08). In contrast, domains and paths exhibited little to no immediate improvement, likely due to their relatively high baseline scores. The delayed post-test further showed sustained or even enhanced performance for subdomains, TLDs, and domains, as reflected by negative retention values (i.e., delayed scores exceeding post-test scores). Overall, the findings suggest that the intervention was particularly effective in improving participants' understanding of less familiar URL components, while knowledge of frequently encountered components was already relatively strong at baseline.

5.3 RQ3: Exploration: URL Literacy and its Effect on URL Legitimacy Judgement

As an exploratory research question, we also examined whether greater URL literacy leads to better accuracy in judging URL legitimacy. Our initial findings

Table 1. Mean correctness of Pre-test, post-test, delayed post-test with learning gain and retention loss across URL components. Here, "RL" denotes retention loss; positive values indicate forgetting, while negative values indicate continued improvement beyond the post-test. All values are rounded to two decimal places for readability.

URL component	**Pre** (x)	**Post** (y)	**Delayed** (z)	**Gain** (y - x)	**RL** (y - z)
Domain	0.83	0.78	0.85	**-0.05**	**-0.08**
Protocol	0.88	0.95	0.90	**+0.08**	**+0.05**
Subdomain	0.33	0.63	0.75	**+0.30**	**-0.13**
TLD	0.35	0.70	0.85	**+0.35**	**-0.15**
Query string	0.75	0.95	0.93	**+0.20**	**+0.02**
Path	0.95	0.95	0.93	**0.00**	**+0.02**
Fragment	0.45	0.80	0.75	**+0.35**	**+0.05**

showed an increase in average correctness for answers about URL legitimacy from pre-test to post-test. Results from a Wilcoxon signed-rank test revealed that post-test scores ($M = 82.5\%, SD = 18.5\%$) were significantly higher than pre-test scores ($M = 77.2\%, SD = 14.1\%$), with a test statistic of $z = 2.1$ and $p < 0.05$. The effect size was moderate ($r = 0.38$). Furthermore, the common language effect size suggested there was a 71% chance that a randomly chosen post-test score would be higher than a randomly chosen pre-test score. Upon analysis, no statistically significant correlation was observed between pre-test scores for URL legitimacy questions and URL component questions ($r(38) = .3, p = .060$). Conversely, post-test results demonstrated that the Pearson correlation revealed a significant moderate positive relationship between these scores ($r(38) = .451, p = .004$).

5.4 RQ4: Attack Pattern Analysis

We compared the participants' overall mean correctness in answering questions related to each of these attack patterns (for suspicious URLs, part of the eight questions on URL legitimacy), across the three tests. Table 2 shows that for typosquatting and combosquatting (with subdomain spoofing included) attack patterns, the post-test showed increase in values as compared to pre-test and then showed a slight reduction in the delayed test. However, for protocol, the trend was declining from pre-test to delayed test. Combosquatting-related questions showed a 12.7% increase (+0.10) from the pre-test to post-test, whereas for typosquatting, it was 11.4% (+0.08) increase. Protocol showed -6.45% (-0.05) as the percentage change. We identified the best retained attack type using these data by computing

$$L_i = C_{i,\text{post}} - C_{i,\text{delayed}} \tag{1}$$

where L_i denotes the retention loss for attack category i, $C_{i,\text{post}}$ denotes the mean post-test correctness score, and $C_{i,\text{delayed}}$ denotes the mean delayed post-test correctness score. Typosquatting showed 0.16 (21.8%) retention loss, whereas combosquatting showed 0.04 (4.4%) and protocol-related questions showed 0.13 (17.8%).

Table 2. Mean correctness of Pre-test, post-test, delayed post-test with learning gain and retention loss across URL attack types. Here, "RL" denotes retention loss; positive values indicate forgetting, while negative values indicate continued improvement beyond the post-test. All values are rounded to two decimal places for readability.

Attack Type	**Pre (x)**	**Post (y)**	**Delayed (z)**	**Gain (y - x)**	**RL (y - z)**
Typosquatting	0.70	0.78	0.61	**+0.08**	**+0.16**
Combosquatting	0.79	0.89	0.85	**+0.10**	**+0.04**
Protocol-related URLs	0.78	0.73	0.60	**-0.05**	**+0.13**

Table 3. Mean correctness of Pre-test, post-test, delayed post-test with learning gain and retention loss across URL Legitimacy questions that had focus on specific URL components. Here, "RL" denotes retention loss. All values are rounded to two decimal places for readability.

