The Impact of Secure Transport Protocols on Phishing Efficacy

Zane Ma Joshua Reynolds Joseph Dickinson Kaishen Wang
Taylor Judd Joseph D. Barnes Joshua Mason Michael Bailey
{zanema2,joshuar3,jddicki2,kwang40,tjudd,jdbarns1,joshm,mdbailey}@illinois.edu
University of Illinois Urbana-Champaign

Abstract

Secure transport protocols have become widespread in recent years, primarily due to growing adoption of HTTPS and SMTP over TLS. Worryingly, prior user studies have shown that users often do not understand the security that is provided by these protocols and may assume protections that do not exist. This study investigates how the security protocol knowledge gap impacts user behavior by performing a phishing experiment on 266 users that A/B tests the effects of HTTP/HTTPS and SMTP/SMTP+TLS on phishing susceptibility. Secure email transport had minimal effect, while HTTPS increased the click-through rate of email phishing links (72.0% HTTPS, 60.0% HTTP) and the credential-entry rate of phishing sites (58.0% HTTPS, 55.6% HTTP). However, our results are merely suggestive and do not rise to the level of statistical significance ($p = 0.17$ click-through, $p = 0.31$ credential-entry). To better understand the factors that affect credential-entry, we categorized differences in browser presentation of HTTP/HTTPS and correlated participant susceptibility with browser URL display features. We administered a follow-up survey for phishing victims, which was designed to provide qualitative insights for observed outcomes, but it did not yield meaningful results. Overall, this study is a suggestive look at the behavioral impact of secure transport protocols and can serve as a basis for future larger-scale studies.

1 Introduction

The rapid adoption of secure network communication protocols (i.e. HTTPS, SMTP over TLS, etc.) in recent years is an encouraging sign for the future of Internet security. When implemented and utilized properly, these protocols provide confidentiality, validate message integrity, and verify identity/authenticity of the sender domain. Together, these measures establish secure communication channels that protect from both active man-in-the-middle attackers and passive eavesdroppers.

Secure data transport is a necessary, but insufficient, piece of overall network security. For example, TLS connections do not protect against drive-by downloads, phishing, or other web maliciousness. TLS does not provide security assertions for the content being delivered or whether delivered content is likely to be abused. In other words, TLS does not handle credibility or trustworthiness of the data or data source/recipient. In an alternate universe, where online identity is directly tied to real-world identity, users might rely on current social structures for determining trustworthiness. However, the identity guarantees provided by TLS only apply to digital identities. Digital identities are fluid—they are easily and frequently acquired, abandoned, or hijacked. Additionally, digital identities can be confusing or misleading due to a combination of poor technological design or limited user expertise (e.g. domain typos [35], Unicode homoglyphs [3], combosquatted domains [22]). These challenges aside, even when users correctly determine digital identity, they often encounter new identities and are expected to decide if they are trustworthy. Will `https://new-estore.com` steal their credit card information? Are users really in financial trouble if they don't pay off `https://irs-enforcement.com`? TLS, by design, provides no notion of credibility in these instances.

Prior work has hinted that some users believe HTTPS, in particular, is an indication of trustworthiness [13, 23, 29]. Although only a handful of users were found to hold this misconception in prior studies, the data were free-form and self-reported and might not represent actual user behavior. In this paper, we attempt to quantitatively measure user behavior in the wild to answer the following: 1) Do users mistakenly assign credibility to secure communication protocols? 2) If so, what causes this misconception? These questions have become increasingly pressing as certificate authorities such as Let's Encrypt have recently made HTTPS more affordable and accessible to the entire web, including phishing operators.

To gain insight into these questions through experimentation, we measured user behaviour as a proxy for the credibility they attributed to different secure protocols. Concretely, we conducted an IRB-approved phishing exercise to A/B test the usage of two applications of the TLS protocol: TLS for SMTP email delivery and HTTPS for website delivery. The experiment was designed to *measure phishing effectiveness (click-through and credential-entry rates) as a proxy for user perception of credibility*. The phishing exercise was performed on 266 employees within a university IT organization and tracked the full phishing pipeline of email delivery, email opening, link clicking, and credential submission.

For users that opened the email, we found that 72% of users receiving an HTTPS link and 60% of users receiving an HTTP link visited the phishing website. However, this trend of HTTPS links increasing phishing email click-through rates was not statistically significant ($p = 0.17$). On the other hand, first-hop email delivery over SMTP+TLS compared to SMTP alone did not influence phishing click-through rate (63% compared to 65%, $p = 0.96$).

For the 92 users that clicked the phishing link (37% click-through rate), we used HTTP User-Agent headers to identify their web browsers. To understand variations in phishing website presentation by different browsers, we then utilized CrossBrowserTesting to generate 2,882 screenshots encompassing the HTTP and HTTPS versions of the experimental phishing website, across multiple platforms and eight browsers. Labeling the visual URL bar features for each screenshot allowed for correlation between phishing website display and user phishing behaviour. We found the most significant features, protocol presence and default favicon, to have likely, but inconclusive, correlation with credential entry to the phishing site ($p = 0.07$ and $p = 0.06$, respectively).

Finally, users that entered their credentials were directed to an informed consent / experiment withdrawal page and then presented with a follow-up user survey. The survey contained questions designed to capture self-reported reasons for phishing susceptibility and questions meant to determine the predisposing conditions, e.g. demographics, secure behaviour awareness, and risk tolerance, that could bias phishing outcomes. We found that the five open-ended explanations for phishing fallibility were vague and did not provide meaningful content. Comparison of additional phishing factors revealed no significant differences or similarities between the different treatment group populations in our study.

