Int. J. Human-Computer Studies 82 (2015) 69–82

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Why phishing still works: User strategies for combating phishing attacks ☆

Mohamed Alsharnouby, Furkan Alaca, Sonia Chiasson *

School of Computer Science, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada K1S 5B6

ARTICLE INFO

Article history:
Received 3 September 2014
Received in revised form
18 March 2015
Accepted 10 May 2015
Communicated by Scott Bateman
Available online 21 May 2015

Keywords:
Phishing
Eye tracking
Usable security
User study

ABSTRACT

We have conducted a user study to assess whether improved browser security indicators and increased awareness of phishing have led to users' improved ability to protect themselves against such attacks. Participants were shown a series of websites and asked to identify the phishing websites. We use eye tracking to obtain objective quantitative data on which visual cues draw users' attention as they determine the legitimacy of websites. Our results show that users successfully detected only 53% of phishing websites even when primed to identify them and that they generally spend very little time gazing at security indicators compared to website content when making assessments. However, we found that gaze time on browser chrome elements does correlate to increased ability to detect phishing. Interestingly, users' general technical proficiency does not correlate with improved detection scores.

1. Introduction

An important aspect of online security is to protect users from fraudulent websites and phishing attacks. *Phishing* is a "criminal mechanism employing both social engineering and technical subterfuge to steal consumers' personal identity data and financial account credentials" (Anti-Phishing Working Group, 2014a). While advances in the automated detection of phishing websites have resulted in improved security, these automated means are not fool-proof and users must be vigilant in protecting themselves in this arms race (Hong, 2012). According to the Anti-Phishing Working Group, phishing attacks remain widespread: 42,890 unique phishing websites were reported in December 2013, with the financial and online payment sectors accounting for nearly 80% of targeted industries (Anti-Phishing Working Group, 2014a).

Modern web browsers provide tools to assist users in making informed security decisions. For example, visual indicators within the URL bar and the SSL padlock have been designed to allow users to judge the legitimacy of websites. Unfortunately, these indicators have been only partially successful at helping to prevent phishing. Poor usability may allow phishing websites to masquerade as legitimate websites and deceive users into divulging their personal information. Earlier browser security indicators have been shown in previous studies to be ineffective, putting users at a higher risk of falling victim to phishing attacks (Whalen and Inkpen, 2005; Lin et al., 2011; Egelman, 2009).

This is compounded by the fact that security is a secondary task for most users (Whitten and Tygar, 1999). Users who are concentrating on the real purpose of their online interaction, such as making a purchase, are unlikely to notice security indicators. Furthermore, some security indicators are visible only when the website is secure. The *absence* of a security indicator, as is possible with phishing websites, is even less likely to be noticed by users. Therefore, developing usable browser security cues to combat phishing attacks remains an important and unsolved problem in usable security, as is understanding how users make determinations about the legitimacy of websites (Purkait, 2012).

To inform the design of improved techniques against phishing, we explored the strategies employed by users to identify phishing attacks. We showed participants a series of websites and asked them to identify whether each one is legitimate or fraudulent. This paper makes several distinct contributions to the literature. First, we evaluate the effectiveness of recent changes that have been made in web browser designs to help users identify fraudulent websites. Secondly, we assess whether users have developed improved detection strategies and mental models of phishing nearly a decade after Dhamija et al. (2006)'s initial phishing study. And finally, we are the first to use eye tracking data to obtain quantitative information on which visual security indicators draw the most attention from users as they determine the legitimacy of websites. Based on our results, we identify aspects in which web browser security indicators have improved in

☆This paper has been recommended for acceptance by Scott Bateman.

* Corresponding author. Tel.: +1 613 520 4333.
E-mail addresses: malsharnouby@ccsl.carleton.ca (M. Alsharnouby), falaca@ccsl.carleton.ca (F. Alaca), chiasson@scs.carleton.ca (S. Chiasson).

http://dx.doi.org/10.1016/j.ijhcs.2015.05.005
1071-5819/

modern web browsers, identify areas for potential improvement, and make recommendations for future designs.

The remainder of this paper is organized as follows: Section 2 reviews related work on phishing detection and tools to aid users in identifying phishing websites. Section 3 details our study methodology. Section 4 provides analysis and interpretation of our quantitative and qualitative data. Section 5 discusses some ideas for future web browser designs, while Section 6 concludes the paper.

2. Related work

Research on protecting users against phishing attacks has taken four complementary approaches: automating phishing detection, providing user interface cues to help users detect phishing, educating users about how to protect themselves, and understanding users' susceptibility to phishing to inform the design of protection mechanisms. Our work falls within scope of the fourth area, but we also provide a brief overview of the other areas to give context to our work. For a general introduction, see Hong (2012)'s article, or for a more complete recent review of the phishing literature, see Purkait (2012)'s literature survey.

2.1. Automated phishing detection

The first line of defense against phishing should be automated detection; users cannot fall for phishing attacks if they never see the attacks. Automatic phishing detectors exist at several different levels: mail servers and clients, internet service providers, and web browser tools. Tools may block access to a detected phishing website and/or request that the website's internet service provider takes down the website (Moore and Clayton, 2007).

Automatic email classification tools commonly use machine learning techniques (Fette et al., 2007), statistical classifiers (Bergholz et al., 2010), and spam filtering techniques (Cormack, 2008) to identify potential phishing messages with varying degrees of effectiveness as the threat continues to evolve. Mis-classifications affect the perceived reliability of the service and users are likely to be quite intolerant to "losing" legitimate messages.

Techniques to detect phishing websites include blacklists, machine learning (Whittaker et al., 2010), URL feature classification and domain name analysis, visual similarity assessment (Fu et al., 2006), contextual analysis and user behavioural prediction (Lee et al., 2014), and crowdsourcing (OpenDNS, 2014). Some blacklists, such as Google's (Whittaker et al., 2010), use automated machine learning. PhishTank (OpenDNS, 2014) offers a blacklist for use by other tools through an API. Its blacklist is populated through crowdsourcing volunteers who submit potential phishing websites and vote on the legitimacy of websites.

Web browsers maintain their own blacklists and heuristics for detecting phishing, displaying warnings to users if they reach a known phishing page. Detection rates have improved considerably over the last 5 years. NSS Labs (2013) conducts independent tests and found that the major browsers had an average phishing detection rate of approximately 90%, with zero-hour block rates above 70%. Third-party add-ons are also available. Sheng et al. (2009) evaluated the effectiveness of eight different browser tools and found them generally slow at detecting new phishing campaigns. This is problematic given that the median lifetime of a phishing campaign is about 12 h (NSS Labs, 2013), with many as short as 2 h.

While successful at stopping a large number of attacks from reaching users, automated methods are insufficient as the sole means of protecting users. Secondary methods involving users are necessary for times when automatic detection fails.

2.2. Security indicators

There have been a number of studies regarding phishing and the usability of browser security cues. Herzberg (2009) provides an overview of several studies.

At its core, phishing is a threat because users are unable to verify the authenticity of the website asking for their credentials. Dhamija and Tygar (2005) first proposed Dynamic Security Skins, a browser extension that allows websites to display a secret image and customizes the browser chrome. Variations of this secret image method have now been deployed by banks and major organizations (e.g., Sitekey Bank of America, 2014; Yahoo Sign-in Seals Yahoo! Inc, 2014). Anecdotal evidence suggests that some users may still fall victim to phishing websites who claim that the image database is down for maintenance or who simply leave out this feature since the absence of a cue may not trigger attention. Many browser toolbars (e.g., Chou et al., 2004; Yee and Sitaker, 2006; Li and Helenius, 2007; Kirda and Kruegel, 2006; Kirlappos and Sasse, 2012) have also been proposed to protect against phishing, each with limited success. User studies by Wu et al. (2006), Li and Helenius (2007), and Li et al. (2014) found that security toolbars intended to prevent phishing attacks were ineffective and identified several usability problems. While users may occasionally pay attention to the indicators, accomplishing that their primary task often gets prioritized, and in these cases users look for visual signs reinforcing the website's trustworthiness rather than heeding warnings to the contrary (Kirlappos and Sasse, 2012). Abbasi et al. (2012) compared users' ability to detect phishing given high- or low-performing browser toolbars and found that users were more successful with the high-performing toolbar. However, users still ignored the toolbar's advice 15% of the time, instead believing that their own intuition was more accurate.

Others have explored the browsers' built-in security indicators. Lin et al. (2011) examined the effectiveness of domain highlighting that is now included in most browsers. They found it to be only marginally successful when users' attention was explicitly drawn to the address bar. Egelman (2009) explored various online trust indicators, including web browser phishing warnings and SSL warnings. They found that 97% of users were fooled by at least one attack, but that active warnings which interrupt users' tasks were more effective than passive warnings.

Although addressing a tangential issue, password managers (Yee and Sitaker, 2006; Ross et al., 2005) can offer protection against phishing by storing both the user's credentials and the legitimate URL at which these credentials should be used. Users attempting to use their password manager at a phishing website will either be warned against a suspicious website or the password manager will supply incorrect credentials.

Efforts to reduce phishing at the email level are also popular, but these typically require minimal user involvement beyond needing to occasionally check spam-filtered mail and potentially update spam filters. Email encryption and digital signing can help protect users against phishing and other attacks, but these are plagued with usability issues and are not widely used (Garfinkel et al., 2005).