Component	**Pre (x)**	**Post (y)**	**Delayed (z)**	**Gain (y - x)**	**RL (y - z)**
Domain	0.76	0.83	0.65	**+0.06**	**+0.18**
Subdomain	0.59	0.79	0.79	**+0.20**	**0.00**
Protocol	0.86	0.78	0.70	**-0.09**	**+0.08**
TLD	0.88	0.91	0.89	**+0.04**	**+0.03**

5.5 RQ4: URL Legitimacy Analysis based on Components

Table 3 presents the mean correctness scores for URL legitimacy questions (eight questions, two each for domain, subdomain, protocol and TLD) based on component identification across the three study phases. At baseline, participants demonstrated the highest correctness for identifying URL legitimacy questions based on TLD (0.88) and protocol (0.86), while subdomain questions had the lowest correctness (0.59). Following the intervention, the largest improvement was observed for questions where subdomain identification was required (gain = 0.20), and this improvement was fully retained in the delayed post-test (loss = 0.00). Domain identification also improved (gain = 0.06), although some decline was observed over time (loss = 0.18). TLD-related questions showed consistently high mean correctness across all phases, with only a modest gain (0.04) and minimal loss (0.03), suggesting strong pre-existing knowledge. In contrast, URL legitimacy questions with protocol identification exhibited a decline in performance (gain = -0.09) and a retention loss of 0.08, indicating that participants may have either overlooked protocols during URL legitimacy judgement or found it more challenging. Overall, the results suggest that subdomain concepts benefited most from the intervention, whereas protocol-related concepts remained difficult despite participants' relatively high baseline familiarity.

5.6 Other Findings

To further characterise participants' URL legitimacy judgements, we calculated the False Positive Rate (FPR) and False Negative Rate (FNR) for each participant using Equations 2 and 3.

$$FPR = \frac{FP}{FP + TN} \tag{2}$$

$$FNR = \frac{FN}{FN + TP} \tag{3}$$

where FP, TN, FN, and TP denote false positives, true negatives, false negatives, and true positives, respectively. The mean FPR decreased from 0.21 (Var = 0.03) in the pre-test to 0.16 (Var = 0.04) in the post-test. Similarly, the mean FNR decreased from 0.23 (Var = 0.07) to 0.18 (Var = 0.05).

To assess participant-level phishing susceptibility, we adapted the industry-standard Phish-Prone Percentage (PPP) metric [8, 21]. While PPP is traditionally computed using responses to phishing simulations, in this study it was defined as the percentage of participants who incorrectly classified at least one phishing URL as legitimate (using Equation 4).

$$PPP = \frac{N_{phish\text{-}prone}}{N_{participants}} \times 100 \tag{4}$$

The baseline PPP was 55.0%, indicating that more than half of the participants fell for at least one phishing URL prior to the intervention. Following the email nudge, PPP decreased to 45.0%, suggesting an immediate reduction in phishing susceptibility. However, the delayed post-test PPP increased to 77.5%, indicating that a substantial proportion of participants still misclassified at least one phishing URL over time. This apparent increase should be interpreted cautiously, as PPP is a stringent metric: a participant is considered phish-prone even if only a single phishing URL is misclassified. Nevertheless, the results suggest that while overall URL literacy and legitimacy-judgement accuracy improved, susceptibility to at least one phishing scenario remained prevalent, highlighting the need for sustained and repeated training interventions.

5.7 Open-ended feedback

We asked the participants for their open-ended feedback regarding the study and their experience with the URLs and the nudge and 30 participants responded. We got mixed reviews from the participants. Participants P15 and P34 said that the content was brief and good for a layman, and would prefer such informative content over email, P18 said that everyone should have basic knowledge regarding the URLs, P32 said that the examples within the nudge made it easier to understand, P27 said that they read about two to three times and absorbed the information very well. A few participants pointed out that it is very easy to "miss the email" and P11 recommended using "catchy titles" within email (email-based nudge). One participant P25 also requested for "phishing vs. legitimate URLs based on fragment and query string" to be included in the nudge. P39 requested for more citations and references regarding the information nudge so that they can learn more about the content.