Our phishing experiments ultimately highlight the credibility that users attribute to network protocols and the content that they are exposed to. This work provides a behavioral perspective that supports the mounting evidence that users often ignore or misunderstand secure transport protocols, which can result in negative security outcomes. Our methodology and initial results signal the need for a larger scale study that identifies specific misconceptions and their causes. Better understanding of this knowledge and behavioral gap can lead to more usable and effective security.

2 Background

Phishing is a social engineering technique designed to obtain sensitive information through disguise as a trustworthy entity. Phishing typically begins through email or instant messaging and directs users to a fraudulent website that engages in harmful activity such as requesting credit card information or installing malware. The fundamental issue underlying phishing is user misidentification or mistrust of online entities. In an attempt to mitigate phishing through better user training [25, 33] or automated technical defenses [2, 21], many studies have identified a multitude of factors around both phishing email features [14, 15, 17, 34] and phishing target characteristics [5, 32]. This study focuses on the phishing impact of secure transport protocols, which do not provide guarantees of end-entity legitimacy.

The concept of HTTPS as a signal of credibility has been discussed by several prior works [7, 13, 16], but its impact on user phishing behavior has never been investigated in depth. Prior work in this area has been largely self-reported and qualitative or conducted on a small laboratory scale [7, 20, 24, 27, 31, 36]. The general result from these studies indicates that users who are on the lookout for phishing often look at HTTPS indicators, but those who are not primed to identify suspicious websites typically ignore security indicators. These studies examine multiple phishing trust factors simultaneously and do not isolate and dive into the specific causes for trust in HTTPS. No studies have examined the effect of secure communication protocols on phishing efficacy outside of a laboratory setting. Additionally, all prior studies related to phishing were conducted in 2007 or earlier, before the upswing of HTTPS and SMTP over TLS in the last decade [8, 12].

Ruoti et al. interviewed suburban adults about their online security posture and discovered that many participants associated TLS indicators with site security, rather than connection security [29]. The authors suggest that phishing operators could potentially abuse this misconception, which we attempt to measure in the wild in this study. Krombholz et al. performed a study of users' HTTPS mental models and found three out of eighteen end-users who mistakenly believed that HTTPS protected against phishing [23]. These academic perspectives are aligned with reports of malware and phishing domains using HTTPS due to an increasingly free and automated web PKI system [1, 18].

HTTPS is not the only secure transport protocol that might be mistaken by users as a sign of credibility. To our knowledge, no prior work has studied SMTP+TLS as a potential phishing influence, likely because few email clients display transport security information, with one notable exception being GMail [19].[1]

[1] As of July 2019, GMail appears to have removed email security indicators.

3 Methodology

This study takes a three-tiered approach to measuring the impact of secure transport protocols (Figure 1) on phishing. First, phishing email click-through rates and phishing site credential-entry rates were measured to indicate the degree of trust/credibility that users assign to email and web content. Second, we labeled browser screenshots to identify the visual differences experienced by users that correlate with phishing credential entry. Third, we conducted a survey for susceptible users to discover causes for the observed phishing outcomes and rule out participant biases that could affect the results.

3.1 Target Population and Test Groups

We examined the impact of two secure transport protocols: HTTPS and SMTP over TLS. Both protocols utilize TLS, which provides integrity, confidentiality, and/or authenticity.[2] HTTP or HTTPS was used for the link embedded in the phishing email and also the corresponding phishing website. Plaintext SMTP and SMTP over TLS were used to deliver mail to the first hop SMTP relay. We modulated these two protocols and randomly divided the users into four equal experimental groups. The phishing experiment targeted 266 staff members of a university IT team.

3.2 Phishing Experiment

We built a customized phishing tool based on the open-source projects Gophish [37], a phishing experiment administration tool, and Tmail [6], an email server written in GoLang.

We coordinated with the main university IT team[3] and sent a phishing email (Appendix A) that employed effective phishing techniques identified by prior research in order to maximize efficacy. The email was sent from an email address belonging to a fabricated member of the main university IT office. The email warned recipients that their device had been potentially compromised, and instructed them to log in and verify device ownership in order to prevent their device from being banned from the campus network. In addition to employing distraction [15], authority [14], and urgency [34], the phishing email was content rich [17] and contained many visual elements used in legitimate university emails, including a university logo image that was used to track email opening. The email contained both a directly embedded link to the phishing site and a plaintext URL that the user could copy-paste into their browser. The embedded link URL and plaintext URL were identical and contained a unique tracking ID for each individual target email address. We refer to both methods of accessing the phishing website (copy-pasting the URL and direct link clicking) as part of the "click-through" rate throughout this work.

The phishing site resided at `illinois-abuse.com`, a non-university domain that was believably related to the content of the phishing email. The site was a visual clone of Shibboleth, the single sign-on (SSO) service used to access most university resources. When a user attempted to login with any username and password, the sensitive credentials were redacted in the browser and replaced with one of three values, based on the length of the username and password: *valid* according to university restrictions on username and password lengths, *invalid*, and *empty*. Upon receiving credential validity, the server immediately redirected users to an informed consent page. This page disclosed details of the phishing experiment, provided educational information, allowed opting-out of the study, and linked to an optional follow-up survey.

