What is this URL's Destination? Empirical Evaluation of Users' URL Reading

Sara Albakry
University of Edinburgh
Umm Al-Qura University
sara.albakry@ed.ac.uk

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

Maria K. Wolters
University of Edinburgh
Edinburgh, UK
maria.wolters@ed.ac.uk

ABSTRACT

Common anti-phishing advice tells users to mouse over links, look at the URL, and compare to the expected destination, implicitly assuming that they are able to read the URL. To test this assumption, we conducted a survey with 1929 participants recruited from the Amazon Mechanical Turk and Prolific Academic platforms. Participants were shown 23 URLs with various URL structures. For each URL, participants were asked via a multiple choice question where the URL would lead and how safe they feel clicking on it would be. Using latent class analysis, participants were stratified by self-reported technology use. Participants were strongly biased towards answering that the URL would lead to the website of the organization whose name appeared in the URL, regardless of its position in the URL structure. The group with the highest technology use was only minorly better at URL reading.

Author Keywords

Uniform Resource Locators; web literacy; URL readability; link destination; online security; technology usage; phishing

CCS Concepts

•Security and privacy → Usability in security and privacy; •Human-centered computing → Usability testing; Hypertext / hypermedia; Empirical studies in HCI; •Social and professional topics → Computing literacy;

INTRODUCTION

Malicious web links embedded in emails and other communications continue to plague companies resulting in compromises and lost revenue. FBI's Internet Crime Report estimates that phishing loses exceeded $29 million in 2017 for US organizations [40]. The Ponemon Institute estimates phishing costs UK organizations an average of $2.01 million per incident [35].

Automatic phishing detection, which is used by most organizations, is the most straight forward solution allowing organizations to detect and remove obviously malicious communication before it reaches users. Browsers also automatically block and provide warnings when they are confident that a URL is phishing [13]. Unfortunately, automatic detection is not perfect, sometimes allowing through malicious links or blocking benign ones [41]. Automatic detection systems also have difficulty identifying targeted communications which are carefully crafted and sent to a single target, known as spear phishing. In 2017, Google and Facebook were both tricked into paying $100 million to a scammer who was impersonating a manufacturer with whom the two companies interact [18].

To handle the fact that some malicious communications get through filters, security experts turn to users as the last line of defense, providing them with training and expecting them to identify phishing attacks, which they are not necessarily good at [14, 15]. Properly training people to detect phishing is also possibly more expensive than it is worth [21]. Knowing what advice to even train users with is also tricky. When security experts were asked to provide advice to internet users, "Don't click on dangerous links" and "Check the URL for an expected site" were common pieces of advice [37]. Both pieces of advice are based on the assumption that if the user pays close attention to the link text, they will be able to determine that it goes to a different website than what the accompanying message claims. The complexity of both the URL and human language processing systems along with the fact that phishers use URLs that contain brand names in different parts of the URL string [34], suggests that users may have trouble with this type of prediction. Hence, a systematic empirical evaluation is critical to form a clear understanding of users' URL reading abilities and to adapt our user-facing approaches accordingly.

In this work, we hypothesize that the majority of web users cannot differentiate between the following two Uniform Resource Locators (URLs): `https://facebook.profile.com` and `https://profile.facebook.com`. We take a slight twist on traditional anti-phishing research. Instead of measuring peoples' ability to identify phishing links, we focus on their ability to predict where a URL is likely to lead. To do so, we designed an online survey where participants were shown 23 URLs with a range of structures. For each URL, the participant was asked via a multiple choice question where the URL will lead and how safe they felt it was to click on, if it was sent from someone they know.

Research Questions

We focus on two high level research questions: ability to read a URL and assessment of the safety of a URL.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ISBN 978-1-4503-6708-0/20/04…15.00

DOI: http://dx.doi.org/10.1145/3313831.3376168

RQ1 Can users accurately predict where a URL will go?

RQ1.1 Can users correctly infer from the URL that it will go to the website of the organization listed in the domain position rather than the subdomain, and what factors affect prediction accuracy?

RQ1.2 Can users recognize that the end destination of shortened URLs is not easy to predict?

RQ1.3 Can users recognize the end destination of complex URL structures?

RQ2 What effects users' assessment of the likely safety of a URL?

To answer our research questions, we conducted an online survey with 1929 participants from both Amazon Mechanical Turk (AMT) [1] and Prolific Acadmic (PA) [2], implemented on LimeSurvey[3], an open source survey tool.

We find that indeed users have difficulty predicting the final destination of URLs. 32.9% of all participants always selected the organization name in the URL regardless of its position. In particular, they struggle to differentiate between a situation where an organization name they recognize is in the subdomain versus the domain, with only 8.3% of our participants able to reliably differentiate. Safety perceptions were also strongly tied to whether the participant thought that the URL would lead to a website that sounded like an organisation name or not.

BACKGROUND

In this paper, we focus on situations where a user is being asked to click on a URL which may or may not be malicious, and the information available to the user in advance of clicking, since once a user has opened a page, they may already be the victim of a malware attack.

URL Structure

Uniform Resource Locators (URLs) are a standardized format for describing the location and access method of resources via the internet [8]. They form a key mechanism for navigating the web and sharing online resources. To handle the breadth of possible addressing, a URL structure has many components that allow it to be flexible and accommodate all sorts of addressing situations. Figure 1 shows the same BBC URL written in two different ways with its component parts highlighted. Both URLs will lead to the same page, though one should generate a phishing warning on most modern browsers. The <user>:<password> and <port> components are valid parts of a URL, though they are not commonly used in user-facing URLs.

In this work, we focus on the <host> component, as it controls the destination computer which the browser will first try and contact. The <host> is further broken up using the '.' character into subdomains, the domain, and the top level domains as such: <subdomain>.<domain>.<top level domain>. Broadly, the host in a URL is read similar to a postal address where the item on the right is the top part of the hierarchy and the item on the left is the most precise. So `facebook.mobile.com` has a top level domain of 'com', a domain of 'mobile', and a subdomain of 'facebook'. When a browser tries to visit this URL, it will first contact the computer associated with 'com' to do a lookup for 'mobile', and then contact 'mobile' to do an internal lookup for 'facebook'. So `mobile.com` gets to decide what 'facebook' means in its local context. This is why the hosts `image.slashdot.org` and `image.google.com` are in very different places despite both URLs having the word "image".

The difference between subdomain, domain, and top level domain is quite important. The domain typically lists the organization's name, while the top level domain indicates the type of organization being contacted and sometimes what country they are in, and the subdomains typically indicate a section of the site. One of the more famous examples of a top level domain mattering is: `http://whitehouse.gov` (official US White House website) vs. `http://whitehouse.com` (porn site). The former has a top level domain of '.gov' indicating a government site, the latter has a '.com' top level domain indicating a commercial site.

On the other hand, URL shortening services have short domain names, such as `bit.ly`, and will provide a short unique alpha numeric string which redirects to another URL which is typically longer. For example, when asked, bit.ly maps `https://google.com` to `http://bit.ly/19BiSHW`.