6 Discussions

In this section, we discuss our findings in relation to the four research questions (RQ1, RQ2, RQ3 and RQ4).

- **RQ1: URL Literacy of Enterprise Participants**: Our initial assessment (RQ1) revealed that enterprise participants possessed only moderate URL literacy, with an average correctness of 64.6% when identifying the seven URL components, with subdomains being identified the least. This aligns with previous research from Alsharnouby et al. [2], which also found limited understanding of subdomains among users. Furthermore, participants with higher age groups showed slightly higher mean correctness.

- **RQ2: URL Literacy Improvement**: Addressing RQ2, we observed a significant improvement in URL literacy following the email-based intervention, with mean correctness increasing from 64.6% to 82.1% and further to 85.0% in the delayed post-test. No significant differences were observed across age groups, consistent with findings from Sarno et al. [39]. The component-wise analysis showed the largest gains for subdomains, TLDs, and fragments, while domain and path recognition remained relatively stable due to high baseline performance. Delayed post-test results further indicated strong retention, particularly for domains, subdomains, and TLDs, suggesting that learning persisted beyond the immediate intervention.

- **RQ3: URL Literacy and Legitimacy Judgement**: For RQ3, we explored the relationship between URL literacy and the ability to judge URL legitimacy. Our results indicated a significant improvement in legitimacy judgement scores after the intervention, rising from 77.2% to 82.5%. While there was no significant correlation between pre-test scores for URL component identification and legitimacy judgement, the post-test results showed a moderate positive correlation ($r = .451, p = .004$), suggesting that the intervention helped participants connect their knowledge of URL components with their ability to assess legitimacy. Additionally, both the FPR and FNR decreased, indicating enhanced discernment after the nudge.

- **RQ4**:
  URL Legitimacy based on attack patterns: The attack-wise analysis revealed differences in users' ability to identify and retain knowledge of URL-based attacks. Combosquatting/subdomain spoofing showed the highest learning gain (+0.10) and lowest retention loss (0.04), whereas typosquatting exhibited smaller gains (+0.08) and greater forgetting (0.16). Protocol-related URLs remained the most challenging category, showing a negative learning gain (-0.05) and a retention loss of 0.13. One possible explanation is that participants focused more on prominent URL features, such as brand names and domains, while overlooking protocol information (HTTP/HTTPS). Overall, the findings suggest that different attack types impose different cognitive demands on users and may require varying levels of training emphasis.

URL Legitimacy based on Components: The component-wise analysis showed subdomain judgements had the largest learning gains and retention, while domain judgements showed modest gains and greater retention loss. TLD performance was high at baseline The protocol-based URL judgements remained challenging and declined over time. These findings suggest that some URL components are more readily learned and retained than others. Since component recognition alone does not guarantee correct legitimacy judgement, future work should examine how contextual cues (e.g., sender, page content) combine with URL literacy to improve accuracy of judgements.

Participants retained over 75% of URL literacy and legitimacy knowledge after two weeks without reinforcement. According to the Ebbinghaus forgetting curve [12,30], memory typically drops off quickly at first and then declines more slowly without continued reinforcement. Our findings aligned with theoretical expectations and also indicated a reduced early memory loss. However, the rise in PPP at the delayed test (55%→45%→77.5%) despite improved literacy suggests component-recognition gains do not fully translate into durable phishing resistance, motivating our proposed spaced-repetition follow-up.

6.1 Suggestions Based on Our Study

- URL literacy training should focus on understanding URL structure, as participants demonstrated only moderate baseline knowledge of URL components.
- Greater emphasis should be placed on subdomains, TLDs, and fragments, which showed the largest learning gains and therefore represent important learning opportunities.
- Improvements in URL literacy were accompanied by improvements in URL legitimacy judgement, suggesting that a better understanding of URL components may support phishing detection.
- Training should explicitly address typosquatting, combosquatting/subdomain spoofing, and protocol-tampering, as susceptibility varied across attack types.
- Reductions in FPR, FNR, and post-test PPP indicate that brief interventions can improve users' ability to distinguish legitimate and phishing URLs.
- The increase in delayed-test PPP suggests that one-time interventions may be insufficient and that periodic reinforcement or spaced learning may be needed.
- Future work should investigate more advanced URL deception techniques, such as homoglyph attacks, shortened URLs, and AI-generated phishing URLs, and evaluate personalised training approaches.