The phishing experiment was tracked through server request logs. Specifically, we recorded four types of events: *email sending*, *email opening* when the email logo image was downloaded, *phishing site access* when an HTTP GET request was made to the phishing site, and *credential entry* when receiving an HTTP POST request with username and password validity parameters. We used HTTP headers to prevent data caching and kept track of repeated events, e.g. a user clicking on their phishing link more than once. We also recorded HTTP headers to understand what email and web clients were used, so that we could connect user software with user behavior (Section 3.4). These tracking techniques are imperfect (e.g. some email clients may not load tracking images or web clients with JavaScript disabled), so we treat our results as a lower bound. Because we randomly assigned users to different treatment groups, we also expect tracking evasion to be evenly distributed across each group.

3.3 Follow-up Survey

To understand why users were vulnerable to the phishing exercise, we included a link to an optional follow-up survey with a $10 participation reward on the final phishing review page. The full survey can be found in Appendix B.1, but broadly speaking, we asked the following questions:

- *Demographics*: standard questions about age, sex, level of education/employment, and university affiliation.
- *Prior Knowledge*: whether the user had previously heard of the study in order to discard responses for users that had prior knowledge.
- *Motivation*: what factors caused users to fall for phishing and if/how security cues influenced their actions.
- *Computer Expertise*: SeBIS survey questions [11] to quantify users' computer security behaviors and three questions from Levesque et al. [26] to measure computer expertise.
- *Risk Attitude*: DOSPERT survey questions [4] to measure a user's likelihood to engage in risky behavior.

We embedded 5 *attention questions* that instruct the user

[2] The certificates used in SMTP over TLS are often self-signed and may not provide authenticity.

[3] Phishing email delivery was ensured by coordinating the circumvention of rate limits and spam filters.

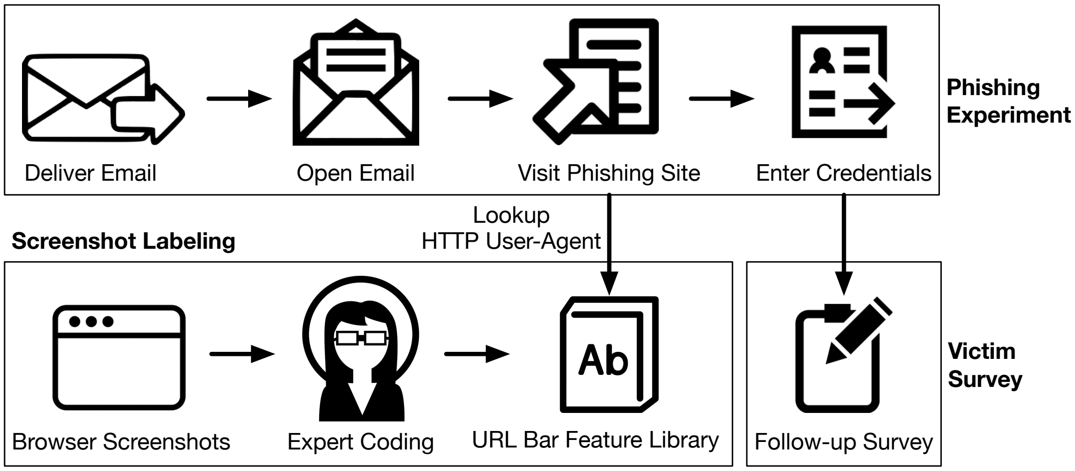


Figure 1: **Experimental Design**—The study consists of an email phishing experiment to measure click-through and credential entry rates, browser screenshot labeling to codify URL bar features, and a user survey to capture behavioural motivation.

to choose a specific answer in order to detect users who were not paying attention. We also followed the standard surveying practice of randomizing the order of question sections and response options, where applicable. To collect baseline values for our target population, we emailed 600 random members of the university community in December 2015, in which we asked users to complete a version of the survey that included the questions related to demographics, computer expertise, and risk attitude. Users who completed the baseline survey were compensated with a $10 Amazon gift card.

3.4 Security Indicators

The way in which phishing emails and websites are presented to end users can influence their susceptibility to phishing, especially when visual warnings are present [9]. Capturing this information for our study was challenging since the experiment was carried out in the wild, not in a lab setting, where users employed a wide range of software to view both the phishing email and phishing webpage. Unfortunately, we could not preemptively inform users of the phishing experiment and request screenshots of their email client / web browser. We likewise could not ethically install software to surreptitiously screenshot each participant's phishing experience. Thus, we took an alternate approach and logged the User-Agent HTTP header to determine what software was being used.

Email clients were not tested/coded, since in brief initial testing, HTTP User-Agent and Referer headers were not a reliable indicator for browser-based email clients. Future work coding the visual appearance of these clients could provide insight explaining user disposition towards phishing emails.