2.3. Anti-phishing education

Although educational efforts are unlikely to solve the phishing problem on its own, vigilant users form an important part of the defensive strategy. Both research efforts and public education campaigns (e.g., Anti-Phishing Working Group, 2014b; Government of Canada, 2014) have focused on teaching users how to protect themselves against phishing attacks. PhishGuru (Kumaraguru et al., 2007, 2009, 2010) embeds phishing education within the primary task of receiving phishing email and results show that the educational material is most impactful if delivered immediately after users have

fallen for a phishing attack, a method now deployed on a large scale by the Anti-Phishing Working Group (2014c)'s landing page.

Sheng et al. (2007) developed Anti-Phishing Phil, a web-based game to teach about phishing attacks, and show that users who played the game were better able to identify phishing websites immediately after playing the game and one week later. Some independent evidence of its effectiveness is provided by Mayhorn and Nyeste (2012). A mobile version of the game was also developed (Arachchilage et al., 2012).

In attempts to formalize the educational process, Arachchilage and Love (2013) are working towards a game design framework based on user motivations. Furthermore, Burns et al. (2013) propose an intervention model describing the most effective types of interventions based on users' stage of knowledge.

2.4. Understanding user behaviour

A significant assumption by attackers is that they will be able to deceive users through websites with visual characteristics sufficiently believable to be accepted as legitimate. As discussed below, early studies showed that users were trusting websites based on quick visual assessments that did not necessarily focus on the most reliable indicators. Some research has focused on characteristics of phishing attacks that have increased likelihood of success while other research seeks to determine characteristics of users that place them at increased risk. It is generally agreed, however, that users are poor at detecting phishing attacks. Users' susceptibility to phishing has been explored using several methods, including lab (e.g., Jakobsson et al., 2007; Dhamija et al., 2006; Whalen and Inkpen, 2005) and field studies (e.g., Wright and Marett, 2010; Jagatic et al., 2007), surveys (e.g., Workman, 2008; Vishwanath et al., 2011; Downs et al., 2007), and Mechanical Turk studies (e.g, Sheng et al., 2010).

As it is the gateway for many phishing attacks, several studies have explored users' likelihood of falling for phishing emails. Personalization of email content, urgency cues, and email load all contribute to increase susceptibility, as does low technical expertise (Vishwanath et al., 2011). Jagatic et al. (2007) offer one of the earliest investigations of user behaviour with respect to phishing. Their field study simulated a targeted phishing attack against unsuspecting university students who received email apparently from an acquaintance. Results show that users were significantly more likely to fall for targeted attacks than generic phishing scams. Vishwanath et al. (2011) completed a survey of intended victims who had recently been targets of two real email phishing campaigns on a university campus. Decisions about phishing were driven by users' motivation, beliefs, prior knowledge and experiences. They further argue that creating habitual rituals of safer behaviour may be more successful than encouraging constant vigilance and alertness.

On the other hand, Downs et al. (2007)'s survey study found that technical knowledge of the web environment led to increased resistance against phishing and suggested education on how to interpret browser cues as a preventative technique. Similarly, Wright and Marett (2010) conducted a field study where university students were sent phishing email purportedly from a system administrator and found that an increased level of web experience and security awareness led users to successfully detect the attack. In a role-playing scenario, Sheng et al. (2010) asked MTurk workers to suggest their likely course of action in response to screenshots of email messages. They evaluated phishing susceptibility against demographic characteristics and found that prior exposure to educational material and a general aversion to risk led people to better detect phishing attempts.

Other studies have explored users' responses to phishing within the web browser. In 2006, Dhamija et al. (2006) conducted a lab-based study where participants were asked to assess the legitimacy of a series of websites. Participants were primed on the purpose of the task and this was clearly a "best-case scenario" which tested users' *ability* to detect phishing rather than the users' *usual practice* when encountering websites. Regardless, 42% of the websites were incorrectly classified by users. Using self-reports and observation, it was determined that 59% of users relied solely on the webpage content and the URL to assess legitimacy, ignoring any security cues provided by the browser. In 2007, Jakobsson et al. (2007) also asked users to assess the legitimacy of emails and websites in a lab environment. Users reported relying on the content of the emails and websites, being suspicious when too much emphasis was placed on security, and trusting signs of personalization or familiar third-party endorsements.

Phishing is now a commonly known attack, discussed in mass media, and most users are familiar with the risk. Have users become more savvy as a result of this familiarity? Are they more capable of protecting themselves than they were a decade prior? We have followed a similar methodology to that used by Dhamija et al. (2006) in their study, but have collected eye tracking data to supplement participants' self-reported data specifically as they were assessing likelihood of phishing. Moreover, since there have been a number of design changes to web browser interfaces in recent years aimed at increasing security, our study examines whether these have led to improved phishing detection by users.

2.5. Eye tracking in phishing studies

To our knowledge, only two related studies have used the eye tracker as a data collection method although neither explicitly looked at phishing. Sobey et al. (2008) used an eye tracker to compare their proposed Extended Validation Certificate interface with Firefox's existing interface. They found that the eye tracker data confirmed users' reported experiences. Users who reported viewing the indicators did gaze at them, but the majority of users' time was spent gazing at the content of the page rather than the browser chrome. Whalen and Inkpen (2005) explored users' use of security cues while completing web transactions. Using eye tracking data, they found that two thirds of users looked at the SSL lock icon when prompted to be security-conscious but rarely used other cues on the browser chrome.

3. Methodology

We conducted an eye tracking usability study to investigate on which strategies users rely to determine the legitimacy of websites. The study's methodology was approved by the Carleton Research Ethics Board.

3.1. Overview of study

Our tests were conducted on a standard Windows XP desktop computer equipped with a Tobii 1750 eye tracker. The eye tracker was used to record the participants' gaze information while viewing the entirety of the websites. The only software which the participants used during the session was a web browser maximized to full-screen. We built a web-based interface, as seen in Fig. 1, to allow participants to navigate easily between the test websites and to launch them one at a time. When designing the web interface, we took the following precautions to reduce possible sources of bias: (1) we randomized the order of the websites for each participant, and (2) we showed the websites one at a time to hide the fact that some websites appeared more than once. We presented participants with a total of 24 websites: 10 were legitimate and 14 were phishing websites. The participants were asked to determine whether each website was legitimate or fraudulent, and asked to explain how they arrived at their decision. They were also asked to rate their level of certainty in their decision on a scale from 1 (not at all certain) to 5 (very certain).

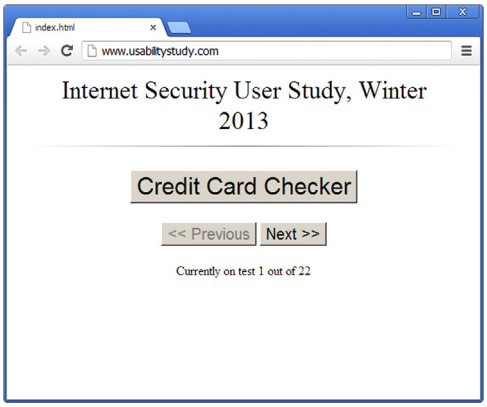

Fig. 1. Web interface where users navigate between test websites.

Each session lasted approximately one hour, with two experimenters present. One experimenter was responsible for the technical aspects of the experiment (e.g., to assist the participant if they closed the main browser window by accident, or to remind the participant to adjust their posture if the eye tracker could no longer detect their eyes) and for asking the appropriate questions for each website (e.g., "Do you think that was a legitimate website or a phishing website? How confident are you in your decision?"). The other experimenter was responsible for recording the participant's answers and noting down real-time observations from the eye tracker when possible. After the participant had viewed all of the test websites, they were asked questions in a semi-structured post-session interview. Each session was audio-recorded to allow missed details to be transcribed at a later time. At the end of the interview, we provided participants with $15 and a debriefing form with our contact information. We also informed them about how well they performed, and explained how they can use browser security cues to protect themselves. We note that this setup represents a best case scenario where participants are primed to detect attacks, representing an upper bound on their ability in real life rather than their usual habits.

3.2. Participants

The 21 participants (12 female, 9 male) were primarily recruited from Carleton University through posters and e-mail. A pre-session questionnaire was emailed to respondents to collect demographic information, web browser preferences, and data on technical proficiency. Portions of this questionnaire were adapted from Egelman (2009)'s questionnaire assessing technical proficiency and a value was assigned to each of them with zero being the lowest and five being the highest technical proficiency. Participants' ages ranged from 18 to 51; the mean age was 27 (σ= 10.19). Fourteen participants were students (undergrad or grad) and seven were professionals. Participants had a wide range of backgrounds; four participants were in engineering fields (electronics, biomedical, physics and electrical) and 17 were in non-technical fields. None had any specialization in computer science or computer engineering.

Additional information about our participants' online habits is summarized in Table 1. The majority of participants used either Chrome or Firefox as their primary browser; 85% of our participants reported also using a secondary browser. All participants reported using online banking, and most do shopping online. Few of our participants have knowingly been victims of online attacks. Their technical proficiency appears in line with everyday computer users; many have designed a website and adjusted firewall settings, but few have used telnet or registered a domain name. Technical proficiency scores ranged from 0 to 5 with a mean score of 1.6 (σ= 1.47).

This questionnaire was completed ahead of the lab sessions and we used it to tweak the test websites used in the study. For example, most of our participants were not users of Instagram, so we excluded it from the study.

Table 1
Demographics and online habits of our participants.