6.2 Limitations and Future Work

This exploratory study should be interpreted in light of a few limitations. First, the findings are based on a relatively small convenience sample ($n = 40$) from a single enterprise, which may limit the generalisability of the results to other

organizational and demographic contexts. Moreover, the IT-centric nature of the participant population may have influenced both baseline URL literacy and responsiveness to the intervention. Future work should replicate across diverse organizations and job profiles, capturing technical background and prior phishing-awareness training. The different URL sets used across tests were balanced by component, legitimacy label, lengths, type of service, and attack type; however, the item difficulty was not independently validated. The absence of a control group limits attribution of improvements solely to the intervention rather than practice effects; future work should include one to confirm causal impact. Participant feedback also indicated that emails during work hours may be overlooked, highlighting the need to consider delivery timing. Future work should validate findings with larger, more diverse samples, assess long-term retention, and explore personalised training for challenging components such as subdomains, TLDs, typosquatting, and protocol cues.

7 Conclusions

This study underscores URL literacy as a human-centred defence against phishing in enterprise settings. Despite only moderate baseline literacy, a lightweight email-based intervention significantly improved URL-component recognition and legitimacy judgement, reducing false positive/negative rates and enhancing discernment. Although the literacy-judgement relationship was correlational, a better URL understanding was linked to a more accurate legitimacy assessment. Improvements were component- and attack-dependent, with greatest gains for subdomains, TLDs, and fragments, while typosquatting and protocol-based URLs remained challenging. These findings suggest that future awareness programmes should place greater emphasis on the URL components and attack patterns that users find most difficult to recognise. Notably, phish-prone percentage rose at the delayed test despite improved URL literacy, indicating that component-recognition gains alone do not ensure durable phishing resistance. As findings stem from a single enterprise, future work should validate results across diverse populations and explore personalised, spaced-learning approaches for lasting URL literacy and phishing resilience.

Appendix

The following subsections present the pre-test, post-test, and delayed post-test questions used to assess URL component identification and URL legitimacy. Portions of some URLs have been redacted to preserve anonymity because they contain region-specific or organisation-specific information.

A.1 Pre-test Questionnaire

Type 1 Questions: Identify "Component" of this URL

1. Identify "domain name" of this URL:
   `https://www.adobe.com/converter/in/acrobat/online/jpgtopdf.html?promoid=[redacted]=other#`
2. Identify "protocol" of this URL:
   `ftp://ftp.microsoft.com/outlook/auth/logon.aspx?replaceCurrent=1&url=https%3a%2f%2fmail.[redacted].com%2fowa%2f%40%2f`
3. Identify "subdomain" of this URL:
   `http://www2.nationalgeographic.com/collaboration/join/marquee/event/register`
4. Identify "top-level domain" (TLD/eTLD) of this URL:
   `https://www.bbc.co.uk/news/world-corporate-[redacted]-sustainathon/58339646`
5. Identify "query string" of this URL:
   `https://www.amazon.in/my-store/online/gift-card/grabcard/my-gift-card?ie=ref=[redacted]&node=[redacted]`
6. Identify "path" of this URL: `https://www.researchgate.net/profile/Zhiqiu-Lin/publication/332745435_Engaging_Through_a_Role-playing_Simulation_software.pdf`
7. Identify "fragment" of this URL:
   `https://[redacted].[redacted].net/enterprise/how-to-update-password/smart-drive/1786?tab=cause#day30`

Type 2 Questions: Identify if this URL is legitimate or suspicious

8. `https://www.indiamart.com/proddetail/dji-mavic-3e-worry-free-basic-[redacted].html` (Legitimate)
9. `https://en.wikipedia.org/wiki/Investment_banking` (Legitimate)
10. `http://ebiz.licindia.in/D2CPM/#Login` (Suspicious)
11. `https://retail.onlinesbi.sbi/retail/.html` (Legitimate)
12. `https://www.facebook.com.esy.pk/index.html` (Suspicious)
13. `https://amazon-online.webs.com/login.html` (Suspicious)
14. `https://www.maxlifeinsurance.com/cs/login` (Legitimate)
15. `https://www.icicibak.com/safe-online-banking/safe-online-banking.page` (Suspicious)