Phishing

Phishing is the practice of sending email, or other communications, with the goal of deceiving the user into clicking on something they shouldn't or giving away valuable data. An in-depth overview of the topic of phishing is beyond the scope of this paper; we refer the interested reader to Hong's overview of phishing [22], and two overview papers on why phishing works [4, 14]. In this section we focus on work which looks at deceiving people with malicious URLs. Numerous studies have shown that people have trouble accurately identifying phishing URLs [14, 31, 11, 15]. Users are generally more likely to draw clues from the page that has been loaded than from the URL or other security indicators [14, 4, 28, 2].

URL Manipulation Techniques

A common phishing tactic is to send a communication claiming that it is vitally important that the user perform an action using the provided URL, such as logging into their bank. The communication claims to be from their bank, but on inspection, the URL does not lead to the bank's actual website. For example, the email could report to be from Bank of the West, but the URL might be `https://bankofthewest.foobar.com`. When the user clicks the link they are taken to a malicious website which may attempt to infect their computer or mimic a real login site and ask for the user's banking credentials.

Phishers typically obscure malicious URLs to make them look safe using several approaches. In particular, malicious actors have been shown to use shortened URLs [30, 12, 20, 32, 19, 5], often also receiving high click-through rates. Chhabra et al. found that URL shorteners are often used in social media

[1] https://www.mturk.com/
[2] https://www.prolific.ac/
[3] https://www.limesurvey.org/

Figure 1. URL scheme. The example top URL contains all the components including the <user>, <password> and <port> which are rare in publicly visible URLs. The bottom URL links to the same location, but is a more typical example of what an average user might see.

phishing and spam attacks, especially on Twitter, to hide the true identity of the phisher [11]. Since users are accustomed to seeing and clicking on short URLs, it is unclear whether they are aware of the security and privacy risks associated with them or not. Possible risks include long-term tracking cookies, phishing, spamming, drive-by-download attacks, and link hijacking. Short URLs have been widely explored from a systems perspective [20] with a focus on analyzing their redirection chain and click-through traffic. However, few studies have explored user awareness of the potential risks. Le-Khac et. al surveyed 100 internet users and found that there was a moderate to high awareness of their danger despite their tendency to propagate short URLs [27].

User-Facing solutions

We are not the first to observe that people have difficulty reading URLs. Tim Berners-Lee, author of the RFC on URL structure [8], once commented about URLs: "I regret that the syntax is so clumsy. I would like `http://www.example.com/foo/bar/baz` to be just written `http:com/example/foo/bar/baz`....But it is too late now." [7]. Most anti-phishing training contains some statement about how to handle URLs in electronic communications. Typical advice includes: "Don't click on links", "Type links in manually", "Avoid links with IP addresses", and "Google for the page instead" [39, 36].

A study by Gavett et al. looked at phishing susceptibility between younger and older adults [15]. They asked participants to visit a set of sites including sites with phishing URLs like `http://www.amazon.jigdee.com`. Participants were initially only shown the URL before clicking and the pages were direct copies of the real ones. Only older adults refused to log into the phishing pages, and even then it was only a small percentage. The vast majority of participants logged into the pages without any concerns. Gavett also found that one of the best predictors of phishing susceptibility was prior experience with a phishing attack.

User Training

Anti Phishing Phil by Sheng et al. [39] used a gamification approach where the player is shown several URLs and has to identify which ones are phishing/fraudulent. Between levels, the player gets feedback on their mistakes and how they could have identified the difference.

NoPhish [9, 10] trained people to identify phishing URLs through a mobile app. Unlike Anti Phishing Phil, the app includes URL structures such as subdomains in their training, with good results.

PhishGuru by Kumaraguru et al. [26] used web comics which appear after a user has clicked on a malicious URL. The work targets "teachable moments" when people are most receptive to anti-phishing training and advised people to simply not click on links in emails at all and instead type the links in. The "teachable moments" concept is now in active use by many anti-phishing training organizations including: `nowbe4.com`, `phishlabs.com`, and `wombatsecurity.com`.

Post-Click Support

Domain highlighting [28] is the practice of only showing the domain, or graying out the non-domain elements of the text in the URL bar of the browser. Domain highlighting has been adopted by several browsers including Chrome and Safari. The approach is intended to help users correctly discern the domain component of the URL and not be confused by other parts of the URL string such as a subdomain or file path. While promising, the approach has one sizable flaw, the user must first load a webpage before the highlighting is shown. As a result, the user must first load the potentially dangerous page, and then remember to check the URL after it has loaded, which is not a common behavior [14].

METHODS

To our knowledge, no large scale study has been conducted that definitively proves that people have difficulty predicting where a URL will lead. Even though it might sound intuitive for researchers who have a first hand experience with user-URL interactions that people find it difficult to read URLs, much of the security advice given online, such as "Don't click on dangerous links", assumes that end-users can read URLs with sufficient skill to be able to detect problematic instances.

In 2014 one of the authors gave out a "fun" worksheet to their Computer Security class which presented a set of increasingly complex URL structures and asked the students to identify elements such as the domain and subdomains. Surprisingly, a large section of the class struggled to answer the question for even simple URLs. Follow-on versions of the worksheet were tried out with a Human Computer Interaction course and a group of Masters students studying Usable Security. In all cases, the students struggled with basic URL reading tasks, such as deciding if `https://facebook.mobile.com` went to Facebook or to Mobile. The outcome of these exercises was used as a starting point for the survey design.

Prestudies

From prior work we know that the recognizability of an organization's name can impact opinions of trust and safety [14]. To control for this issue, we ran a prestudy to find a set of organizations that were familiar and that were unfamiliar to our target populations. We brainstormed a list of globally well-known organizations and used online lists of start ups and small newspapers to find the lesser-known organizations. We surveyed participants on both AMT (n=50) and PA (n=50) to be sure both target populations had similar opinions about the organization recognizability. We then selected the most and

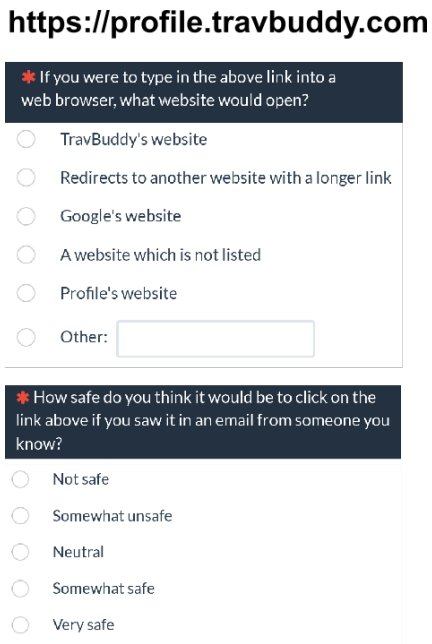

Figure 2. Example of one of the URL subdomain questions. Participants are shown the URL at the top and then asked a prediction and safety question about it. The questions are identical for all URLs, but the provided answers to the first question change based on the URL content.

least recognizable for use in the main the survey. The survey took an average of 1.5 minutes (AMT) and 2.5 minutes (PA) to complete. Participants were compensated $0.50 (AMT), €0.50 (PA).

We also conducted a second pre-study to test the main survey design, with nearly identical wording and structure to what is described in Section 3.2 below. We used the pre-study to trial some wording variations and to make sure that there were no unexpected differences between how the AMT and PA populations interacted with the survey. We sampled 100 participants each from AMT and PA. The survey took 9 min (AMT), 7 min (PA) to complete on average and participants were compensated $1.50 (AMT), €1.00 (PA).