Online habits category	Characteristic	Percentage
Primary browser	Chrome	52
	Firefox	28
	Internet explorer	10
	Safari	10
Online use	Banking	100
	Shopping	86
Operating system	Windows	86
	Mac OS	14
Online attack	Account break-in	19
	Credit card fraud	14
	Identity theft	0
Proficiency	Designed website	57
	Changed firewall settings	43
	Installed OS	38
	Registered domain name	10
	Used telnet/SSH	10
Website	Facebook	86
	Amazon	76
	cuLearn	72
	Kijiji	67
	Paypal	38
	Ebay	38
	Twitter	33
	Instagram	14
Demographics		
Sex	Male	9
	Female	12
Age (mean)		27
Field	Technical	4
	Non-Technical	17

3.3. Tasks

The participants received verbal instructions during the test. We briefly explained that phishing is typically characterized by a fraudulent website which mimics a legitimate website in an attempt to trick users into revealing personal information such as passwords or banking information. We asked participants to imagine that they had visited each of the 24 websites through an email message or some other link. Their task was to decide whether each was a legitimate website or a phishing attempt, and explain the reasons behind their decision. If we noticed participants trying to enter their own credentials into any of the websites, we advised against it and told them to enter fake information if they wished. We did not log any credentials entered. This scenario is similar to the original study by Dhamija et al. (2006).

3.4. Experimental setup

We tailored our websites to ensure that several would be familiar to participants. We included the Carleton University website, major Canadian banks, and a number of websites from the Alexa top 500 list in Canada.[1] The complete list of websites used in the experiment is available in Table 2.

All of the phishing websites were hosted from an Apache web server running on a laptop. We also set up a certificate authority on our laptop, and used it to issue certificates for some of our

[1] http://www.alexa.com/topsites/countries/CA.

Table 2
Websites shown to participants, with corresponding phishing techniques. (B) = bank website, (U) = university website. Types are described in Section 3.4.2.

	Type	Website	Description
Phishing	1	Scotiabank (B)	Real website in iframe, overlaid by another iframe with malicious login link
	1	TD (B)	Real website in iframe, overlaid by another iframe with malicious login link
	1	RBC (B)	Home and login pages replicated, all other links point to real website
	1	CIBC (B)	Home and login pages replicated, all other links point to real website
	1	HSBC (B)	Home and login pages replicated, all other links point to real website
	1	cuLearn (U)	Mistyped URL (cartelon.ca instead of carleton.ca)
	1	Carleton Portal (U)	Mistyped URL (cartelon.ca instead of carleton.ca)
	1	Paypal (Favicon)	Favicon set to green padlock
	1	Gov. of Canada (theft)	Sign-up page for identity theft protection. Asks for Social Insurance Number
	2	Paypal (IP)	IP Address as URL
	3	Amazon	Browser chrome replicated in page contents to spoof a new window
	4	Carleton University (U)	Legitimate home page with fraudulent popup asking for login credentials
	5	Facebook (popup)	Legitimate website overlaid with a phishing pop-up window
	6	Credit Card Checker	Asks for credit card details to check if card has been compromised
Legitimate	7	Research Survey (U)	Legitimate research survey page with self-signed certificate
	7	Gov. of Canada	Legitimate, non-SSL
	7	Facebook	Legitimate, with SSL
	7	HSBC (legit) (B)	Legitimate, with EV SSL certificate
	7	LinkedIn (non-SSL)	Legitimate, non-SSL
	7	LinkedIn (SSL)	Legitimate, with SSL
	7	Pinterest	Legitimate, non-SSL
	7	Kijiji	Legitimate, non-SSL
	7	Netflix	Legitimate, non-SSL
	7	Twitter	Legitimate, with EV SSL certificate

phishing websites. In fact, in order to show the ease with which it is possible to obtain a domain-validation certificate from a reputable provider without extensive verification, we purchased http://www.cartelon.ca as well as a corresponding SSL certificate (both from GoDaddy.com). To prepare the desktop computer used for our experiment, we used our own certificate authority by adding its certificate to the web browser and modified the hosts file to route requests to the phishing websites to the laptop's web server.

3.4.1. Web browser selection

At the time of the study, there were four major web browsers: Microsoft Internet Explorer 10, Mozilla Firefox version 10.0.12, Google Chrome version 26.0.1410.43, and Apple Safari version 6.0.3. We chose to use Google Chrome. Since our eye tracker could only run on Windows XP, it was technically infeasible to create a remote desktop connection to an Apple OS X based machine from the eye tracker; network latency would have caused delays and undesirable experiences for our participants. We also ruled out Internet Explorer, due to its apparent unpopularity among members of the Carleton University community. To choose between Firefox and Chrome, we performed a user interface comparison and studied the differences between how they display visual security indicators (although we had already ruled out Internet Explorer, we have included it in our comparison because of its high market share).

SSL lock: We found that while Chrome uses a vivid green SSL lock with well-defined edges, Firefox and Internet Explorer both use a dull grey lock which could easily go unnoticed. As can be seen in Fig. 2, Firefox and Chrome both display the SSL lock on its own on the left-hand side of the URL box, whereas Internet Explorer displays it on the right-hand side of the URL box, buried between a number of other icons which appear in the same shade of grey (strangely, however, the Internet Explorer SSL lock is more prominent in the Windows XP theme as compared to the more recent Windows 7 theme).

Https and URL: Chrome displays the https portion of the URL in green, whereas Firefox displays it in grey and Internet Explorer displays it in black. All three browsers display the domain name in black text (but Chrome also includes the subdomains) and the remainder of the URL in grey text.

EV certificates: We compared SSL indicators for extended validation (EV) certificates, and found that they are virtually identical in Chrome and Firefox; both display a prominent green rectangle containing the company name in the left-hand side of the URL box. Internet Explorer turns the entire URL bar green and places the company name on the right-hand side of the URL box.

Certificate warnings: Regarding SSL certificate warnings, we found that Firefox does not allow the user to proceed to the website without adding a permanent exception for that website, and never again presents any warning indicators for subsequent visits. Chrome and Internet Explorer, however, do not prompt the user to add an exception to proceed. Even after proceeding, Chrome overlays a red cross over the https portion of the URL and Internet Explorer turns the entire URL bar red. Moreover, we found that Chrome's certificate warning page, which has a deep red background, seems much more intimidating than in the other two browsers.

Favicon: It was reported by Dhamija et al. (2006) that many users were easily deceived when phishing websites set their favicon to be a padlock. This problem appears to be taken into account by Chrome and Firefox, which have both moved the favicon from the URL box to the tab corresponding to the page, thereby separating the locations of the favicon and SSL padlock. However, Internet Explorer still places the favicon on the left-hand side of the URL bar, and it is much more prominent than the dull-grey SSL lock on the right-hand side of the URL.

Overall summary: We believe that Chrome and Firefox do a better job of locating the SSL indicators, since the https, lock icon, and EV certificate are displayed cleanly on the left-hand side before the URL. Chrome, however, makes more extensive use of colour. A final difference that we noticed regarding the URL bar with Chrome is that it uses a visibly larger font when compared to Firefox and Internet Explorer. We also compared the layout of user interface in Chrome with that of Firefox and found them to be virtually identical. Given the nature of our experiment, where we specifically instructed our participants to identify phishing websites, we decided that it would be consistent with our objectives to

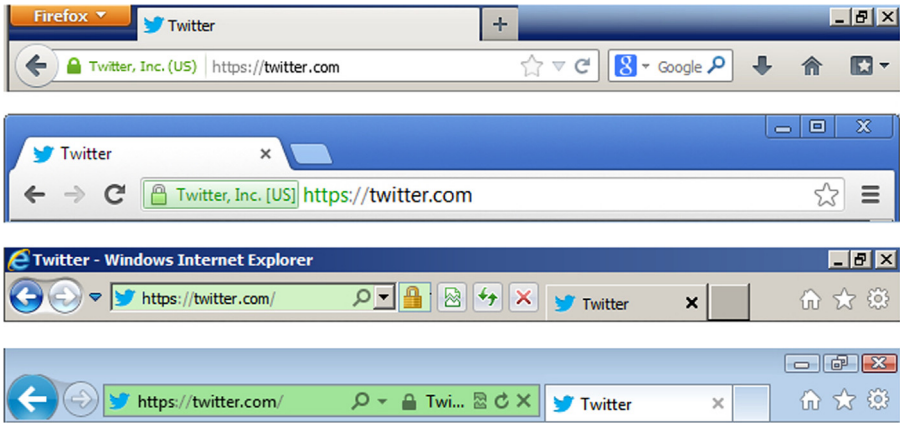

Fig. 2. Comparison between browser chromes (Top to bottom: Mozilla Firefox, Google Chrome, Internet Explorer 10 with Windows XP theme, Internet Explorer 10 with Windows 7 theme).

pick the browser with the most prominent security indicators, which we believe is Chrome.

3.4.2. *Phishing techniques and categories of attacks*

The following is an overview of the general categories of websites which we tested in our study. In an effort to test for all possible web browser security cues, we employed a number of common techniques such as incorrect and typejacked URLs, overlaid iframes and popups, missing SSL certificates, and SSL certificate errors. Table 2 summarizes which test website fell into each category.

Type 1: *Spoof websites with incorrect URL*. For phishing websites in this category, we downloaded the legitimate website to our laptop computer using HTTrack.[2] We then modified the code as necessary to ensure that the main page would load properly when served by our web server, and that all links on the main page led to the legitimate website. Therefore, participants paying attention to the URL bar would notice that clicking on any link from the main page causes the domain name of the website to change. In some cases, even the SSL indicator would change (e.g., the website would change from having no SSL to having SSL, or would change from having a domain-validation certificate to an EV certificate). Some websites in this category had mistyped domains such as cartelon.ca (instead of carleton.ca), whereas others had completely erroneous domains such as scotiabank.secure-encrypt05.com.