A.2 Post-test Questionnaire

Type 1 Questions: Identify "Component" of this URL

1. Identify "domain name" of this URL:
   `https://www.linkedin.com/linkedin/answer/[redacted]/linkedin-groups-member-join?lang=en`
2. Identify "protocol" of this URL:
   `https://mail.outlook.com:443/personal/[redacted]/_?id=[redacted]file&ga=1#sharepoint`

3. Identify "subdomain" of this URL: `https://developers.facebook.com/article/developers/hackaton?gotosite=1&url=https3a#eventatmeta`
4. Identify "top-level domain" (TLD/eTLD) of this URL: `https://www.online.citibank.in`
5. Identify "query string" of this URL: `https://www.w3schools.com/html/tryit.asp?filename=tryhtml_default_default`
6. Identify "path" of this URL: `https://www.makemytrip.com/holidays-india/char_dham-travel-packages.html`
7. Identify "fragment" of this URL: `https://developer.mozilla.org/en-US/docs/Web/HTML/Element/a#compatibility`

Type 2 Questions: On URL Legitimacy

8. `https://netbanking.hdfcbank.com/netbanking/` (Legitimate)
9. `https://linkedin.ctoliifso.repli.xyz/` (Suspicious)
10. `https://www.google.co.in/search?q=investment+banking` (Legitimate)
11. `https://www.myntra.com/shop/women?utm_source=dms_google` (Legitimate)
12. `https://www.axissbank.com/make-payments/pay-bills/postpaid-mobile-bill-payment` (Suspicious)
13. `https://customer.starhealth.in/sso/login` (Legitimate)
14. `http://www.acko.com/authn/v1/signin` (Suspicious)
15. `https://flipkart-online.yolasite.com/login.html` (Suspicious)

A.3 Delayed Post-test Questionnaire

Type 1 Questions: Identify "Component" of this URL

1. Identify "domain name" of this URL: `https://www.youtube.com/feed/trending?bp=4gIcGhpnYW1pbmdf[redacted]`
2. Identify "protocol" of this URL: `https://play.google.com/store/apps/details?id=cris.org.in.prs.ima&hl=en&pli=1`
3. Identify "subdomain" of this URL: `https://apps.apple.com/in/app/[redacted]-bank-mobilebanking/id515891771`
4. Identify "top-level domain" (TLD/eTLD) of this URL: `https://www.irctc.co.in/nget/enquiry/PNR/PnrEnquiry.html?locale=en`
5. Identify "query string" of this URL: `https://www.news18.com/news/india-news-today-1234567.html?src=website&medium=campaign=news`

6. Identify "path" of this URL:
`https://www.reddit.com/r/india/comments/abc123/title_of_the_post`
7. Identify "fragment" of this URL:
`https://in.yahoo.com/tech-updates-20-feb-2025-1234567.html?src=referral&campaign=news#latest-updates`

Type 2 Questions: On URL Legitimacy

8. `https://www.snapdeal.com/products/fashion-women?sort=plrty#bcrumbLabelId`(Legitimate)
9. `https://Itwitter.tk/login.html`(Suspicious)
10. `https://ebay.latestock.worker.dev/` (Suspicious)
11. `https://customer.kotaklifeinsurance.com/kliportal`(Legitimate)
12. `http://myaccount.hdfclife.com/login` (Suspicious)
13. `https://application.axisbank.co.[redacted]/webforms/[redacted]/index.aspx?utm_source=ws&utm=product=lap` (Legitimate)
14. `https://www.amazon.in/amazonprime?ref_=nav_cs_primelink_nonmember` (Legitimate)
15. `https://www.standardchartred.com/en/banking-services/personal-banking/` (Suspicious)