We used CrossBrowserTesting to test 8 browsers, which amounted to 1,441 total configurations across different browser and OS versions. We generated a screenshot of our phishing website loaded in each browser, once for HTTP and once for HTTPS. These 2,882 screenshots were manually coded to identify different features of the browser URL bar. Example features include the type and color of URL bar icons, or the presence of additional text (e.g. "Secure") and its background color. Three coders first agreed to a codebook of 19 feature categories based on a random 50 screenshot sample and then independently labeled two-thirds of the overall screenshots, so that each screenshot was coded twice. The average Cohen's kappa across 19 coding categories was 0.855 (Appendix C). We performed manual conflict resolution to create the final ground truth labeling. Our automatically-generated screenshot dataset was missing data for Chrome 64/65, so we manually verified Chrome 64/65 on Windows 10 and Mac OS 10.11.6. We extrapolated these results to other platforms, since we observed version consistency across platforms for previous Chrome versions. This data is available open-source to the broader research community.[4]

3.5 Ethics

We received IRB approval for our phishing experiment, addressing the following concerns:

- *Consent*: As part of their employment contracts, university employees agree to comply with "Periodic [*sic*] security assessments" [28]. Additionally, as soon as users entered their credentials to the phishing site, we presented an informed consent page that explained the phishing research and allowed users to opt out of the study. We additionally sent a follow-up email to all participants at the conclusion of the study detailing the phishing experiment and results.
- *Minors*: We automatically excluded any individuals who reported to be under 18 years of age.
- *Privacy*: University IT staff members operated the phishing tool and only exported anonymized results to researchers.

[4] https://github.com/teamnsrg/url-bar-coding

Each individual was assigned a random pseudonymous identifier to link phishing vulnerability logs with survey responses. None of the survey questions or HTTP headers were deemed to contain personally identifying information.

4 Results

We performed the phishing test over a 24 hour period beginning on March 28, 2018 at 17:30 local time. We initiated the phishing emails thirty minutes after the end of the workday in order to avoid the effects of office gossip that would inform participants of the research study prior to examining the email. While the timing of the experiment could skew the types of devices that participants used to view and access phishing content, we prioritized the maximization of uninformed participants. After completing data collection, we first sanitized the data and then performed three analyses: phishing efficacy across treatment groups, browser indicator correlation with phishing outcomes, and survey evaluation of user behavior.

4.1 Data Sanitization

Prior to data analysis, we sanitized the data to remove noise generated by curious or mischievous actors. First, we removed all user events after first data submission, since users were immediately informed that the phishing email was a research exercise. Second, we removed all user events for users who only submitted invalid data, i.e. usernames or passwords that did not fit within university imposed character limits. This behavior indicates that the user had prior knowledge or suspicion of the phishing exercise. Finally, we removed all events for users who accessed the phishing site five or more times within our twenty-four hour study period, as this could indicate deviousness rather than organic phishing susceptibility. In total, we removed 428 of 1633 (26.2%) overall user events, which were associated with 55 of 266 (20.5%) total users, including 19 users (7.1%) for whom all events were removed. Five surveys were discarded from these curious/mischievous actors, and an additional two surveys were discarded for missing more than half of the five attention-check questions. All subsequent analysis reflects the data after sanitization.

4.2 Phishing Efficacy

Overall, the phishing campaign produced a 37% click-through rate and 23% credential submission rate (Table 1). Notably, of the 247 emails that were sent, only 56.7% were detectably opened, reducing the effective sample size by nearly half. To quantify the significance of the two test features, we analyzed two perspectives of the data: split by HTTP versus HTTPS and split by SMTP versus STMP over TLS. We apply both chi-squared (χ^2) and Fisher's exact tests to each perspective (Table 2).

We found that the presence of HTTPS links in phishing emails, compared to HTTP links, resulted in higher observed email open rates and click-through rates. However, the likelihood that both these outcomes arose from different underlying distributions is 83%, which does not satisfy the bar for statistical significance. While we did not study how email clients handled our phishing emails, one possibility for the perceived email opening rate difference is due to variations in email clients, such as browser-based web clients that may selectively ignore HTTP email images due to mixed content concerns.

To understand the differences in click-through and data submission rate, we measured the possible correlations to visual browser cues in Section 4.3. The usage of SMTP over TLS had minimal effect on the effectiveness of any of the three stages of phishing. One unexplored possibility is the lack of any differences observed by the end-user, since many email clients anecdotally do not display indicators of email transit security.

4.3 Browser URL Indicators

To better understand the causes for potential phishing susceptibility differences between HTTP and HTTPS, we examined how protocols are visually distinguished by different browsers. Specifically, we codified browser URL bar features and measured their correlation to the efficacy of credential entry on our phishing website. Our feature codebook contained browser information for 69 out of 92 (75%) users who accessed the phishing site. In particular, we computed Pearson's chi-squared test to measure whether credential-entry correlates with browser URL indicators, from two perspectives: the URL security indicators that were present at the time of credential submission, which is dependent on a user's treatment group, and the URL security indicators that a user would typically see for an HTTPS site (Table 3). We found no statistically significant correlations ($p < 0.05$), although protocol presence and default favicon could potentially correlate with phishing ($p < 0.25$). Most desktop browsers only display the URL protocol for HTTPS, and not HTTP. The two primary mobile browsers, Mobile Safari and Mobile Chrome, behave similarly with the exception that Mobile Safari never displays the protocol even for HTTPS. The influence of the default favicon on credential-entry rate is likely due to the close mapping between browser and default favicon. Most browsers use the same default favicon, with the exception of some versions of Internet Explorer and Opera.

Treatment Group	Emails	Emails Opened		Links Visited		% Prev.	Credentials Entered		% Prev.	Surveys Filled		% Prev.
HTTP + SMTP	61	36	(59.0%)	21	(34.4%)	58.3%	13	(21.3%)	61.9%	3	(4.9%)	23.1%
HTTP + SMTP/TLS	61	39	(63.9%)	24	(39.3%)	61.5%	12	(19.7%)	50.0%	1	(1.6%)	8.3%
HTTPS + SMTP	62	33	(53.2%)	24	(38.7%)	72.7%	14	(22.6%)	58.3%	2	(3.2%)	14.3%
HTTPS + SMTP/TLS	63	32	(50.8%)	23	(36.5%)	71.9%	18	(28.6%)	78.3%	4	(6.3%)	22.2%
Total	247	140	(56.7%)	92	(37.2%)	65.7%	57	(23.1%)	62.0%	10	(4.0%)	17.5%

Table 1: **Phishing Efficacy**—The phishing campaign produced a 37% click-through rate and 23% credential entry rate. The four treatment groups have slightly different sizes due to data sanitization removal.