For some websites, we created a page which displays the legitimate website in an iframe that covers the entire width of the browser window. We then overlaid the website with iframes on top of the login button(s), to redirect the user to our spoofed login page. Since the legitimate page is displayed in an iframe, it is guaranteed that the user will never see an outdated page. Secondly, the URL bar does not change as the user clicks on different links. However, if they visit a page which does not normally contain a login button, an observant participant should notice that our iframe with the login button still appears when it should not. Interestingly, we noticed that some websites (e.g., Gmail) made use of the X-Frame-Options header forbidding the web browser from displaying it inside of a frame, whereas the majority of Canadian banks did not take advantage of this feature.

Lastly, we set the favicon to the green SSL lock on the Paypal website to see if that would deceive participants into thinking that they were on a secure website.

Type 2: *Spoof websites with IP address as URL*. Another classic phishing technique is to use an IP address instead of a domain. We suspected that users would more easily notice this type of attack, and therefore chose to use this technique for only one website (Paypal).

A technique which previously was commonly used in conjunction with this category (and also the previous category) of spoofs was to embed a long user name and password before the URL in the form of http://username:password@domain. Phishing URLs would take advantage of this format by filling the username segment with a string long enough to displace the actual domain or IP address, and beginning with what appears to be a legitimate URL. However, modern web browsers do not display the user name and password in the URL bar, rendering this type of attack obsolete.

Type 3: *Fake chrome*. There have been phishing attacks which attempt to spoof the *browser chrome*.[3] In the past, when browsers supported toolbarless popups, it was possible to reproduce a remarkably realistic spoof of the browser chrome complete with the URL bar and navigation buttons. However, modern web browsers display at least the URL bar in all popups, which makes it more difficult for phishers to spoof the browser chrome in a non-obvious way. For this category, we opted for a relatively primitive attack which we hoped that most participants would catch. We constructed a page with two horizontal frames: the top frame displayed an image of a browser chrome with the Amazon Canada URL, complete with the green SSL lock, and the bottom frame displayed our fake Amazon login page. The real chrome showed a fake URL (http://secure-signin_amazon.ca) with no SSL lock.

Type 4: *Popups asking for credentials*. For this category, we simulated rogue popups which appear over the legitimate website and ask for the user's credentials. As mentioned above, since modern web browsers always include a URL bar in popup windows, these attacks may be less effective than they were previously. We designed a popup window for the Carleton University website which prompts the student for their user name and password to log in.

Type 5: *Overlaid popup windows*. We also tried another type of popup attack, which overlays a rogue website over the entire contents of the legitimate website, while leaving the chrome and URL bar visible. We again predicted that the URL bar on the popup window would make it easier for participants to detect this type of attack. We used Facebook for this attack, since we wanted to see if users would detect this relatively large inconsistency on a website that they use often.

[2] An open source web crawler: http://www.httrack.com/.

[3] The borders of the browser, including the menus, toolbars, and buttons.

Type 6: *Fraudulent based on context*. We included one website which did not attempt to masquerade as any other organization, but instead was fraudulent based on context. We constructed a page which claimed to check if a user's credit card has been stolen.[4] We included a "Verified by Visa" logo in order to make the website appear more credible. We included this website with the expectation that all participants would recognize it as a phishing attempt.

Type 7: *Legitimate websites*. While Dhamija et al. (2006) tested legitimate websites which looked suspicious, such as a legitimate website hosted by a third-party service or mobile websites with no graphics, we omitted such tests for two reasons: (1) due to the difficulty of finding such a legitimate website which would be of any relevance to our participant base, and (2) due to our belief that people should in fact be suspicious when they see such cues, particularly in cases where the website is hosted by a third-party service. Instead, we used popular websites which should be obviously legitimate based on cues from both the browser chrome and content. One variable which we did include in the legitimate category was to test both the http and https versions of LinkedIn, to see if participants would distinguish between the two in the controlled setting of our lab study.

Given the ease with which an SSL certificate can be obtained for a fraudulent website, we believe that it is possible for the lock icon to give a false sense of security to the users. For this reason, we generated SSL certificates for a number of our phishing websites. We suspected that if none of the phishing websites had SSL locks, some participants may have found that to be enough of a reason to declare the website as a phishing attempt, without bothering to read the URL bar. Therefore, by generating an SSL certificate for domains such as cartelon.ca, we aimed to eliminate all possible phishing indicators other than the mistyped domain name.

3.5. Data collection and analysis protocols

We collected both qualitative and quantitative data using several methods. We recorded users' self-reported decision about whether a website was fraudulent or legitimate and comments on why they reached this decision. Participants also rated their certainty level with each decision on a scale of 1 (not at all certain of their choice) to 5 (completely certain of their choice). We recorded eye tracking data as they interacted with each website and the time it took for participants to judge the legitimacy of each website. We held a short post-session semi-structured interview to understand participants' knowledge and experiences with phishing and their knowledge of browser security cues. During the interview, we prompted participants about whether they knew the meaning of specific indicators (e.g., SSL lock, https) and why they had made certain decisions during the test.

3.5.1. Interview protocol

The interview revolved around the following themes:

- Knowledge and experience with phishing: e.g., whether or not they have heard of phishing before, if they understand what it means, if they recall ever having encountered phishing attempts, and the strategies that they employ to identify phishing attempts.
- Knowledge of browser security cues: e.g., checking the address bar for SSL lock icon, correct URL, and https.
- Knowledge of certificates and SSL and their indicators: e.g., whether they understand the purpose and meaning of certificate warnings presented by the browser and the glowing green bar shown for EV certificates.

We recorded each session with a digital audio recorder and transcribed them for our analysis. We believe that conducting an interview-style session where we asked users to verbally explain their choices and answer our questions allowed us to collect more valuable data in comparison to a written or electronic questionnaire, where it would take more effort for participants to give detailed answers.

3.5.2. Eye tracking analysis

Participants' interactions with the computer were recorded using the eye tracker. Multiple steps were taken to prepare the eye tracking data for analysis. For each participant, we divided the recording of the entire eye tracking session into separate recordings per website. This eliminated the eye tracking data recorded during the use of the main experiment interface (Fig. 1) and also allowed us to filter actions that are considered irrelevant to our analysis.

To associate the raw eye tracker data with specific areas on the screen, we defined Areas of Interest (AOIs) for each website using the analysis software bundled with the eye tracker, as shown in Fig. 3. The software generated data specific to the AOI such as the timestamp and total dwell time for that area. AOIs were grouped into two main categories: *browser chrome* and *webpage content*. The browser chrome category included areas such as the lock icon within the chrome (Fig. 3, Label A), the https/http section of the URL address bar (Fig. 3, Label B), the address bar (Fig. 3, Label C), and the bottom section of the chrome (Fig. 3, Label E) where information such as hyperlinks are displayed. The content category included only one area of interest: the content of the pages displayed within the chrome (Fig. 3, Label D). Since some AOIs overlapped, for example the lock icon AOI and the address bar AOI, care was taken when the data was being prepared to subtract these values accordingly.

The eye tracking data was imported into a MySQL database, and SQL procedures were written to format the data into a more usable form. This intermediate data was further analyzed and combined with data from the interview such as participants' certainty levels and decisions to produce the final results.

4. Results

We report on participants' success at identifying fraudulent websites, the time they spent looking at the different aspects of the screen, and their reasoning for making decisions. Where appropriate, we conducted statistical analysis to see whether demographic characteristics affected results. We used independent sample t-tests and Pearson's Moment Product Correlation Coefficient (i.e., Pearson's r) as appropriate to assess whether certain participant characteristics affected their performance scores. In all cases, a $p < 0.05$ was considered statistically significant.

4.1. Performance scores

We assigned each participant a score, which we defined as the sum of the number of websites which they identified correctly as either legitimate or phishing websites. The scores ranged from 9 to 22, out of a total of 24 websites ($\mu = 15.28$, $\sigma = 3.97$). The performance results are summarized in Table 3. The table includes the success rate for each website (i.e., the percentage of participants who correctly identified whether the website was legitimate or fraudulent) and the average time it took participants to make a decision, broken down into time spent on AOIs and total decision time. It also includes the number of participants who thought each

[4] Based on http://www.ismycreditcardstolen.com/.

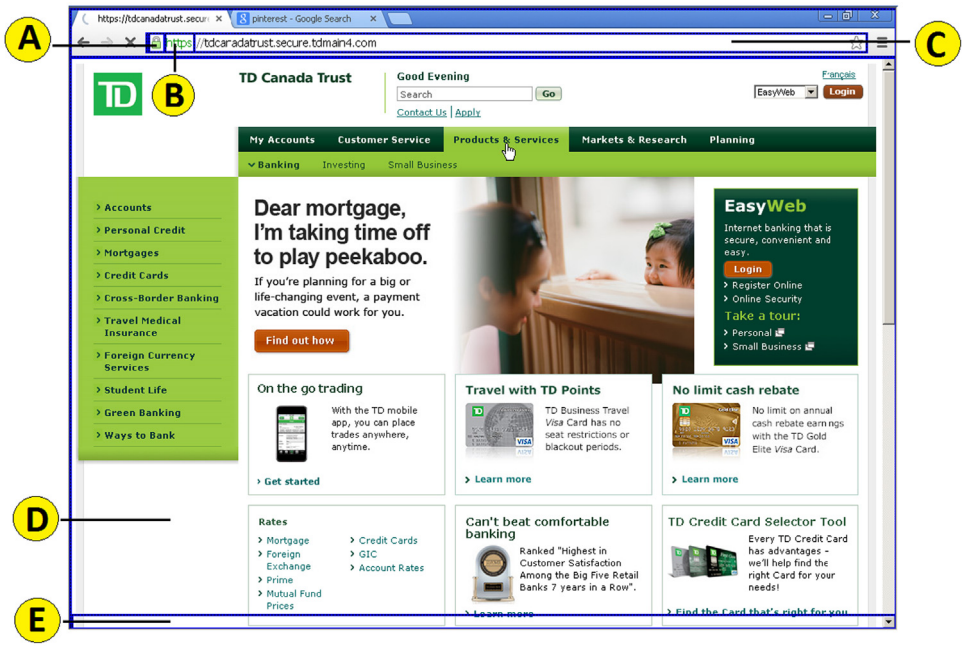

Fig. 3. Areas of interest (AOIs) for eye tracking.