Survey Instrument

The survey had three sections: introduction, URL questions, and demographics which were presented to participants in that order. We detail key parts of the survey below and provide a full version in [3].

Section 1: Consent Form and Task Clarity Checkup

The survey started with a consent form and a set of instructions asking participants to answer questions based on reading the URLs and not typing them into a browser. To ensure that the participants had read the instructions, we asked them: "What do you need to do in this survey?". The correct answer was "Read links and predict where they will go", with the other options being typing in links and clicking on links. Participants who answered incorrectly were not allowed to progress till they provided the correct answer. The consent form and instructions avoided any mention of privacy or security.

Section 2: 23 Randomised URL Blocks (2 Questions/URL)

Participants were shown 23 URL question blocks in random order. Each block was presented on its own page with an image of the URL on the top followed first by a destination prediction question and then by a safety question (Figure 2).

Q1. URL Destination (RQ1): This question asked: *"If you were to type in the above link into a web browser, what website would open?"* The provided answers were in a randomized order and include two parts of the URL (typically the subdomain and domain), a distractor answer (i.e. Google), *"redirects to another website with a longer link"*, *"a website which is not listed"*, and an *"other"* option where the participant could provide a free-text response. In cases where a second part of the URL did not exist (i.e. `microsoft.com`) a second distractor answer was used instead.

Q2. URL Safety (RQ2): This question asked: *"How safe do you think it would be to click on the link above if you saw it in an email from someone you know?"* Answer options were a 5-level likert scale ranging from "Not safe" to "Very safe".

Wording for the safety question was highly debated as perceived safety is dependent on context, not just the URL text. We settled on the above wording using the second prestudy described above by comparing the above wording with the alternative: *"Would you click on the link above if you saw it in an email from someone you know?"* Both versions provided a spread of answers, but participants seemed more willing to answer with strongly agree or disagree to the safety question, while the clicking question had a narrower range of answers with neutral being selected frequently.

Section 3: Demographics Questions

Demographics had 12 questions listed on a single page asking about participants' gender, age, native language, level of education, computing devices used, how often they visit different types of websites (social media, financial, online games, news, company), how often they ask for help with computers and how often others ask them, Westin Privacy index (Section 2.6, [25]), web development or system administrator experience, and optionally free-text comments.

We also asked them to report if they had typed in any of the URLs during the length of the study, making it clear that answers would not impact payment. Comments in the second prestudy described above suggested that taking the survey was causing people to question their initial URL reading strategy and switch strategies mid-survey. In the final survey we asked them what strategy they were using at the beginning of the survey and what strategy they used at the end. Provided answers were drawn from our own experience and from comments made by prestudy participants.

URLs Tested

Each participant saw 23 URLs representing four URL structures which align with RQ1 sub questions. The full set of URLs used is listed in Table 1.

Controlling Confounds

We wanted to test peoples' ability to parse URLs in best case situations. Therefore, we intentionally avoided most of the

URL Structure	Orgnization Industry	Orgnization Recognizablity	Organization Name	URL	
				Group 1	Group 2
Domain Only			Microsoft	https://microsoft.com	
			Google	https://google.com	
			AMT	https://mturk.com (AMT participants only)	
			PA	https://prolific.ac (PA participants only)	
Single Subdomain	Social	Well known	Facebook	https://facebook.profile.com	https://profile.facebook.com
			Twitter	https://mobile.twitter.com	https://twitter.mobile.com
		Less known	Travbuddy	https://profile.travbuddy.com	https://travbuddy.profile.com
			Weheartit	https://weheartit.mobile.com	https://mobile.weheartit.com
	News	Well known	BBC	https://bbc.profile.com	https://profile.bbc.com
			CNN	https://mobile.cnn.com	https://cnn.mobile.com
		Less known	Dunfermlinepress	https://profile.dunfermlinepress.com	https://dunfermlinepress.profile.com
			Haysfreepress	https://haysfreepress.mobile.com	https://mobile.haysfreepress.com
	Financial	Well known	Paypal	https://paypal.profile.com	https://profile.paypal.com
			Western Union	https://mobile.westernunion.com	https://westernunion.mobile.com
		Less known	Purepoint	https://profile.purepoint.com	https://purepoint.profile.com
			Revolut	https://revolut.mobile.com	https://mobile.revolut.com
Short		Well known	Bit.ly	https://bit.ly/1bdDlXc	
			Goo.gl	https://goo.gl/fJOIAv	
		Less known	Po.st	https://po.st/If6RgX	
			U.to	https://u.to/SbwC	
Complex			Google	https://facebook.com@google.com	
			Twitter	https://twitter.com/facebook.com	
			Facebook	https://facebook.com/picture.html?a=twitter.com	
			Facebook	https://facebook.com/?url=twitter	

Table 1. URLs used in the study. For the single subdomain questions participants were divided into two groups to ensure that each participant only saw each company name once while still seeing both order combinations.

phishing tricks detailed in Section 2.2 and made the following simplifying decisions:

Https Protocol Thanks to the work of groups like Let's Encrypt, any site owner, including phishers, can get a valid security certificate and have "https" in their URL [34]. We wanted participants to answer the safety questions based on the content of the URL not the absence of the https protocol, so all URLs in the study start with https.

End in .com We limited all URLs to ones ending in .com, with the exception of the URL shorteners. The .com top level domain is recognizable by consumers, finding companies who have a .com domain is easy, and using only one top-level domain limits study confounds.

Real URLs We base all the URLs on real organizations' URLs. While we ask participants not to type in URLs, some percentage are likely to search the Internet for them. This decision ensures that any URL found through searching will match, at least in part, the one we present.

Recognizable Names We only used organizations which have their commonly used name in their URL. For example, CNN was included in the study as their URL (`https://cnn.com`) is easily identifiable with CNN. The New York Times was not considered because their URL (`https://nyt.com`) uses an abbreviation which is not trivially identifiable with the organization.

URL Structures

The use of multiple structures was to test a variety of situations, ensure that participants were seeing a good mix of different URLs, and prevent habituation to a single URL structure or answer shape.

Domain Only (Baseline) URLs in this category have no subdomain and are the most simple. We included three URLs which participants should be able to easily recognize: `https://google.com`, `https://microsoft.com`, and depending on the recruitment platform, either `https://mturk.com` (AMT) or `https://prolific.ac` (PA).

The Microsoft and Google URL questions did not contain the company name as one of the multiple choice options, participants were instead expected to answer "A website which is not listed" or "other". The company names were not included for two reasons: 1) to ensure that all answer options were used at least twice, and 2) the prestudies and work with students suggested that both URLs are fairly easy to predict even when students were asked to provide free-text answers. The AMT and PA URL questions did contain the organization names within the answer options.

Single Subdomain (RQ1.1) These URLs tested our hypothesis that people "read" URLs by looking for a recognizable word regardless of whether it is positioned in the domain or subdomain. Their form was: `https://<subdomain>.<domain>.com`.