Treatment Group	Opened Email				Visited Link				Entered Credentials			
	Ratio	%	$p_{\tilde{\chi}^2}$	$p_{Fisher's}$	Ratio	%	$p_{\tilde{\chi}^2}$	$p_{Fisher's}$	Ratio	%	$p_{\tilde{\chi}^2}$	$p_{Fisher's}$
HTTP	75/122	61.5%	0.17	0.16	45/75	60.0%	0.17	0.15	20/45	55.6%	0.31	0.28
HTTPS	65/125	52.0%			47/65	72.3%			32/47	68.0%		
SMTP	69/123	56.1%	0.96	0.90	45/69	65.2%	0.96	1.00	27/45	60.0%	0.87	0.83
SMTP+TLS	71/124	57.3%			45/71	63.3%			30/47	63.8%		

Table 2: **Security Protocol Significance**—We find that phishing is potentially more effective with HTTPS than HTTP, but unlikely to be more effective with first-hop SMTP+TLS compared to SMTP.

Category	DoF	$p_{\text{exp}\ \tilde{\chi}^2}$	$p_{\text{https}\ \tilde{\chi}^2}$
Any Icon?	1	0.25	–
Lock Icon?	1	0.32	0.71
Lock Position	1	0.98	0.87
Lock Color	3	0.55	0.66
Detailed Lock?	1	0.54	0.87
Lock Additions	1	0.27	–
Favicon?	1	0.56	0.23
Favicon Position	1	0.32	0.67
Default Favicon	2	0.06	0.06
Protocol Visible?	1	0.07	0.46
Protocol Emphasis	2	0.63	0.12
Additional Text?	2	0.62	0.22
Add. Text Emphasis	2	0.62	0.22
Add. Text Background	1	0.97	0.22
Icon/URL Separator?	1	0.42	0.42

Table 3: **Browser feature phishing correlation**—$\tilde{\chi}^2$ correlation between URL bar features and credential submission rates. We consider the features observed by each user during the actual experiment, and the HTTPS features a user would normally see.

DOSPERT [4]						
Comparison	μ_1	σ_1	μ_2	σ_2	t	p
HTTP/HTTPS	91.25	9.46	86.33	12.42	0.71	0.50
SMTP/SMTP+TLS	87.4	8.08	89.2	14.39	-0.24	0.82
SEBIS [11]						
Comparison	μ_1	σ_1	μ_2	σ_2	t	p
HTTP/HTTPS	57.25	3.86	55.5	6.77	0.52	0.62
SMTP/SMTP+TLS	53.0	5.83	59.4	3.29	-2.14	0.07

Table 4: **DOSPERT / SeBIS treatment group comparison**—Welch's t-test was performed to compare the distribution of risk perception and security intentions of the different treatment groups in our study. No significant similarity or difference was found for either test, although the SMTP/SMTP+TLS division had a potential bias towards more security aware users in the latter group.

4.4 Survey of Phishing Victims

The survey served two primary purposes. First, using Welch's t-test, we attempted to identify characteristics of our target populations that have been previously revealed to influence phishing outcomes (Table 4). DOSPERT, which tests user's risk behaviors [4], revealed no statistical significance regarding the similarity between our both HTTP versus HTTPS and SMTP vs SMTP+TLS test groups. Likewise, SeBIS, which tests security behavior intentions [11] and has been linked to phishing susceptibility [10], achieved similar inconclusive results. Compared to the university population and reference populations (Table 5), our study population of IT professionals did not differ in risk perception, but differed in two aspects of security behavior intentions.

The second purpose of the survey was to understand user-reported causes for their susceptibility to the phishing experiment. Two researchers independently reviewed the open-ended questions to generate codebooks, which were manually consolidated into a final codebook. All ten questions were subsequently coded by two researchers, and differences were resolved through discussion until there was complete agreement in coding. We found that phishing victims were primarily motivated by a sense of concern/importance (6/10) and legitimate visual imitation (6/10). These responses demonstrated a reliance on visual features and concern-fueled susceptibility. Only three out of ten victims mentioned the URL or sender email address, which are the only technical mechanisms for identifying information veracity. Unfortunately, email addresses are still spoofable for a large number of domains [8], and the URL in the URL bar is the only reliable way to assess the credibility of potential phishing emails.