References

1. Albakry, S., Vaniea, K., Wolters, M.K.: What is this url's destination? empirical evaluation of users' url reading. In: Proceedings of the 2020 CHI conference on human factors in computing systems. pp. 1–12 (2020)
2. Alsharnouby, M., Alaca, F., Chiasson, S.: Why phishing still works: User strategies for combating phishing attacks. International Journal of Human-Computer Studies **82**, 69–82 (2015)
3. Althobaiti, K., Vaniea, K., Zheng, S.: Faheem: Explaining urls to people using a slack bot. In: AISB 2018 Symposium on Swarm Intelligence & Evolutionary Computation. pp. 1–8 (2018)
4. Berners-Lee, T., Masinter, L., McCahill, M.: Uniform resource locators (url). Tech. rep. (1994)
5. Cisco Talos Intelligence Group: Phishtank (2026), `https://phishtank.org/`, retrieved Apr. 20, 2026
6. CJ, G., Pandit, S., Vaddepalli, S., Tupsamudre, H., Banahatti, V., Lodha, S.: Phishy-a serious game to train enterprise users on phishing awareness. In: Proceedings of the 2018 annual symposium on computer-human interaction in play companion extended abstracts. pp. 169–181 (2018)
7. Claravine: A query on using query strings/parameters (2026), `https://www.claravine.com/a-query-on-using-query-strings-parameters/#:~:text=A%20query%20string%20is%20a,separates%20each%20key%20and%20value.`, retrieved Apr. 20, 2026
8. CyberWire: Phish-prone percentage definition. `https://thecyberwire.com/glossary/phish-prone-percentage` (2020), accessed: 2026-07-08

9. De Silva, R., Nabeel, M., Elvitigala, C., Khalil, I., Yu, T., Keppitiyagama, C.: Compromised or {Attacker-Owned}: A large scale classification and study of hosting domains of malicious {URLs}. In: 30th USENIX security symposium (USENIX security 21). pp. 3721–3738 (2021)
10. Dhamija, R., Tygar, J.D., Hearst, M.: Why phishing works. In: Proceedings of the SIGCHI conference on Human Factors in computing systems. pp. 581–590 (2006)
11. Dutton, Sam: What are the parts of a url? (2026), `https://web.dev/articles/url-parts`, web.dev. Retrieved Apr. 20, 2026
12. Ebbinghaus, H.: A contribution to experimental psychology. New York, NY: Teachers College, Columbia University (1913)
13. Fernando, M., Arachchilage, N.A.G.: Why johnny can't rely on anti-phishing educational interventions to protect himself against contemporary phishing attacks? arXiv preprint arXiv:2004.13262 (2020)
14. Garera, S., Provos, N., Chew, M., Rubin, A.D.: A framework for detection and measurement of phishing attacks. In: Proceedings of the 2007 ACM workshop on Recurring malcode. pp. 1–8 (2007)
15. GeeksforGeeks: Components of a url (2026), `https://www.geeksforgeeks.org/computer-networks/components-of-a-url/`, retrieved Apr. 20, 2026
16. Google LLC: Domain name basics (2026), `https://support.google.com/a/answer/2573637?hl=en#S`, retrieved Apr. 20, 2026
17. HubSpot, Inc.: Parts of a url: A short guide (2026), `https://blog.hubspot.com/marketing/parts-url`, retrieved Mar. 25, 2026
18. IBM: The components of a url (2026), `https://www.ibm.com/docs/en/cics-ts/6.x?topic=concepts-components-url`, retrieved Apr. 20, 2026
19. ICANN: The domain name system (2026), `https://www.icann.org/resources/pages/dns-2022-09-13-en`, retrieved Apr. 20, 2026
20. Jabir, R., Le, J., Nguyen, C.: Phishing attacks in the age of generative artificial intelligence: A systematic review of human factors. AI **6**(8), 174 (2025)
21. KnowBe4: Phishing benchmarking analysis center. `https://www.knowbe4.com/phishing-benchmarking-analysis-center` (2026), accessed: 2026-07-08
22. Kolenbrander, J., Rheault, E., Michaels, A.J.: Privacy risks of cybersquatting attacks. Journal of Cybersecurity and Privacy **6**(1), 38 (2026)
23. Lain, D., Nakatsuka, Y., Kostiainen, K., Tsudik, G., Capkun, S.: {URL} inspection tasks: Helping users detect phishing links in emails. In: 34th USENIX Security Symposium (USENIX Security 25). pp. 1435–1454 (2025)
24. Lin, E., Greenberg, S., Trotter, E., Ma, D., Aycock, J.: Does domain highlighting help people identify phishing sites? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 2075–2084 (2011)
25. Ma, J., Saul, L.K., Savage, S., Voelker, G.M.: Beyond blacklists: learning to detect malicious web sites from suspicious urls. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1245–1254 (2009)
26. Mozilla Corporation: Public suffix list (2026), `https://publicsuffix.org/`, retrieved Apr. 20, 2026
27. Mozilla Corporation: Tld (2026), `https://developer.mozilla.org/en-US/docs/Glossary/TLD`, retrieved Apr. 20, 2026
28. Mozilla Corporation: Url fragment (2026), `https://developer.mozilla.org/en-US/docs/Web/URI/Reference/`, retrieved Apr. 20, 2026
29. Mozilla Corporation: What is a url? (2026), `https://developer.mozilla.org/en-US/docs/Learn_web_development/Howto/Web_mechanics/What_is_a_URL`, retrieved Apr. 20, 2026