We therefore designed a set of URLs varying: the location of the organization name in a URL (subdomain, domain), filler word for the other position (mobile, profile), how recognizable the organization name is (well known, less known), and the category of the organization (social media, news, financial). The result was 24 URLs, with each organization name appearing twice, once in the domain and once in the subdomain position. To ensure that each participant saw each organization name only once, we divided the subdomain questions into two groups as shown in Table 1. Each participant was randomly assigned to group 1 (n=984) or

	MTurk		Prolific		All Participants	
Gender						
Male	543	(55.9%)	465	(48.5%)	1008	(52.3%)
Female	423	(43.6%)	486	(50.7%)	909	(47.1%)
No Answer	3	(0.3%)	1	(0.1%)	4	(0.2%)
Other	2	(0.2%)	6	(0.6%)	8	(0.4%)
Education						
Some High School	6	(0.6%)	22	(2.3%)	28	(1.5%)
High School	107	(11.0%)	141	(14.7%)	248	(12.9%)
Some College	285	(29.4%)	210	(21.9%)	495	(25.7%)
College	446	(45.9%)	359	(37.5%)	805	(41.7%)
Postgrad	123	(12.7%)	220	(23.0%)	343	(17.8%)
Other	4	(0.4%)	6	(0.6%)	10	(0.5%)
Native Language						
English	915	(94.2%)	686	(71.6%)	1601	(83.0%)
Other Languages	56	(5.8%)	272	(28.4%)	328	(17.0%)
Westin Index						
Fundamentals	282	(29.0%)	239	(24.9%)	521	(27.0%)
Pragmatic	622	(64.1%)	676	(70.6%)	1298	(67.3%)
Unconcerned	67	(6.9%)	43	(4.5%)	110	(5.7%)
Web Development						
No Experience	760	(78.3%)	734	(76.6%)	1494	(77.4%)
Experience	211	(21.7%)	224	(23.4%)	435	(22.6%)
Asking for tech help						
I ask others	49	(5.0%)	40	(4.2%)	89	(4.6%)
Others ask me	456	(47.0%)	447	(46.7%)	903	(46.8%)
Technology Use						
Desktop	648	(66.7%)	567	(59.2%)	1215	(63.0%)
Laptop	791	(81.5%)	804	(83.9%)	1595	(82.7%)
Tablet	529	(54.5%)	511	(53.3%)	1040	(53.9%)
Smartphone	907	(93.4%)	903	(94.3%)	1810	(93.8%)
Website Use						
Company	703	(72.4%)	676	(70.6%)	1379	(71.5%)
Financial	806	(83.0%)	743	(77.6%)	1549	(80.3%)
Game	504	(51.9%)	416	(43.4%)	920	(47.7%)
News	789	(81.3%)	818	(85.4%)	1607	(83.3%)
Social Media	853	(87.8%)	833	(87.0%)	1686	(87.4%)

Table 2. Demographic information. Numbers indicate either the number of participants who provided that answer or, for Likerts, the number who indicated "agree" or "strongly agree", or for usage frequency, "daily" or "weekly". Percentages are taken from the total number of participants from that platform.

group 2 (n=950) and only saw the URLs associated with their group.

Short URLs (RQ1.2) The majority of short URLs have the name of the shortening service in the <host> position and a unique alpha numeric string in the <query-string> location, for example: `https://bit.ly/1bdDlXc` will redirect to `http://www.google.com`.

We included four short URLs, two from commonly used shortening services, and two from relatively unknown shortening services. All participants were exposed to all four shortened URLs.

Complex URLs (RQ1.3) Complex URLs contained two well known organization names in the URL string and use more complex structures. These questions had three purposes: 1) see how participants handled having two recognizable names in the URL, 2) determine if users can differentiate between the <host>, <user>, <url-path>, and <query-string> elements, and 3) identify participants that have advanced URL reading skills.

Recruitment

Participants were recruited from Amazon Mechanical Turk (AMT) and Prolific Academic (PA), both of which are regularly used for advertising academic surveys [24, 33]. Using both also resulted in a more representative sample as AMT is primarily North Americans and PA is primarily European.

The survey was advertised as "Opinions on Weblinks" with an estimated time of 15 minutes and compensation of $1.75 (AMT), €1.50 (PA). All advertisement materials avoided priming words like "security" or "privacy". The compensation amounts were selected to be roughly equal based on the exchange rates [4] at the time of the study. Advertised time estimates were calculated based on the time required by slower pre-study participants to ensure that most participants would finish within the advertised time.

We screened participants to ensure they were 18 or above, had not taken our pre-study, and had a minimum approval rate of 95% on AMT and 90% on PA. Both settings produced high-quality pre-study answers. We posted tasks on AMT and PA in multiple batches at different times of the day and days of the week to ensure we were targeting a range of potential respondents. The survey was conducted in early April 2017. The study complied with the Ethics procedure of the School of Informatics, University of Edinburgh.

Participants

A total of 2030 respondents completed the survey, 1016 (50%) from AMT and 1014 (50%) from PA. After data cleaning and coding of "other" answers, we excluded 101 (5%) participants. Exclusions were done for the following three reasons. 1) Incorrect answers to the AMT or PA URL questions (attention check, n=62). 2) The distractor answer was selected on more than two of the subdomain or domain questions (n=14). Complex and short URL questions were not used for exclusion because they are both harder to answer correctly and more open to interpretation. 3) There was a temporary issue with the server which provided the URL images. Any participant mentioning issues with seeing the URL images was excluded (n=25). Three people declared themselves to be below 18 years of age; given that all of these ages were below 10, we regard these as typographical errors or intentional false replies and include the participants.

There were no differences between included and excluded participants in age (Wilcoxon test, $p<0.5$), education (Fisher test, $p<0.8$), gender (Fisher test, $p<0.6$), or survey platform (Fisher test, $p<0.3$).

Of the final 1929 participants, 52.3% were male. The age range was 18-100 years (the second highest age was 77) with a median of 32 years and an average of 34.5 years (σ=11.5 years). Participants' native language was primarily English (AMT=915, 94.2%, PA=686, 71.6%). The next most popular languages were in order: Portuguese (n=45), Bulgarian (n=40), Spanish (n=26), Italian (n=23), and German (n=22). 1903 (98.7%) reported using either a desktop or laptop on at least a weekly basis. Detailed demographics can be found in Table 2.

Survey Analysis

Statistical Analysis: All statistical analyses were conducted in R. Differences in demographics between platforms and user groups were analysed using standard non-parametric tests as

[4] XE http://www.xe.com/

implemented in coin [23]. The Akaike Information Criterion and the Bayes Information Criterion were used for assessing model fit, and χ^2 tests for determining significant differences between two statistical models. For space reasons, model coefficients and model comparisons are not reported in full.

Establishing User Groups: Use of technology and self-perceived technology skills are covered by four sets items in the demographics section (c.f. Table 2). We used latent class analysis (PoLCA,[1, 29]) to reduce these items to a single variable, User Group, for further statistical analysis.

Correctness of Decisions: We used generalised linear mixed models (GLMMs, [16]) to investigate the effects of participant-level and URL-level characteristics on participants' ability to read a URL. Whereas in normal generalised linear models (e.g., linear or logistic regression), the model coefficients are the same for all participants, GLMMs allow some coefficients to vary by participant or participant group. The R packages used were lme4 [6] and arm [17].