We asked follow-up questions to the five respondents that noticed the presence/absence of security indicators in the URL bar, to gauge user understanding of the security indicators. User responses (listed in Appendix B.2) were vague and

DOSPERT Comparison with University Population						
Subscale	μ_S	σ_S	μ_P	σ_P	t	p
Ethical	11	21.0	12.0	4.2	-0.15	0.89
Financial	15	12.6	13.9	6.2	0.28	0.79
Health/Safety	13.7	29.0	16.1	6.3	-0.27	0.80
Recreational	19.5	25.3	18.2	6.4	0.16	0.88
Social	29.9	22.5	27.3	6.6	0.36	0.73
DOSPERT Comparison with Reference Population [4]						
Ethical	11	21.0	12.0	2.0	-0.89	0.40
Financial	15	12.6	13.9	2.5	-1.15	0.28
Health/Safety	13.7	29.0	16.1	2.5	-0.76	0.47
Recreational	19.5	25.3	18.2	2.5	-0.37	0.72
Social	29.9	22.5	27.3	2.6	-0.38	0.72
SEBIS Comparison with Reference Population [11]						
Device Securement	4.5	.43	3.2	1.5	8.57	6.4e-07
Password Generation	3.4	.43	3.3	1.1	0.69	0.50
Proactive Awareness	2.8	.18	3.7	1.0	-12.4	8.6e-12
Updating	3.5	.29	3.5	1.1	0.0	1.0

Table 5: **DOSPERT / SeBIS population comparison**—We compared the risk perception and security behavior intentions of our population sample, S, to the general university population and reference populations, P. Using a threshold of $p < 0.05$, significant differences were observed for SeBIS device securement and proactive awareness.

varied: one did not know what the indicators meant, others associated HTTP indicators with general danger or a need to be cautious, and one user described HTTPS content as "secure and protected". This portion of the survey did not provide insight into which specific security properties users understand and attribute to security protocols. The open-ended questions were an attempt to collect organic responses and avoid biasing users towards specific answers, but the responses were too general to be useful. Future survey design should consider testing user agreement with provided descriptions of different security properties.

5 Discussion

One of the main limitations of this study was its limited sample size. The nuanced effects we observed, coupled with a progressively diminishing sample size at each stage in the phishing pipeline, led to a string of suggestive, but statistically inconclusive ($p > 0.05$) results. Performing this study at a larger scale would likely improve confidence in the observed impact of HTTPS on phishing efficacy and could uncover subtle new effects as well. The second limitation of the study was its non-representative participant population, who were mostly technology professionals, and exhibited more security aptitude via SeBIS than a reference population (Table 5). It is possible that a more representative population would have greater technical security misconceptions and exhibit even more pronounced phishing susceptibility differences.

Assuming that user misunderstanding of HTTPS does manifest in increased phishing susceptibility, the next logical step would be to identify the origins of misunderstanding in order to correct it. One hypothesis that could be substantiated through a larger-scale browser coding experiment is that browser indicators inadvertently habituate users to associate their own colloquial definition of security with HTTPS. Coincidentally, Chrome recently shifted away from positive signaling around HTTPS to neutral signaling (removed "Secure" text in URL bar), and has shifted signaling around HTTP from neutral to negative [30].

Erroneous attribution of credibility/trustworthiness to HTTPS points to a fundamental divide between what users consider to be secure and the security actually provided by HTTPS. This gap presents two paths forward: educate users to understand HTTPS better and make better security decisions, or adopt new security protocols that match the existing expectations of users. Parallel research efforts into both approaches can help mitigate the core issues underlying phishing attacks.

6 Conclusion

This study attempts to empirically evaluate the notion within the security community that users erroneously attribute credibility and trustworthiness to secure transport protocols. We designed and executed a three-pronged phishing experiment that A/B tests phishing susceptibility rates in the wild for secure/insecure transport protocols, performs correlation tests between user behavior and phishing website display features, and corroborates results with a user survey. Ultimately, our results are merely suggestive and do *not* statistically prove that users are more likely to click and enter credentials to phishing links/sites that are served over HTTPS, while no effect is observed for TLS-secured email delivery. We hope to provide the kindling and spark for future research into the security problems and opportunities surrounding the user-centric concept of trustworthiness on the internet.

Acknowledgments

The authors thank Eric Frahm, Charles Geigner, and the Technology Services Security team for their feedback and help deploying this study. We wish to also thank our anonymous reviewers who provided helpful comments on earlier drafts of the manuscript. This work was supported in part by the National Science Foundation under contract CNS 1518741, and the State Farm Companies Foundation Doctoral Scholar Program. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not reflect the views of their employers or sponsors.

References

[1] J. Aas. The CA's role in fighting phishing and malware. https://letsencrypt.org/2015/10/29/phishing-and-malware.html.

[2] S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair. A comparison of machine learning techniques for phishing detection. In *Anti-Phishing Working Group 2nd Annual eCrime Researchers Summit*, 2007.

[3] P. Agten, W. Joosen, F. Piessens, and N. Nikiforakis. Seven months' worth of mistakes: A longitudinal study of typosquatting abuse. In *Network & Distributed Systems Symposium*, 2015.

[4] A.-R. Blais and E. U. Weber. A domain-specific risk-taking (DOSPERT) scale for adult populations. 2006.

[5] M. Blythe, H. Petrie, and J. A. Clark. F for fake: four studies on how we fall for phish. In *SIGCHI Conference on Human Factors in Computing Systems*, 2011.

[6] S. Depierrepont. tmail. https://github.com/toorop/tmail.

[7] R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. In *SIGCHI Conference on Human Factors in Computing Systems*, 2006.

[8] Z. Durumeric, D. Adrian, A. Mirian, J. Kasten, E. Bursztein, N. Lidzborski, K. Thomas, V. Eranti, M. Bailey, and J. A. Halderman. Neither snow nor rain nor MITM...: An empirical analysis of email delivery security. In *15th ACM Internet Measurement Conference*, 2015.

[9] S. Egelman, L. F. Cranor, and J. Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *SIGCHI Conference on Human Factors in Computing Systems*, 2008.