Table 3
Success rate, mean time spent on all AOIs, mean time to make a decision per website, number of participants who decided a website was "legitimate" (out of 24) and their mean certainty level (out of 5), number of participants who decided a website was "phishing" (out of 24) and their mean certainty level. For phishing websites, success rate (TP – true positive) is the percentage of correctly identified phishing websites. For legitimate websites, success rate (TP) is the percentage of correctly identified legitimate websites. TP are identified in green and in bold roman. FN (false negatives) are identified in blue and in italic. FP (false positives) are identified in orange and in bold italic. (For interpretation of the references to colour in this table caption, the reader is referred to the web version of this paper.)

	Type	Website	Success rate (%)	AOI time (s)	Total time (s)	"Legitimate" answers		"Phishing" answers	
						Num	Certainty	Num	Certainty
Phishing	1	CIBC	33	42	73	*14*	4.71	**7**	4.29
	4	Carleton University	38	53	89	*13*	4.62	**8**	4.13
	1	cuLearn	38	55	89	*13*	4.62	**8**	3.88
	2	Paypal (IP)	38	42	77	*13*	3.85	**8**	4.25
	1	RBC	43	71	109	*12*	4.25	**9**	4.11
	1	HSBC	48	47	82	*11*	3.91	**10**	3.70
	1	Paypal (Favicon)	48	58	95	*11*	4.09	**10**	3.90
	1	Carleton Portal	48	62	105	*11*	4.82	**10**	3.10
	1	Scotiabank	52	46	89	*10*	4.20	**11**	4.36
	3	Amazon	62	49	90	*8*	4.50	**13**	4.38
	5	Facebook (popup)	62	31	61	*8*	4.38	**13**	4.62
	1	TD Bank	62	62	94	*8*	4.25	**13**	4.46
	1	Gov. of Canada (theft)	71	54	111	*6*	3.67	**15**	4.07
	6	Credit Card Checker	95	24	47	*1*	1.00	**20**	4.70
Legitimate	7	Research Survey	14	48	123	**3**	3.67	***18***	4.50
	7	Pinterest	67	47	89	**14**	4.00	***7***	3.86
	7	Kijiji	81	57	98	**17**	4.29	***4***	3.75
	7	LinkedIn (non-SSL)	81	55	96	**17**	4.06	***4***	4.25
	7	Gov. of Canada	86	52	97	**18**	4.50	***3***	2.67
	7	HSBC (legit)	86	50	83	**18**	4.44	***3***	2.67
	7	LinkedIn (SSL)	90	43	85	**19**	4.37	***2***	4.50
	7	Netflix	90	39	75	**19**	4.37	***2***	4.00
	7	Facebook	95	31	57	**20**	4.70	***1***	3.00
	7	Twitter	100	42	73	**21**	4.33	***0***	0.00

website was "legitimate" or "phishing" and the mean certainty level for each type of decision.

Participants took 87 s on average to make a decision whether the website is real or fake, with 48 s of that time spent specifically on the chrome and content AOIs. Levene's test to assess the homogeneity of variance was passed ($F=0.003$, $p=0.96$). T-test results showed that there was no statistically significant difference in the mean decision time between legitimate and fraudulent websites ($t=-0.13$, $p=0.90$, $\alpha=0.05$ two-tailed) with the mean time taken for the phishing websites (87 s) virtually identical to that of legitimate websites (88 s). The average success rate is 53% for the phishing websites and 79% for the legitimate websites, indicating that participants were relatively poor at detecting fraudulent websites even when primed to look for signs of deception. Participants were confident in their decisions, with a mean certainty level of 4.25, regardless of whether they had made the correct choice or not.

We next examined whether demographic characteristics may have influenced results and found no significant effects.

Gender: The mean performance score was 14.0 for males (σ=4.21) and 16.2 for females (σ=3.67). The homogeneity of variance between the two groups was confirmed using Levene's test (F=0.54, p=0.47). Once more the t-test showed no statistical significance between the scores for the male and female groups ($t = -1.305$, p=0.21, α=0.05, two-tailed).

Age: A Pearson's r was conducted to analyze the correlation between participants' ages and their scores. No statistical significance was found (N=21, $r = -0.24$, p=0.30, α=0.05, two-tailed).

Technical expertise: A technical proficiency score out of 5 was calculated for each participant, based on the number of "yes" responses given to the pre-test proficiency questions from Table 1. No statistically significant correlation was found between the participants' proficiency score and their performance score (N=21, r=0.19, p=0.40, α=0.05, two-tailed).

We also informally explored the performance of the four participants with engineering background to see if general technical experience impacted scores. We found no relationship. These four participants placed 5th, 10th, 16th, and 21st out of 21 participants with respect to their performance scores.

Familiarity with websites: Although all participants reported using online banking at least once a month, 52% of the participants failed to recognize phishing attempts with their own banking website. One participant was able to correctly identify all banking websites as phishing websites except for their own bank, whereas others were not able to recognize any of the banking websites as phishing attempts at all. We found no relationship between participants' ability to identify phishing attempts on their own banking website with their ability to identify phishing attempts on unfamiliar banking websites.

Although this is difficult to measure, we noticed that some participants were overconfident with their most familiar websites, such as with their own bank or cuLearn. They were quick to judge and make statements such as "I know this website, it looks exactly like the last time I used it" and seemed to take less precautions in comparison to the rest of the websites.

4.2. Eye tracking results

Table 4 shows the total percentage of time each participant spent looking at browser chrome elements versus webpage content, along with the total score of each participant. On average, participants spent 6% of the time looking at chrome AOIs, 9% of the time on other chrome elements, and 85% of the time looking at the webpage contents. A positive significant correlation was found between the time participants' spent looking at the chrome and performance scores (r=0.55, p=0.01, α=0.01, one-tailed). A positive correlation was also found between time specifically spent on the chrome AOIs and the performance scores (r=0.40, p=0.04, α=0.05, one-tailed). This suggests that more time spent observing the chrome, and the chrome AOIs in particular, led to an increased ability to correctly assess the legitimacy of websites.

Fig. 4 shows the time that each participant spent looking at specific elements in the browser chrome. Each chrome AOI is listed separately and the remainder of the chrome elements such as the back and forward buttons are included in the "other chrome" category. We exclude the chrome footer (Fig. 3, Label E) since Google Chrome does not reserve a space for the footer and only overlays data on top of the main page content. It was technically infeasible to distinguish eye tracking data in that area from the webpage content area (Fig. 3, Label D). When considering only the time spent looking at chrome elements, 2% of the time was spent looking at the lock icon, 5% was spent looking at the http/https area, 31% of the time was spent looking at the address bar, and the remaining time was spent on irrelevant chrome elements.

Table 4
Percentage of time spent looking at the browser's chrome AOIs, all chrome elements, and webpage content versus performance scores, ordered by score (out of 24).

Participant ID	Time (percentage) spent on			Score
	Chrome AOIs	All chrome	Website content	
09	4	13	87	9
19	4	8	92	9
16	1	4	96	10
01	2	6	94	11
06	3	7	93	11
07	6	12	88	12
02	11	18	82	13
18	4	14	86	13
17	2	4	96	15
20	3	5	95	15
03	15	25	75	16
08	2	8	92	16
11	6	18	82	17
21	3	16	84	17
10	10	20	80	18
13	1	5	95	18
14	5	32	68	18
04	3	16	84	20
12	4	11	89	20
05	13	43	57	21
15	12	23	77	22

4.3. Attention to URL

As shown in Fig. 4, all participants spent some time looking at the URL address bar according to the eye tracking data. Of the chrome AOIs, the address bar was most popular. However, not all users reported using the URL in their decision-making process. It is unclear from the data whether they glanced at it subconsciously, whether they discounted it when making a decision, or whether they simply forgot to mention it.

While 14 participants (67%) reported looking at the URL at least once, the degree to which they understood its contents and relied on it varied significantly. For example, one participant only got suspicious when they noticed after visiting a number of links that the main page had a different domain name from all the pages that it linked to. Another participant cited the URL as important, but wrongly stated that it could be "manipulated by cookies", and incorrectly classified both LinkedIn websites as phishing attempts. Two participants were comfortable with identifying websites as legitimate if they had "simple and clear" URLs such as netflix.com or twitter.com, but on the other hand they wrongly identified the banking websites as legitimate, even though they did not have "simple" URLs. Therefore, it became clear to us that although the majority of participants acknowledged the role of a website's URL as an important cue for identifying a website's legitimacy, they had many different interpretations of its meaning and level of importance.

Our most deceptive URL proved to be cartelon.ca, in place of carleton.ca. In fact, we used the same misspelling in two tests (cuLearn and Carleton Portal). Only three participants (14%) noticed the misspelling. However, despite this, cuLearn and Carleton Portal were correctly identified as phishing websites by 38% of participants. These additional participants labelled the websites as phishing for completely erroneous reasons. Two participants stated that the cuLearn URL should have been carleton.ca/culearn instead of culearn.carleton.ca, and a number of students identified the student portal as fraudulent for reasons such as outdated content (when in reality, the content was identical to the real website).