Perceived URL Safety: Since perceived safety is an ordinal variable, we used proportional odds logistic regression (POLR [1]) to investigate the effects of decision correctness, URL characteristics, and user group on safety judgments.

Coding Free Text Answers: All the URL questions had a possible answer of "other" with a free text box. For two of the domain questions, the correct answer was not listed and participants were expected to either select the not listed option, or provide an answer of "other".

One researcher went through all the "other" answers and marked those that could be trivially mapped onto one of the existing answer options, and, in the case of the domain questions, whether the answer was correct. For example, a participant answered "other" for the URL `https://haysfreepress.mobile.com` stating "haysfreepress on the website mobile.com". From the comment it is clear that the participants understands that the URL will attempt to go to the mobile.com website, so their answer was re-mapped to the provided answer "Mobile's website". A second researcher then reviewed all decisions. Any disagreements were handled through discussion. Answers which were ambiguous or expressed confusion were left as "other". For example, for `https://weheartit.mobile.com`, a participant responded "music website". Henceforward, we use the term *other* to refer to answers where the user both indicated the "other" option, and provided an answer which could not be trivially mapped to an existing answer option using the above procedure. Excluding the single domain answers where "other" was the expected response, 188 "other" answers were re-mapped onto other codes, 114 of which were re-mapped as subdomain.

Limitations

Our work recruited participants from the online crowd sourcing platforms AMT and PA. Research on AMT workers has shown them to generally be more privacy conscious and more technically skilled than average Americans [24, 38]. Participants on both platforms are also likely professional survey takers with extensive experience answering common survey questions such as Westin's Privacy Index. The above-average technical skill and experience with the internet also likely impacted our results as participants would have been more familiar with concepts like links and safety. However, we argue that while not representative, these participants represent a best case scenario and that we would expect the general population to have less knowledge of URL reading.

Repeated use of the filler words "profile" and "mobile" could have caused participants to notice the repetition and change their answers. However, we feel that the impact was minimal since the repeated words could be interpreted as either the correct or wrong answer. Our results below also show that people struggle to read the subdomain URLs correctly, so it is unlikely that the repetition improved accuracy.

Our study focuses on simple URL reading rules that we might reasonably expect the general public to know. This was intentional, as we wanted to create a baseline for future work. There are likely many more factors that impact a user's ability to read a URL, such as more complex URL structures.

Participants may have had a range of interpretations of the term "safety" in the second question asked for each URL. Safety in URL reading is often contextual in ways that are challenging to replicate in a survey. While we tried to give minimal context, it is still very likely that participants interpreted "safety" in different ways ranging from computer security to privacy.

RESULTS

User Groups

We compared the fit of polynomial latent class models for 2–8 classes. For each class, we determined the best model out of a sample of 100, to avoid local minima. 1000 iterations proved sufficient for parameter estimation. The most parsimonious model consisted of three groups of users. In the following, we will call them mobile users, desktop users, and power users.

Mobile Users ($n = 761, 39.3\%$) have the least coding experience (8.0%). They use laptops (92.3%) and smartphones (90.3%) every day, and are least likely to be asked for help. *Desktop Users ($n = 367, 19.0\%$)* use desktops every day (95.6%), but only 3.8% use laptops regularly. They are also the least likely to use a smartphone every day (77.4%). Finally, *Power Users ($n = 806, 41.7\%$)* are most likely to have coding or web development experience (37.8%), most likely to be asked for help, and the heaviest smartphone (98.5% daily) and social media (84.9% daily) usage.

The three categories have distinct gender and educational profiles. 60.6% of mobile users are female, whereas desktop users tend to be male (54.0%), and power users are mostly male (64.4%). 59.7% of mobile users and 64.9% of power users have at least a graduate or postgraduate degree, compared to only 47.1% of desktop users. The three classes do not differ by Westin privacy type (χ^2 test, $\chi^2(4) = 6.35, p < 0.18$) or age (Kruskal-Wallis test, $\chi^2(2) = 3.8, p < 0.15$).

Single Domain URLs

Participants were shown three URLs which contained only a domain (Table 1). Here, we report the results for Microsoft and Google. The most obvious correct answer was

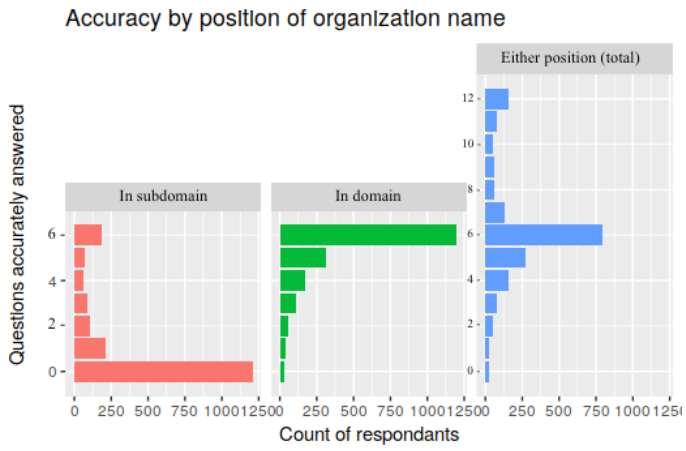

Figure 3. Number of single subdomain questions each respondent correctly answered when the organization was in the subdomain (left, out of 6), domain (middle, out of 6), or in total (right, out of 12).

"not listed". "Redirect" was also acceptable, given that both `https://microsoft.com` and `https://google.com` often redirect to `https://www.microsoft.com` and `https://www.google.com`.

As expected, both questions were fairly easy to answer with 1793 (92.7%) answering both by either indicating the domain, or by indicating that the correct destination was not listed.

Single Subdomain URLs

The 12 URLs used for this part of the study are summarised in Table 1. On average participants answered 6.4 ($\pm$ 2.6) out of 12 of these questions correctly. Looking at the effect of the position of the organization name in the URL, we find that when the organization was in the domain (i.e. `https://profile.facebook.com`), participants generally answered correctly (5.2$\pm$1.4 out of 6), and when it was in the subdomain, they did not (1.2$\pm$2 out of 6).

Of the 10840 total incorrect answers provided, 9418 (86.9%) indicated that the URL led to the subdomain. 7.9% ($n = 851$) indicated that the correct answer was not listed, and 3.5% ($n = 384$) thought that the URL redirected.

While answers other than the domain are technically inaccurate, an argument could be made that thinking the URL redirects or is not listed is safer than thinking the URL leads to the subdomain. Selecting "redirect" or "not listed" was also somewhat common with 32% (n=618) of respondents providing at least one such answer. If we considered these answers correct, then respondents would still only average 7 $\pm$ 2.5 out of 12 correct.

Factors Influencing Participant Judgements.
Using GLMMs, we modelled the effect of three variables on the correctness of participant judgements: position of organization name (domain versus subdomain), organization name, and user group. To estimate the effect of adding each term, we compared models with and without the term. The models assessed and their AIC values are listed in Table 4. The key predictors of correct responses are position (model B versus model A: $\chi^2(1) = 10885, p < 0.0001$) and the interaction between position and organization (model E versus model B: $\chi^2(22) = 194.72, p < 0.0001$). The model is further improved by letting the position effect vary by user group (model F versus model E: $\chi^2(2) = 31.675, p < 0.0001$).