[10] S. Egelman, M. Harbach, and E. Peer. Behavior ever follows intention?: A validation of the security behavior intentions scale (SeBIS). In *SIGCHI Conference on Human Factors in Computing Systems*, 2016.

[11] S. Egelman and E. Peer. Scaling the security wall: Developing a security behavior intentions scale (SeBIS). In *SIGCHI Conference on Human Factors in Computing Systems*, 2015.

[12] A. P. Felt, R. Barnes, A. King, C. Palmer, C. Bentzel, and P. Tabriz. Measuring HTTPS adoption on the web. In *USENIX Security Symposium*, 2017.

[13] A. P. Felt, R. W. Reeder, A. Ainslie, H. Harris, M. Walker, C. Thompson, M. E. Acer, E. Morant, and S. Consolvo. Rethinking connection security indicators. In *Symposium on Usable Privacy and Security SOUPS*, 2016.

[14] A. Ferreira, L. Coventry, and G. Lenzini. Principles of persuasion in social engineering and their use in phishing. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*, 2015.

[15] A. Ferreira and G. Lenzini. An analysis of social engineering principles in effective phishing. In *Workshop on Socio-Technical Aspects in Security and Trust STAST*, 2015.

[16] S. Gastellier-Prevost, G. G. Granadillo, and M. Laurent. Decisive heuristics to differentiate legitimate from phishing sites. In *Conference on Network and Information Systems Security*, 2011.

[17] B. Harrison, A. Vishwanath, Y. J. Ng, and R. Rao. Examining the impact of presence on individual phishing victimization. In *System Sciences (HICSS)*, 2015.

[18] C. Hassold. A quarter of phishing attacks are now hosted on HTTPS domains: Why? https://info.phishlabs.com/blog/quarter-phishing-attacks-hosted-https-domains.

[19] G. Help. Email encryption in transit. https://support.google.com/mail/answer/6330403?hl=en:.

[20] M. Jakobsson, A. Tsow, A. Shah, E. Blevis, and Y.-K. Lim. What instills trust? A qualitative study of phishing. *Financial Cryptography and Data Security*, 2007.

[21] M. Khonji, Y. Iraqi, and A. Jones. Phishing detection: a literature survey. *IEEE Communications Surveys & Tutorials*, 2013.

[22] P. Kintis, N. Miramirkhani, C. Lever, Y. Chen, R. Romero-Gómez, N. Pitropakis, N. Nikiforakis, and M. Antonakakis. Hiding in plain sight: A longitudinal study of combosquatting abuse. In *ACM Conference on Computer & Communications Security*, 2017.

[23] K. Krombholz, K. Busse, K. Pfeffer, M. Smith, and E. von Zezschwitz. "If HTTPS were secure, i wouldn't need 2FA"- end user and administrator mental models of HTTPS. In *IEEE Security & Privacy*, 2019.

[24] P. Kumaraguru, A. Acquisti, and L. F. Cranor. Trust modelling for online transactions: a phishing scenario. In *International Conference on Privacy, Security and Trust*, 2006.

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

[26] F. Lalonde Levesque, J. Nsiempba, J. M. Fernandez, S. Chiasson, and A. Somayaji. A clinical study of risk factors related to malware infections. In *ACM SIGSAC Conference on Computer & Communications Security*, 2013.

[27] E. Lin, S. Greenberg, E. Trotter, D. Ma, and J. Aycock. Does domain highlighting help people identify phishing sites? In *SIGCHI Conference on Human Factors in Computing Systems*, 2011.

[28] U. of Illinois. Information security policy. http://cam.illinois.edu/policies/fo-36/.

[29] S. Ruoti, T. Monson, J. Wu, D. Zappala, and K. Seamons. Weighing context and trade-offs: How suburban adults selected their online security posture. In *Symposium on Usable Privacy and Security SOUPS*, 2017.

[30] E. Schechter. A secure web is here to stay. https://security.googleblog.com/2018/02/a-secure-web-is-here-to-stay.html.

[31] S. E. Schechter, R. Dhamija, A. Ozment, and I. Fischer. The emperor's new security indicators. 2007.

[32] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs. Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In *SIGCHI Conference on Human Factors in Computing Systems*, 2010.

[33] S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L. F. Cranor, J. Hong, and E. Nunge. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In *Symposium on Usable Privacy and Security SOUPS*, 2007.

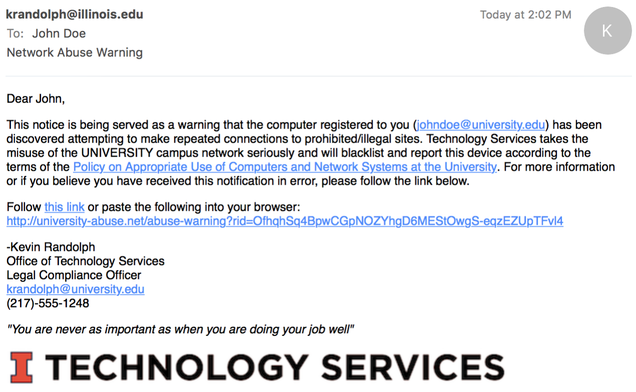
[34] J. Wang, R. Chen, T. Herath, and H. R. Rao. An exploration of the design features of phishing attacks. *Information Assurance, Security and Privacy Services*, 2009.

[35] Y.-M. Wang, D. Beck, J. Wang, C. Verbowski, and B. Daniels. Strider typo-patrol: Discovery and analysis of systematic typo-squatting. *SRUTI*, 2006.