We frequently observed that since all of the links on our phishing websites redirected to legitimate websites, many participants

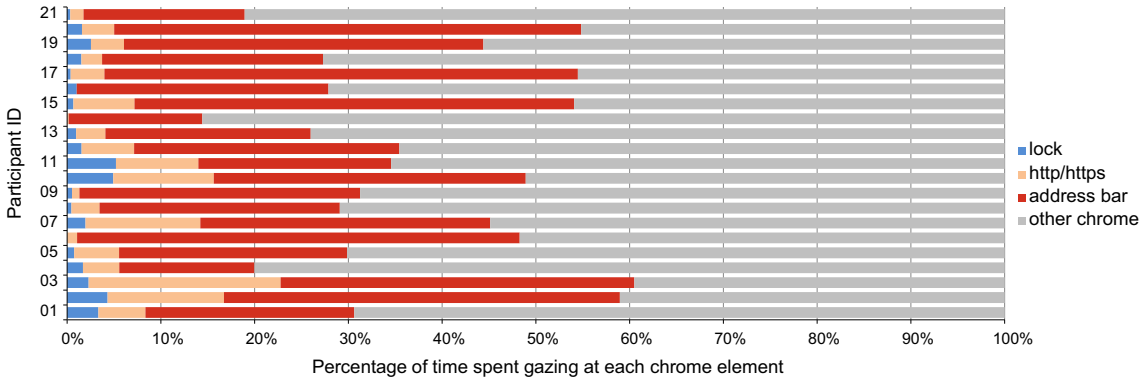

Fig. 4. Time spent looking at the lock, http/https, URL bar, and other chrome elements.

incorrectly classified these phishing websites as legitimate after having been redirected to the legitimate website. This re-emphasizes our eye tracking data and observations that many users infrequently observe the chrome, do not notice changes in information located in the chrome such as changing URLs, or do not understand the changes even when they do observe them.

4.4. Attention to SSL indicators

According to the eye tracking data, participants spent much less time observing SSL indicators such as the lock and https than domains (Fig. 4). Granted both of these can be observed relatively easily and a quick decision can be made as to whether the indicator is present.

Lock: The eye tracking data shows that 90% of users at least glanced briefly at the lock icon during the session, but it is unclear how many actually consciously noted its state. Twelve (57%) of our participants referred to the SSL lock at least once when justifying the legitimacy of a website. However, we observed a wide range of interpretations for the meaning of the lock. Interestingly, in the previous study by Dhamija et al. (2006), only 5 out of 22 participants knew the correct location of the SSL padlock in the browser.

We generated SSL certificates for all of the bank phishing websites (except Scotiabank, below), and therefore they all displayed a green padlock in the chrome. Since the legitimate Scotiabank home page does not support https and we included it in an iframe for our attack, this website displayed a grey padlock with a yellow yield sign indicating that there is unsecured content on the page. The participant who paid the most attention to the SSL locks incorrectly identified all of the banking websites as legitimate with the exception of Scotiabank, on which he cited the yellow yield sign as his reason for identifying it as a phishing website. Another participant stated that "the lock just means that there is a secure connection, but the server could still be harming you".

For the single website where we set the favicon to imitate an SSL lock, only one participant noticed it. They commented that the lock was in the wrong location and cited that as a reason for why the website was a phishing attempt. None of the other participants noticed the favicon.

Https: The eye tracker recorded 95% of participants looking at the https identifier. Ten participants (48%) reported actively checking for https in the URL and stated that its existence is a positive security indicator. One participant mentioned that https in an address bar would lead them to trust a website with their personal information.

When prompted during the post-session interview, three participants reported knowing that "the s in https means secure" but that they had completely forgotten to check for it during the experiment. One participant knew the details of the https protocol implementation and that "https is just the server securing itself. Server could still be doing something that can harm you". This participant still completely overlooked this indicator and forgot to check during the experiment. Four participants (19%) specifically reported not knowing what https means or what it does.

We found that 80% of participants who reported knowing the importance of https prepended to a URL did not truly understand what it implies. When asked, none could explain how https could potentially be used at both phishing and legitimate websites.

EV certificate highlighting: Four (19%) participants referred to the EV certificate at least once. The first participant (the same one who trusted all of the banking websites except for Scotiabank) stated that when he goes on Paypal he usually sees the company name in a green box, and he cited the absence of the box as his reason for correctly identifying the site as a phishing attempt. Two participants mentioned the EV certificate as one of their reasons for trusting the legitimate Twitter website. However, in our debriefing sessions, none of the participants knew the difference between an EV certificate and a regular certificate until it was explained to them.

Certificate warning: Fifteen (71%) participants (incorrectly) did not proceed when confronted with a certificate warning when viewing the Carleton Research Survey link. When questioned, one participant stated that if he was browsing a trusted website such as Carleton, and he arrived at an SSL certificate warning page through an internal link, he would most likely proceed. On the other hand, if he arrived at the page through an external link such as through a search engine or an e-mail from a friend, he would turn back. Another participant stated that he would contact the person who sent the link. Although not completely secure, these are likely reasonable strategies given that most of the time users will see such warnings for a legitimate page. The remaining participants stated that if they "knew where they were going", they would proceed, or otherwise turn back. These responses are potentially more problematic because it is unclear how users would assess their destination.

4.5. Strategies employed by users

We discuss qualitative data representing the general strategies that we observed our participants using to identify phishing attempts. We analyzed our observation notes and the transcripts of the sessions where participants explained their decision-making process. We coded the emerging strategies and grouped them into broader categories.

Many participants were inconsistent with the way in which they applied their strategies. For example, some participants commented on the simplicity of a URL as their reason for trusting one website, but did not raise any objections for other websites with clear alterations to the URL.

The strategies are presented in order of popularity, with Strategy A being by far the most common strategy. However, given that participants changed strategies throughout their sessions and may not have reported them each time, it is infeasible to accurately classify

exactly how many participants used each strategy. Furthermore participants sometimes combined two distinct strategies during their assessments.

Strategy A: *Website content*. Some participants judged the websites purely based on content, despite glancing at the chrome according to the eye tracking data. For example, they cited the layout and professional appearance of the website to support their decision. One participant repeatedly stated "this is what I would expect a banking website to look like". Two participants also paid attention to links to Facebook and Twitter, with one participant saying that since the website "has a lot of likes on Facebook, it means that a lot of people have been here so it should be trustworthy". These participants explored the website and casually clicked on some of the links to gain a better idea of the website. Participants who solely relied on this strategy were more susceptible to falling for phishing websites that closely matched the look of their legitimate counterparts. Participants usually reported that the phishing website looks "very familiar" or "just like the one I've seen this morning", and that they saw no alarming difference. Clearly, this approach is relatively naive and leaves users susceptible to attack. Unfortunately, it was the most popular strategy.

Strategy B: *Brute-forcing website functionality*. Some participants tested a number of features on each website. They were actively trying to find glitches and hints that the website was broken. For example, they tried functionality such as changing the language, changing locations, trying the mobile version of the website, testing the forms and login fields with false information, clicking on many different links on the main page, and searching for contact information. These participants performed better than we expected. While we paid great attention to detail when developing the phishing websites, there were small details on a number of the test websites that we missed (e.g., the ability to change languages) and these raised suspicion. We suspect that the majority of phishers would not spend a greater level of effort than we did in designing their phishing websites, which would likely make this a relatively effective defense strategy in real life if users adhered to this practice regularly but this seems unlikely given the amount of effort involved.

Strategy C: *Paying attention to the URL*. While many participants reported paying attention to the URL, all of them used this strategy in combination with other strategies. Participants either tried to recall familiar URLs or used heuristics such as assessing the simplicity of the URL. Those who questioned the URL often mentioned that the "URL looks weird" or that "it contains random stuff". One participant correctly noticed that the domain changed every time she was redirected from a phishing website to a legitimate website by clicking on links and deemed this behaviour malicious. However, most participants reported that they did not recognize the difference between a domain and a sub-domain, and were unable to identify the important parts of a URL.

Strategy D: *Using a search engine*. Three (14%) participants used Google to search for the website being tested, and compared both the URL and page content with the search results. Each of these participants also applied Strategy B, actively trying to find flaws in the websites. These participants performed the best, with an average score of 22.3. In fact, two participants could have potentially scored even higher, but they did not begin using this strategy until the third or fourth website. In practice, users are unlikely to apply this strategy to every website they visit. They may verify a website if they become suspicious, but this relies on users first recognizing questionable websites.

Strategy E: *Exclusive reliance on SSL indicators*. Two participants reported that they based their decisions exclusively on SSL indicators. Eye tracking data shows considerable time spent on the webpage content as well, so it is unclear whether this subconsciously also influenced decisions. These participants missed cues such as the misspelled URL (cartelon.ca) or the overlaid popups because they assumed that a signed certificate meant that the website was safe. This strategy highlights misunderstandings that occur when interpreting security indicators. It also highlights how phishing attacks can trick users by intentionally using the exact visual cues that are meant to indicate security.

4.6. Challenges and limitations

We faced a number of challenges during our study. Since security in general is a secondary task, it is difficult to create an ecologically valid scenario while collecting eye tracking data and not putting users' personal data at risk. We chose a controlled lab experiment where users were primed to complete the security task, clearly sacrificing ecological validity but allowing us to collect more precise data. Our results provide an upper-bound on users' ability to detect phishing websites, which is somewhat concerning given that users correctly identified only 53% of phishing websites on average. Moreover, participants frequently identified phishing websites correctly but for completely wrong reasons. For example, they would say that the content on the page was outdated or poorly designed, when in fact we had constructed an exact replica of the website in question. We also note that users' technical proficiency scores provide only a rough approximation of users' technical experience and that a more thorough assessment may lead to different outcomes.