User Group	Position	Correct N	Correct %
Mobile User	Subdomain	727	15.9%
	Domain	3859	84.5%
	Total	4586	50.2%
Desktop User	Subdomain	441	20.0%
	Domain	1942	88.2%
	Total	2383	54.1%
Power User	Subdomain	1226	25.4%
	Domain	4173	86.3%
	Total	5399	55.8%

Table 3. Correct responses to single subdomain questions by position of organization name in URL by User Group.

ID	Model	Df	AIC
A	1 + (1 | User Group)	3	32027
B	1 + Position + (1|UG)	3	21144
C	1 + Organization + (1|UG)	13	32026
D	1 + Pos. + Org. + (1|UG)	14	21147
E	1 + Pos. x Org. + (1|UG)	25	20993
F	1 + Pos. x Org. + (1 + Pos.|UG)	27	20966

Table 4. Logistic Mixed Regression Models for the Single Subdomain data. Position = organization name in domain versus subdomain, Organization = organization name, 1 = intercept, x = includes interaction term. Df = degrees of freedom, AIC = Akaike Information Criterion. The model coefficients for the first set of variables are the same for all cases, the coefficients for the variables in brackets vary by user group.

The position of the organization name had the largest impact on judgement accuracy. Participants tended to assume that the URL led to the named organization's website, regardless of whether it was in the domain or in the subdomain, with 32.2% (n=623) always indicating that the URL went to the organization name in the URL, and only 8.3% (n=161) always indicating the correct answer of domain.

The effect of organisation name was moderated by the position of that name in the URL. For Twitter, CNN, and Western Union, participants were more likely to be correct if the name was in the domain position and less if it was in the subdomain.

Finally, there is a difference in accuracy between user groups. Power users are overall more accurate than desktop users, who are in turn more accurate than mobile users (Table 3). Power users are also more likely to correctly read URLs where the organisation name is in the subdomain. However, overall, performance is still relatively poor for all three groups.

Short URLs

For short URLs, the "correct" answer is challenging to determine since Redirect (the URL will redirect to a different website), Not Listed (the target URL is not listed in the responses), and domain name (`bit`, `goo`, `po`, or `u`) are all plausibly correct. 93.4% of all answers were for one of these options.

Factors Influencing Participant Judgements.
For this analysis, we only consider the answers Redirect and Not Listed, since both answers reflect that it is impossible to determine the exact final destination from the URL. Using GLMM, we constructed three models, one looking at Redirect judgements, one looking at Not Listed judgements, and one combining both judgements into a single response variable.

Overall, shortener type (Well Known or Less Known) did not influence whether the answer was redirect or not listed. However, participants are more likely to identify `bit.ly` links as redirecting, and other link shorteners as not listed. In general, mobile users are less likely to choose one of the two options Redirect / Not Listed than desktop or power users.

Complex URLs

The `https://facebook.com@google.com` URL was confusing for participants. They were split on if it would go to a site not listed (30.1%), Facebook (28.3%), Google (14.7%), Redirect (10.8%), or Other (16.1%). The correct answer is that it would go to Google and attempt to login as the user "facebook.com", likely generating a warning on most modern browsers.

The 16.1% who wrote in responses strongly felt that this URL would result in an error due to it either not being a real URL, or because there is no such website. As one participant put it: "almost looks like an email attached to a web link."

Respondents did somewhat better on the `https://twitter.com/facebook.com` question, with 62.2% correctly answering Twitter and very few thinking that it would lead to Facebook (6.2%). 16.5% were likely to indicate that the URL led to a page that was not listed, while 8.4% answered "other" (8.4%). These responses typically indicated that the URL wasn't real, would result in an error page, or simply couldn't exist: "It wouldn't work because there's no URL that includes both Twitter and Facebook."

Perceived Safety

For analysis purposes, the ratings of perceived safety were converted to a numerical scale, with 1 corresponding to "Not Safe", and 5 corresponding to "Very Safe". Overall, participants rated Single Subdomain URLs as somewhat safe (median: 4, inter quartile range: 3–4), and Shortened URLs as somewhat unsafe (median: 2, inter quartile range: 1–3).

Single Subdomain.
The best fitting model for safety judgments for Single Subdomain URLs, with an AIC of 63385, consists of the variables Correctness (whether the correct destination was given), Position, Organization, all two-way interactions, and the three-way interaction between all terms. Overall, participants are more likely to rate a URL safe if they think that it goes to the organization (median: 4; inter quartile range: 3–4 if the organization is in the subdomain, and 3–5 if the organization is in the domain). There are strong company-specific effects. For example, as Figure 4 shows, URLs involving PayPal are generally regarded as somewhat unsafe when participants do not think that the URL will lead them to PayPal, whereas their median rating of URLs involving Western Union, another well known payment provider, is neutral in the same context. The differences between user groups are significant (Kruskal-Wallis test, $\chi^2(2) = 25.7, p < 0.0001$), but small. Mobile Users tend slightly more towards neutral ratings than other user groups.

Short URLs.
For Short URLs, the best fitting model, with an AIC of 20355, includes the variables Answer (for answer given), Shortener Type, and their two-way interaction. When participants choose one of the correct answers, Redirect or Not Listed, they are far more circumspect in their judgment. We also find that `bit.ly` and `goo.gl` are regarded as safer than `po.st` and `u.to`. There are no differences between user groups (Kruskal-Wallis test, $\chi^2(2) = 1.0, p < 0.62$).

SUMMARY OF FINDINGS

Being able to accurately compare a URL to an expected destination is an important skill needed in detecting fraudulent URLs, such as those sent in phishing communications. Our findings clearly illustrate that participants, even those with extensive computer experience, are unable to accurately predict the destinations of relatively simple and trick-free URLs. Instead of recognising that URLs typically go to the organization listed in the domain position, our participants tended to select the recognisable organization name in the URL, even if it was located in non-domain positions, such as the subdomain.

However, users were not necessarily consistent in their URL reading approaches, with only 32.2% consistently selecting the organization name and 8.1% consistently selecting the domain. Technology use impacted their ability to correctly read URLs, but the impact was fairly low, with mobile users correctly answering only 15.9% of the questions with the organization in the subdomain, and power users only 25.4%. The results are concerning as they indicate that even technically skilled users struggle to accurately predict the destination of a basic URL.

Participants were more likely to rate a URL as safe to click on if they thought that its destination was the organisation named in the URL. They were also more dubious of short URLs than the domain URLs, single subdomain URLs, and complex URLs. This result is somewhat good as it indicates that users can identify that the destinations of short URLs are more difficult to predict than the destinations of longer ones.