[36] T. Whalen and K. M. Inkpen. Gathering evidence: use of visual security cues in web browsers. In *Proceedings of Graphics Interface*, 2005.

[37] J. Wright. Gophish. https://github.com/gophish/gophish.

Appendix A Phishing Email



krandolph@illinois.edu Today at 2:02 PM

To: John Doe

Network Abuse Warning

Dear John,

This notice is being served as a warning that the computer registered to you (johndoe@university.edu) has been discovered attempting to make repeated connections to prohibited/illegal sites. Technology Services takes the misuse of the UNIVERSITY campus network seriously and will blacklist and report this device according to the terms of the Policy on Appropriate Use of Computers and Network Systems at the University. For more information or if you believe you have received this notification in error, please follow the link below.

Follow this link or paste the following into your browser:
http://university-abuse.net/abuse-warning?rid=OfhqhSq4BpwCGpNOZYhgD6MEStOwgS-eqzEZUpTFvl4

-Kevin Randolph
Office of Technology Services
Legal Compliance Officer
krandolph@university.edu
(217)-555-1248

"You are never as important as when you are doing your job well"

I TECHNOLOGY SERVICES

Appendix B Survey

B.1 Questions

Are you male or female? [Female, Male, Other, Prefer not to answer]
What is your age? [17 or younger, 18-20, 21-29, 30-39, 40-49, 50-59, 60 or older, Prefer not to answer]
What is the highest level of school you have completed or the highest degree you have received? [Less than high school degree, High school degree or equivalent (e.g. GED), Some college but no degree, Associate degree, Bachelor degree, Graduate degree, Prefer not to answer]
Which of the following categories best describes your employment status? [Employed, working full-time, Employed, working part-time, Not employed, looking for work, Not employed, NOT looking for work, Retired, Disabled / not able to work, Prefer not to answer]
Please select your affiliation with the REDACTED, if any. [Faculty, Staff, Graduate Student, Undergraduate Student, No affiliation, Prefer not to answer]
Had you heard any information about this specific research study in the past? [Yes, No, Prefer not to answer]
Why did you open the phishing email?
Why did you click on the link in the phishing email?
Why did you enter your credentials on the phishing website?
What security indicators did you notice in the phishing email? [HTTP/HTTPS URL, URL bar lock icon, Other (please specify), None, Prefer not to answer]
{If indicator specified} How did [security indicator] influence your decision to click or not click on the link in the email? {If indicator specified} What does [security indicator] mean to you?
What security indicators did you notice on the phishing website? [HTTP/HTTPS URL, URL bar lock icon, Other (please specify), None, Prefer not to answer]
If indicator specified How did [security indicator] influence your decision to submit your credentials on the phishing website?
If indicator specified What does [security indicator] mean to you?
SEBIS survey [11]
DOSPERT survey [4]

B.2 Responses

	Motive Questions				Indicator Questions	
Code	Why Opened Email	Why Clicked Link	Why Entered Credentials	**Code**	Email Security Indicators	Site Security Indicators
Concern/Importance	6 / 10	6/10	2 / 10	Not Sure	1/5	1/5
Urgency	1 / 10	0/10	0 / 10	Means Danger	2/5	2/5
Authority	1 / 10	0/10	0 / 10	Means Secure	1/5	1/5
Trusted Sender	3 / 10	2/10	2 / 10	Means Be Cautious	1/5	0/5
Looked Legitimate	2 / 10	4/10	6 / 10	Look at URL & Protocol	0/5	1/5
Lack of Attention	0 / 10	0/10	2 / 10			
Failed to Look at URL Bar	0 / 10	0/10	2 / 10			
Trusted Email Filter	0 / 10	1/10	0 / 10			
Errors More Likely Than Attacks	0 / 10	0/10	1 / 10			

Table 6: **Coded Survey Response**—Open-ended questions were coded to categorize the motives for phishing susceptibility and user understanding of security indicators. Security indicator questions were only asked if participants indicated they had noticed one in the email or website.

Appendix C URL Bar Security Indicator Codebook

Category	Values	Cohen's Kappa	Category	Values	Cohen's Kappa
Any Icon?	True/False	0.970	Protocol Separate?	True/False	0.985
Lock Icon?	True/False	0.999	Protocol Emphasis	Bold/Green/None	0.685
Lock Position	Left/Right of URL	0.999	Protocol Obscured	Truncated/None	0.613
Lock Color	Black/Gold/Green/Grey	0.994	Additional Text?	True/False	0.996
Detailed Lock?	True/False	0.773	Add. Text Position	Left URL	0.614
Lock Additions	Red slash/None	0.999	Add. Text	Not Secure/ Secure/ Web/ HTTPS/ Domain	0.686
Favicon?	True/False	0.975	Add. Text Emphasis	Blue/Bold/Green/None/White	0.449
Favicon Position	Left URL/Left Tab/Both	0.980	Add. Text Background	Blue/Grey/None/Yellow	0.981
Default Favicon	Square/ Page Icon/ Globe/ IE Icon/ Opera Icon	0.775	Icon/URL Separator?	True/False	0.800
URL Visible?	True/False	1.0			
Protocol Visible?	True/False	0.985	*19 categories*	–	*0.855*

Table 7: **Initial Browser URL Indicator Coding**—21 categories describing URL bars and security-relevant features were established by three expert coders. 2,882 screenshots were each independently coded by at two coders, resulting in an average Cohen's kappa of 0.855. Inconsistencies were identified and manually resolved.