30. Murre, J.M., Dros, J.: Replication and analysis of ebbinghaus' forgetting curve. PloS one **10**(7), e0120644 (2015)
31. Oest, A., Zhang, P., Wardman, B., Nunes, E., Burgis, J., Zand, A., Thomas, K., Doupé, A., Ahn, G.J.: Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale. In: 29th {USENIX} Security Symposium ({USENIX} Security 20) (2020)
32. OWASP: Path traversal (2026), `https://owasp.org/www-community/attacks/Path_Traversal`, retrieved Apr. 20, 2026
33. Prakash, P., Kumar, M., Kompella, R.R., Gupta, M.: Phishnet: predictive blacklisting to detect phishing attacks. In: 2010 Proceedings IEEE INFOCOM. pp. 1–5. IEEE (2010)
34. Proofpoint: The human factor 2025 – vol. 2: Phishing and url-based threats. `https://www.proofpoint.com/us/resources/threat-reports/human-factor-url-phishing` (2025), accessed: 2026-08-05
35. Ramraje, S., Sable, P.: Comparison of the effect of post-instruction multiple-choice and short-answer tests on delayed retention learning. The Australasian medical journal **4**(6), 332 (2011)
36. Reynolds, J., Kumar, D., Ma, Z., Subramanian, R., Wu, M., Shelton, M., Mason, J., Stark, E., Bailey, M.: Measuring identity confusion with uniform resource locators. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–12 (2020)
37. Sahidjuan, A.S.I., Amin, M.A., Ahaja, L.I., Ahog, A.A., Ladjahasan, R.T., Jul, R.K., Radjail, N.J., Sayadi, B.J., Sarabi, A.J., Tahil, S.K.: Understanding the impact of phishing attacks on organizational, security and trust. International Journal of Scientific Research and Modern Technology (IJFMR) (2024), `https://www.ijfmr.com/papers/2024/6/34230.pdf`, accessed: Apr. 20, 2026
38. Sameen, M., Han, K., Hwang, S.O.: Phishhaven—an efficient real-time ai phishing urls detection system. Ieee Access **8**, 83425–83443 (2020)
39. Sarno, D.M., Black, J.: Who gets caught in the web of lies?: Understanding susceptibility to phishing emails, fake news headlines, and scam text messages. Human Factors **66**(6), 1742–1753 (2024)
40. Semrush: Top websites ranking (2026), `https://www.semrush.com/trending-websites/in/all`, accessed: April 20, 2026
41. Sharma, P., Nagpal, B.: A study on url manipulation attack methods and their countermeasures. International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN pp. 0976–1353 (2015)
42. Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L.F., Hong, J., Nunge, E.: 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. pp. 88–99 (2007)
43. Shrestha, L., Balogun, H., Khan, S.: Ai-driven phishing: Techniques, threats, and defence strategies. In: International Conference on Global Security, Safety, and Sustainability. pp. 121–143. Springer (2023)
44. Tian, K., Jan, S.T., Hu, H., Yao, D., Wang, G.: Needle in a haystack: Tracking down elite phishing domains in the wild. In: Proceedings of the Internet Measurement Conference 2018. pp. 429–442 (2018)
45. WHATWG: Url living standard (2026), `https://url.spec.whatwg.org/`, retrieved Apr. 20, 2026
46. Xiong, A., Proctor, R.W., Yang, W., Li, N.: Is domain highlighting actually helpful in identifying phishing web pages? Human factors **59**(4), 640–660 (2017)