DISCUSSION AND IMPLICATIONS

Our findings illustrate that participants, even heavy technology users (power users), are unable to accurately predict the destination of a clearly written URL. Based on the "other" answers, as well as our own experience, users are "reading" URLs by looking for large familiar names and assuming that these are the true destinations. The tactic is not irrational, as the vast majority of URLs users likely encounter on a daily basis have only one large organization name and that name is the page destination. Large organizations also protect their web brand by buying or otherwise removing similar-looking domains, reducing the number of legitimate similar domains a user is likely to see. The result is that users believe that URLs will go to the large familiar looking word because it is the only possible destination. The observation has implications for

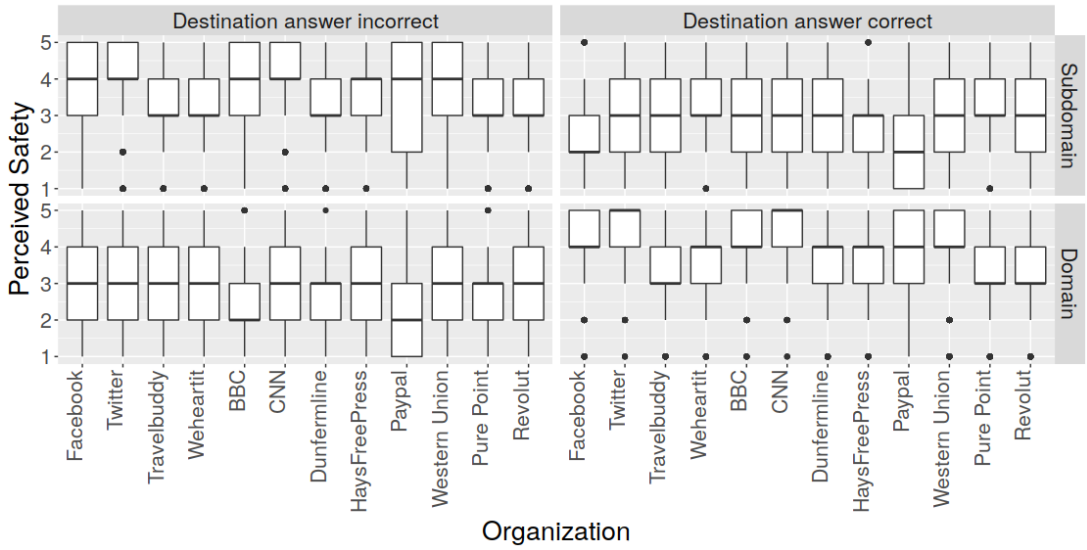

Figure 4. Perceived Safety of Single Subdomain URLs. 1= Not Safe, 5 = Very Safe. Subdomain/Domain: Position of the Organization Name.

URL	Facebook	Twitter	Google	Distractor	Redirect	Not listed	Other
`https://facebook.com@google.com`	569	-	**296**	0	217	606	325
`https://twitter.com/facebook.com`	125	**1252**	-	11	125	332	169
`https://facebook.com/?url=twitter`	**1527**	124	-	3	166	140	53
`https://facebook.com/picture.html?a=twitter.com`	**1636**	107	-	1	118	102	49

Table 5. Answers to questions involving URLs with more complex structures and two company names.

both how we present URLs to people, and the misconceptions that must be corrected during training. Simply telling users to look at the URL is not sufficient, they need training, or better, support, to be able to differentiate between the actual destination and distracting familiar words. It is also unclear how effective embedded approaches, such as domain highlighting, are at correcting misconceptions, especially in user groups like power users who feel confident at using the Internet or users of smartphones who may be unaware that they can view URLs.

We would like to challenge the reader to look past the problem of teaching users to correctly read a URL and instead focus on the problem of how to present users with information about a link's true destination. The two may sound similar but the problems are actually quite different. The question of where a URL leads is surprisingly complex. Even answering a basic question such as: "does `microsoftemail.com` belong to Microsoft?", is not easy even for a computer, and gets harder the less popular a website is. Examples like `windows.net` or `googleusercontent.com` are even more challenging to understand as both are hosting services owned by Microsoft and Google respectively, where clients pay to put their content. So the domains have valid certificates but the subdomains can contain user-controlled content, including phishing pages. To solve the phishing problem, we need to not just help users read a URL, but help them accurately understand who controls the content at the destination they are visiting. To accomplish that task, we need a stronger understanding of how people with all types of backgrounds think about website content ownership.

URL shortening services and redirects make URL reading even more complex. A URL might lead to a website owned by a known trusted organization, say Google, but then redirect to a malicious website. Our results show that people are generally aware that the destinations of short URLs are challenging to predict. But redirecting URLs are virtually impossible to identify from the URL itself, and even computers can't identify them without first visiting the URL's first destination.

Finally, the concept of safety when visiting a link is complex, contextual, and individual. Phishing or URLs that lead to malicious sites are considered bad by most people. But it is more challenging to automatically classify URLs that lead to websites that sell visitors' information to marketers, contain fake news, or contain content that a portion of the population might find offensive. For these situations, users need to be aware of where they are going and to be able to make individual decisions about how "safe" it is to go there. As mentioned earlier, knowing who controls the content on a page is valuable for users not only to compare the expected destination with the actual one, but also to be aware of who created the content they are ingesting and who might be recording information about the user.

ACKNOWLEDGMENTS

We want to thank the members of the TULiPS lab for their feedback on survey design. We also thank our anonymous reviewers for their insightful comments. This research is funded in part by the UK National Cyber Security Center and by a Google Research Award. MW acknowledges the Alan Turing Institute ((EPSRC, EP/N510129/1)).

REFERENCES

[1] Alan Agresti. 2002. *Categorical Data Analysis* (2 ed.). John Wiley, New York, NY.

[2] Aiping Xiong, Robert W. Proctor, Weining Yang and Ninghui Li. 2017. Is Domain Highlighting Actually Helpful in Identifying Phishing Web Pages? Aiping. *Human Factors* (2017). http://journals.sagepub.com/doi/pdf/10.1177/0018720816684064

[3] Sara Albakry, Kami Vaniea, and Maria Wolters. 2020. Opinions on Weblinks. (2020). DOI: http://dx.doi.org/10.7488/ds/2749

[4] Mohamed Alsharnouby, Furkan Alaca, and Sonia Chiasson. 2015. Why Phishing Still Works. *Int. J. Hum.-Comput. Stud.* 82, C (Oct. 2015), 69–82. DOI: http://dx.doi.org/10.1016/j.ijhcs.2015.05.005

[5] Demetris Antoniades and Thomas Karagiannis. 2011. we.b: The web of short URLs. In *e International World Wide Web Conference Committee (IW3C2)*. http://bit.ly/dv82ka.

[6] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. DOI:http://dx.doi.org/10.18637/jss.v067.i01

[7] Tim Berners-Lee. 2000. Frequently Asked Questions. *World Wide Web Consortium* (2000).

[8] T. Berners-Lee, L. Masinter, and M. McCahill. 1994. RFC1738: Uniform Resource Locators (URL). (December 1994). https://www.w3.org/Addressing/rfc1738.txt https://www.w3.org/Addressing/rfc1738.txt.

[9] Gamze Canova, Melanie Volkamer, Clemens Bergmann, and Roland Borza. 2014. NoPhish: an anti-phishing education app. In *International Workshop on Security and Trust Management*. Springer, 188–192.

[10] Gamze Canova, Melanie Volkamer, Clemens Bergmann, and Benjamin Reinheimer. 2015. NoPhish app evaluation: lab and retention study. In *NDSS workshop on usable security*.