Further investigation is needed to determine how to effectively use eye tracking data in phishing studies. First, given that some of the AOIs are relatively small on the screen, the eye tracking margin of error may have impacted results. Furthermore, dwell times on an area of the screen do not necessarily reflect users' level of understanding of security cues. Conversely, a short fixation on an area does not necessarily indicate that a user has missed that element. While we analyzed how long participants fixated on various areas of interest and supplemented this data with our observations and participant comments, there may also be different ways of interpreting the data.

The use of Windows XP on our eye tracking computer caused some minor technical issues. One participant noticed a visual glitch in the title bar of the browser, which was caused by Windows XP. Another participant also became suspicious when she noticed that the alert boxes produced by the browser looked slightly different than on her computer, not knowing that the difference was due to the operating system. A final issue related to the way in which Windows handles multiple monitors. The desktop was equipped with two displays, with the main display being used by the experimenters and the secondary display (containing the eye tracker) being used by the participant. Since Windows does not display a taskbar on the secondary display, this was one less visual cue for participants when they viewed phishing websites with a popup window. Windows 7, however, groups all windows into a single taskbar button, so it is not clear whether the taskbar would have helped in this regard.

5. Discussion

In this section, we present our thoughts with respect to the different types of attacks tested, discuss how our study compares to previous work, reflect on the usefulness of including eye tracking data in this type of study, and suggest a few recommendations for future designs.

5.1. Performance compared to expectations

For each type of phishing website tested, we now discuss whether the results of our study matched our initial expectations. First, we

found (see Table 3) that users quite consistently spent nearly a minute specifically examining the websites (chrome and content AOIs) for clues about their legitimacy and spent additional time on other aspects before coming to a decision. Based on this and on our observations, participants were really paying attention and trying to make correct decisions. Participants were capable of detecting certain types of phishing attacks, but the majority of attacks went undetected even in our study where users were specifically attempting to detect phishing websites.

Type 1: *Spoof website with incorrect URL*. Although users spent more time looking at the URL than any other chrome security indicator, it did not necessarily translate into sound phishing determinations. Participants did poorly in identifying fraudulent banking websites (even their own bank) and did not recognize erroneous Carleton URLs. Fewer than half noticed the fake Paypal website with the lock Favicon. Participants were most successful at recognizing that the Government of Canada website asking them enroll in Identity Theft protection was malicious, although this may be because the website asked for a social insurance number rather than because participants recognized a suspicious URL.

Type 2: *Spoof website with IP address in URL*. We had expected that most participants would recognize this well-established phishing technique. However, only 38% of participants noticed that the Paypal website had an IP address instead of a URL and deemed this suspicious. This is particularly concerning because this type of attack is fairly apparent even with little knowledge of how to parse URLs. This is also a common tip given to users on how to detect phishing attacks; it seems that the message has not reached this audience or is an unreasonable expectation of users.

Type 3: *Fake chrome*. 62% of participants recognized that the Amazon website was a phishing attempt. We had expected a higher success rate given the obvious double chrome and URL. It appears that many users still do not have clear expectations of how the chrome should look and do not find a double URL bar to be suspicious. This may be due to the popularity of browser toolbars; users are so accustomed to a cluttered chrome that they have little expectation of what *should* be there.

Type 4: *Popups asking for credentials*. Participants were even more likely to be fooled by the fake popup login window appearing on the Carleton website (38% success rate), despite the fact that the window had a clearly erroneous URL. Participants spent nearly a minute examining the page content and chrome but still incorrectly deemed them trustworthy.

Type 5: *Overlaid popup windows*. Once again, participants saw a double URL bar as the popup phishing Facebook page covered all of the real page content. The legitimate URL was displayed in the main chrome, but a fake URL was visible in the popup window. We found that 62% of participants recognized this attack. Although this was one of the higher detection rates for phishing websites, it is still lower than expected given that 86% of our participants were regular Facebook users.

Type 6: *Fraudulent based on context*. 95% of participants successfully identified the Credit Card Checker as a phishing website. We suspect that most users have heard about phishing through the media and that the typical message is to be careful of websites asking for credit card numbers or personal information. It is encouraging to note that all but one user recognized this scam, but it is also worrying that none of the other attacks were similarly detected.

Type 7: *Legitimate websites*. Eight out of 10 legitimate websites had success rates of over 80%. This emphasizes that users may be better at recognizing when things appear “normal” than at identifying when things are suspicious. Especially noteworthy is the Research Survey website. Although the Research Survey was a legitimate website, it had several indicators that made it questionable: it was hosted at a relatively obscure and unfamiliar sub-domain of Carleton.ca, searching with Google did not return any related search results, and it warned users about a self-signed certificate. This website serves to emphasize that these decisions are sometimes nearly impossible for users to make; phishing websites give no warnings and appear familiar, while legitimate websites may prompt warnings and be relatively unpopular. In this particular case, participants were wise to be cautious.

5.2. Comparison with previous work

We set out to investigate whether improvements to browser security indicators and phishing awareness over the last decade have resulted in decreased susceptibility to phishing for everyday users. Nearly a decade ago, Dhamija et al. (2006) demonstrated that users were easily fooled and that security indicators needed improvement. Their study relied on observation and user comments to determine how users made decisions.

In our study, we additionally collected eye tracking data to help understand how users determine the legitimacy of websites. Overall, our results show that users still pay very little attention to security indicators, give most weight to website content when making decisions, and are frequently deceived by phishing websites. Our eye tracking data provides empirical evidence that participants spend only 6% of their time glancing at security indicators, with most of that time spent examining the URL. Interestingly, our participants were also very confident in their decisions, with similar certainty scores to Dhamija et al.'s study.

We noted some progress since the earlier study, in particular with respect to knowledge about phishing. Dhamija et al. found that some participants were unaware that spoofing websites was even possible. In contrast, our participants were all familiar with the idea of phishing and knew that this type of attack was possible. Browsers have made considerable changes to their security indicators to help users detect phishing and these appear to have had some effect on users' ability to detect phishing. Our participants performed modestly better than in the earlier study, with an average success rate of 64% compared to 58% in Dhamija et al.'s study. A direct comparison is difficult since the test websites differed, however, we believe that our websites were similarly representative of phishing attacks. It is also unclear how much of this improvement is due to improved browser interfaces as opposed to increased awareness of the threat.

5.3. Usefulness of eye tracking data

We found the eye tracking data to be a valuable source of information for this particular context. Given the already artificial scenario of asking users to detect phishing websites, relying on users' self-reflection of their decision-making process seemed problematic since they were aware that we were looking for specific behaviours and may have modified their answers accordingly. The eye tracker provided a more direct source of data. Participants may still have paid more attention to security indicators than they would normally, but they actually had to look at the indicators rather than simply telling us that they had. The eye tracker also captured passing glances that may have helped in the decision making process even if participants were unaware.

Our eye tracking results show that users paid little attention to security indicators. It is worth noting that even when they see the indicators, there is no guarantee that this translates into useful information for the user. Participants may not have known how to interpret what they were seeing or may have misunderstood its meaning. Secondly, some of our AOIs were small and the accuracy of the eye tracker may have impacted the results. However, overall we found it to be a useful tool for assessing phishing susceptibility when combined with our other data collection methods. The

different sources of data largely aligned with each other, giving us some confidence in the reliability of our results.

5.4. Recommendations

The majority of our participants were pleased to have completed our study, and expressed their gratitude for having learned how to better protect themselves online. They were very willing to provide us feedback both during and immediately after our session. Based on participant feedback, our observations, and on the eye tracking data, we devised some suggestions which we believe have a potential to improve the usability of web browsers for identifying phishing websites. These suggestions should be investigated in future work, since further user studies would be required to assess their effectiveness.

User-friendly URLs: Although the majority of participants at least occasionally attempted to use the URL, they did not have enough knowledge of the structure of URLs to make an informed decision. In fact, only one participant understood how a sub-domain worked. The remainder were surprised during the debriefing session when we informed them that anybody could buy an arbitrary domain such as evil.com and then set up a subdomain such as paypal.evil.com with little effort. Therefore, we believe that the URL bar should be made more user-friendly. Current browsers (and the version of Chrome used in this study) differentiate the top-level domain by colouring it black and the remainder of the URL in grey. None of our participants noticed this visual cue. We suggest that the domain names needs to be significantly more visually distinct in order to be effective as a security cue. Alternatively, browsers could use "breadcrumbs", as in file managers of many operating systems (e.g., Windows 7 and Ubuntu). In this way, the actual domain name of the website could be displayed more prominently, and users who wish to view the entire URL could simply click on the URL bar.

Legitimate websites that use different domains for different sections of their website are also problematic. In these cases, users become accustomed to ignoring the URL since it provides no useful information to them. If inspection of URLs is to be advocated as a way of detecting phishing, then URLs must be reliable and consistent to avoid training users to avoid this cue.

Visual aids for browsing: Many participants made decisions on a phishing website's legitimacy while they were actually examining the legitimate website itself. This was because all of the links on our phishing pages pointed to the original website. Therefore, we believe that it would be beneficial to develop an indicator which informs the user when they move from one domain to another. We concede that it is difficult to come up with an indicator that is both effective and non-obtrusive. Many websites warn users when they click on a link that leads to an external website. This addresses the reverse issue, where fraudulent links may be posted on a legitimate website. Nevertheless, if this idea were implemented on every website, it would be much more cumbersome to browse the web. We believe that a better implementation of this suggestion could instead make use of a visualization (possibly in the browser's chrome) where it is easy for users to notice changes even if they are not paying close attention. This approach would also avoid situations where phishing websites intentionally leave out warnings about navigating to a different domain.