[11] Sidharth Chhabra, Anupama Aggarwal, Fabricio Benevenuto, and Ponnurangam Kumaraguru. 2011a. Phi.Sh/$oCiaL: The Phishing Landscape Through Short URLs. In *Proceedings of the 8th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS '11)*. ACM, New York, NY, USA, 92–101. DOI: http://dx.doi.org/10.1145/2030376.2030387

[12] Sidharth Chhabra, Anupama Aggarwal, Fabricio Benevenuto, and Ponnurangam Kumaraguru. 2011b. Phi.sh/$oCiaL: The Phishing Landscape through Short URLs. In *CEAS*. ACM. http://www.barracudalabs.com/downloads/

[13] Chromium. 2017. Chromium Project: IDN in Google Chrome. (2017). https://www.chromium.org/developers/design-documents/idn-in-google-chrome Accessed 28 Aug, 2017.

[14] Rachna Dhamija, J Doug Tygar, and Marti Hearst. 2006. Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 581–590.

[15] Brandon E. Gavett, Rui Zhao, Samantha E. John, Cara A. Bussell, Jennifer R. Roberts, and Chuan Yue. 2017. Phishing suspiciousness in older and younger adults: The role of executive functioning. *PLOS ONE* 12, 2 (feb 2017), e0171620. DOI: http://dx.doi.org/10.1371/journal.pone.0171620

[16] Andrew Gelman and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge, UK.

[17] Andrew Gelman and Yu-Sung Su. 2018. *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models*. https://CRAN.R-project.org/package=arm R package version 1.10-1.

[18] Samuel Gibbs. 2017. Facebook and Google were conned out of $100m in phishing scheme | Technology | The Guardian. (2017). https://www.theguardian.com/technology/2017/apr/28/facebook-google-conned-100m-phishing-scheme

[19] Chris Grier, Kurt Thomas, Vern Paxson, and Michael Zhang. 2010. @spam: The Underground on 140 Characters or Less * General Terms. In *Proceedings of the 17th ACM conference on Computer and communications security*. ACM, 27–37. http://apiwiki.twitter.com/Twitter-API-Documentation

[20] Neha Gupta, Anupama Aggarwal, and Ponnurangam Kumaraguru. 2014. bit.ly/malicious: Deep Dive into Short URL based e-Crime Detection. In *eCrime Researchers Summit, eCrime*, Vol. 2014-Janua. 14–24. DOI:http://dx.doi.org/10.1109/ECRIME.2014.6963161

[21] Cormac Herley. 2009. So Long, And No Thanks for the Externalities: The rational rejection of security advice by users. In *Proceedings of NSPW'09*.

[22] Jason Hong. 2012. The State of Phishing Attacks. *Comm. ACM* 55, 1 (Jan. 2012), 74–81. DOI: http://dx.doi.org/10.1145/2063176.2063197

[23] Torsten Hothorn, Kurt Hornik, Mark van de Wiel, and Achim Zeileis. 2008. Implementing a class of permutation tests: The coin package. *Journal of Statistical Software* 28, 8 (2008), 1–23.

[24] Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. 2014. Privacy Attitudes of Mechanical Turk Workers and the U.S. Public. In *Symposium On Usable Privacy and Security (SOUPS '14)*. 37–49.

[25] Ponnurangam Kumaraguru and Lorrie Faith Cranor. 2005. *Privacy indexes: A survey of Westin's studies*. Technical Report. Carnegie Mellon University.

[26] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. 2008. Lessons From a Real World Evaluation of Anti-Phishing Training. *e-Crime Researchers Summit, Anti-Phishing Working Group* (October 2008). http://precog.iiitd.edu.in/Publications_files/eCrime_APWG_08.pdf

[27] Nhien-An Le-Khac and Tahar Kechadi. 2015. Security Threats of URL Shortening: A User's Perspective. *Journal of Advances in Computer Networks* 3, 3 (2015).

[28] Eric Lin, Saul Greenberg, Eileah Trotter, David Ma, and John Aycock. 2011. Does Domain Highlighting Help People Identify Phishing Sites?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2075–2084. http://grouplab.cpsc.ucalgary.ca/grouplab/uploads/Publications/Publications/2011-DomainHighlighting.CHI.pdf

[29] Drew A. Linzer and Jeffrey B. Lewis. 2011. poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software* 42, 10 (2011), 1–29. http://www.jstatsoft.org/v42/i10/

[30] Federico Maggi, Alessandro Frossi, Stefano Zanero, Gianluca Stringhini, Brett Stone-Gross, Christopher Kruegel, and Giovanni Vigna. 2013. Two Years of Short URLs Internet Measurement: Security Threats and Countermeasures. In *Proceedings of the 22nd international conference on World Wide Web - WWW '13*. 861–872. http://point.to/redir.html

[31] Max-Emanuel Maurer and Lukas Höfer. 2012. Sophisticated Phishers Make More Spelling Mistakes: Using URL Similarity against Phishing.. In *CSS*. Springer, 414–426.

[32] Nick Nikiforakis, Federico Maggi, Gianluca Stringhini, M. Zubair Rafique, Wouter Joosen, Christopher Kruegel, Frank Piessens, Giovanni Vigna, and Stefano Zanero. 2014. Stranger Danger: Exploring the Ecosystem of Ad-based URL Shortening Services. *Proceedings of the 23rd international conference on World wide web - WWW '14* (2014), 51–62. DOI: http://dx.doi.org/10.1145/2566486.2567983

[33] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70 (2017), 153–163.

[34] PhishLabs. 2018. *2018 Phishing Trends & Intelligence Rport: Hacking the Human*. Technical Report. Ecrime Management Strategies, Inc.

[35] Ponemon. 2018. *Cost of Cyber Crime Study: Insights on the security investments that make a difference*. Technical Report.

[36] Elissa M Redmiles, Amelia R Malone, and Michelle L Mazurek. 2016. I Think They're Trying to Tell Me Something: Advice Sources and Selection for Digital Security. In *2016 IEEE Symposium on Security and Privacy (SP)*. IEEE, 272–288. DOI: http://dx.doi.org/10.1109/SP.2016.24

[37] R. Reeder, I. Ion, and S. Consolvo. 2017. 152 Simple Steps to Stay Safe Online: Security Advice for Non-tech-savvy Users. *IEEE Security Privacy* PP, 99 (2017), 1–1. DOI: http://dx.doi.org/10.1109/MSP.2017.265093101

[38] Joel Ross, Lilly Irani, M. Six Silberman, Andrew Zaldivar, and Bill Tomlinson. 2010. Who Are the Crowdworkers?: Shifting Demographics in Mechanical Turk. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10)*. ACM, New York, NY, USA, 2863–2872.

[39] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. 2007. Anti-Phishing Phil: The Design and Evaluation of a Game That Teaches People Not to Fall for Phish. In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS '07)*. ACM, New York, NY, USA, 88–99. DOI: http://dx.doi.org/10.1145/1280680.1280692

[40] The Federal Bureau of Investigation (FBI), Internet Crime Complaint Center. 2017. *2017 Internet Crime Report*. Technical Report. https://pdf.ic3.gov/2017_IC3Report.pdf.

[41] Yue Zhang, Serge Egelman, Lorrie Cranor, and Jason Hong. 2007. Phinding Phish: Evaluating Anti-Phishing Tools. (2007). http://lorrie.cranor.org/pubs/ndss-phish-tools-final.pdf