Moving authentication to the chrome: Although we have seen that users pay some attention to the browser chrome in a controlled lab study environment, they may not always be as attentive during regular Internet use. One way of reinforcing the trustworthiness and importance of the chrome would be to move some important tasks to the chrome, such as user authentication. Moving authentication to the chrome may make it easier to inform the user about the legitimacy of the current website. This is, however, a difficult task which faces many obstacles due to the extensive collaboration that would be required between browser developers and web developers. Mozilla was pursuing a browser-based authentication approach with Mozilla Persona (Mozilla, 2014), but this no longer appears to be under active development.

Automate as much as possible: Our results confirm that identifying phishing at the user interface is an extremely difficult task. A decade of improvements in browser security indicators and in user education campaigns has yielded only a 6% increase in detection rates by users in the best case. Each of the cues and heuristics that users may use to identify fraudulent websites can also be used maliciously to deceive users in phishing attacks. If we cannot provide users with tools and cues that are sufficiently reliable, then users should not be entrusted with making these decisions.

6. Conclusion

We have conducted an empirical study to gather evidence on what strategies users employed to determine the legitimacy of websites. We have relied both on self-reported qualitative data and on eye tracking data to determine whether improved browser security cues have led to improved ability to detect phishing attacks. Contrary to Vishwanath et al.'s suggestion, users were unable to reliably detect phishing even when alert and vigilant. We found that even in our controlled lab environment, participants had an average success rate of 53% for identifying phishing websites, essentially the same as if users took a random guess. When making decisions about the legitimacy of websites, participants spent only 6% of the time looking at security indicators in the browser chrome and 85% of the time looking at the contents of the webpage.

We found a correlation between participants' performance scores and the time which they spent looking at chrome elements, indicating mildly positive results for security indicators. Other variables, such as users' general technical proficiency, did not correlate with improved performance scores. The most effective strategy for detecting phishing websites combined searching for the website using a search engine and testing for broken website functionality. However, the vast majority of users still relied primarily on the superficial appearance of the website content. Although modest improvements were observed compared to Dhamija et al. (2006)'s earlier work, we find that existing browser cues remain insufficient to help users protect themselves against phishing. We identified areas for potential user interface improvements, and made recommendations for future designs.

Acknowledgements

Sonia Chiasson holds a Canada Research Chair in Human Oriented Computer Security. She acknowledges the lNatural Sciences and Engineering Research Council of Canada (NSERC) for funding the Chair and her Discovery Grant. Furkan Alaca acknowledges a graduate scholarship from NSERC.

References

Abbasi, A., Zahedi, F., Chen, Y., 2012. Impact of anti-phishing tool performance on attack success rates. In: International Conference on Intelligence and Security Informatics (ISI), IEEE, Arlington, USA, pp. 12–17.

Anti-Phishing Working Group, 2014a. Phishing Activity Trends Report—4th Quarter 2013. ⟨http://apwg.org/resources/apwg-reports⟩.

Anti-Phishing Working Group, August 2014b. ⟨http://apwg.org/⟩.

Anti-Phishing Working Group, August 2014c. ⟨http://phish-education.apwg.org/r/about.html⟩.

Arachchilage, N.A., Love, S., Scott, M.J., 2012. Designing a mobile game to teach conceptual knowledge of avoiding 'phishing attacks'. Int. J. e-Learn. Secur. 2, 127–132.

Arachchilage, N.A.G., Love, S., 2013. A game design framework for avoiding phishing attacks. Comput. Hum. Behav. 29, 706–714.

Bank of America, August 2014. Sitekey. ⟨https://www.bankofamerica.com/privacy/online-mobile-banking-privacy/sitekey.go⟩.
Bergholz, A., De Beer, J., Glahn, S., Moens, M.F., Paaß, G., Strobel, S., 2010. New filtering approaches for phishing email. J. Comput. Secur. 18, 7–35.
Burns, M.B., Durcikova, A., Jenkins, J.L., 2013. What kind of interventions can help users from falling for phishing attempts: a research proposal for examining stage-appropriate interventions. In: 46th Hawaii International Conference on System Sciences, IEEE, Wailea, USA.
Chou, N., Ledesma, R., Teraguchi, Y., Mitchell, J.C., 2004. Client-side defense against web-based identity theft. In: Proceedings of Network and Distributed System Security (NDSS) Symposium, Internet Society, San Diego, USA.
Cormack, G.V., 2008. Email spam filtering: a systematic review. Found. Trends Inf. Retr. 1, 335–455.
Dhamija, R., Tygar, J.D., 2005. The battle against phishing: dynamic security skins. In: Proceedings of the Symposium on Usable Privacy and Security (SOUPS), ACM, Pittsburgh, USA, pp. 77–88.
Dhamija, R., Tygar, J.D., Hearst, M., 2006. Why phishing works. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI), ACM Press, Montreal, Canada.
Downs, J.S., Holbrook, M., Cranor, L.F., 2007. Behavioral response to phishing risk. In: Proceedings of the APWG eCrime Researchers Summit, ACM, Pittsburgh, USA.
Egelman, S., 2009. Trust Me: Design Patterns for Constructing Trustworthy Trust Indicators, (Ph.D. Thesis), University of California Berkeley.
Fette, I., Sadeh, N., Tomasic, A., 2007. Learning to detect phishing emails. In: Proceedings of the International Conference on World Wide Web (WWW), ACM, Banf, Canada, pp. 649–656.
Fu, A.Y., Wenyin, L., Deng, X., 2006. Detecting phishing web pages with visual similarity assessment based on earth mover's distance (emd). Trans. Dependable Secur. Comput. 3, 301–311.
Garfinkel, S.L., Margrave, D., Schiller, J.I., Nordlander, E., Miller, R.C., 2005. How to make secure email easier to use. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI), ACM, Oregon, USA, pp. 701–710.
Government of Canada, August 2014. ⟨https://www.getcybersafe.gc.ca/index-eng.aspx⟩.
Herzberg, A., 2009. Why Johnny can't surf (safely)? attacks and defenses for web users. Comput. Secur. 28, 63–71.
Hong, J., 2012. The state of phishing attacks. Commun. ACM 55, 74–81.
Jagatic, T.N., Johnson, N.A., Jakobsson, M., Menczer, F., 2007. Social phishing. Commun. ACM 50, 94–100.
Jakobsson, M., Tsow, A., Shah, A., Blevis, E., Lim, Y.K., 2007. What instills trust? A qualitative study of phishing. In: Proceedings of the International Workshop on Usable Security (USEC), Springer-Verlag, Scarborough, Trinidad/Tobago, pp. 356–361.
Kirda, E., Kruegel, C., 2006. Protecting users against phishing attacks. Comput. J. 49, 554–561.
Kirlappos, I., Sasse, M.A., 2012. Security education against phishing: a modest proposal for a major rethink. IEEE Secur. Priv. 10, 24–32.
Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M.A., Pham, T., 2009. School of phish: a real-world evaluation of anti-phishing training. In: Proceedings of the Symposium on Usable Privacy and Security (SOUPS), ACM, Pittsburgh, USA.
Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., Cranor, L.F., Hong, J., 2007. Getting users to pay attention to anti-phishing education: evaluation of retention and transfer. In: Proceedings of the Anti-phishing Working Group's Annual eCrime Researchers Summit (eCrime), ACM, Pittsburgh, USA, pp. 70–81.
Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L.F., Hong, J., 2010. Teaching johnny not to fall for phish. ACM Trans. Internet Technol. 10, 7:1–7:31.
Lee, L.H., Lee, K.C., Juan, Y.C., Chen, H.H., Tseng, Y.H., 2014. Users' behavioral prediction for phishing detection. In: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, pp. 337–338.
Li, L., Berki, E., Helenius, M., Ovaska, S., 2014. Towards a contingency approach with whitelist-and blacklist-based anti-phishing applications: what do usability tests indicate?. Behav. Inf. Technol., 1–12.
Li, L., Helenius, M., 2007. Usability evaluation of anti-phishing toolbars. J. Comput. Virol. 3, 163–184.
Lin, E., Greenberg, S., Trotter, E., Ma, D., Aycock, J., 2011. Does domain highlighting help people identify phishing sites? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI), ACM, pp. 2075–2084.
Mayhorn, C.B., Nyeste, P.G., 2012. Training users to counteract phishing. Work: A Journal of Prevention, Assessment and Rehabilitation 41, 3549–3552.
Moore, T., Clayton, R., 2007. Examining the impact of website take-down on phishing. In: Proceedings of the Anti-phishing Working Group's Annual eCrime Researchers Summit, ACM, pp. 1–13.
Mozilla, August 2014. ⟨https://www.mozilla.org/en-US/persona/⟩.
NSS Labs, 2013. Evolutions in Browser Security: Trends in Browser Security Performance. Technical Report. NSS Labs.
OpenDNS, August 2014. ⟨http://www.phishtank.com/⟩.
Purkait, S., 2012. Phishing counter measures and their effectiveness-literature review. Inf. Manag. Comput. Secur. 20, 382–420.
Ross, B., Jackson, C., Miyake, N., Boneh, D., Mitchell, J.C., 2005. Stronger password authentication using browser extensions. in: Usenix Security Symposium, pp. 17–32.
Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L.F., Downs, J., 2010. Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI), ACM, Atlanta, USA, pp. 373–382.
Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L.F., Hong, J., Nunge, E., 2007. Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish. In: Proceedings of the Symposium on Usable Privacy and Security (SOUPS), ACM, pp. 88–99.
Sheng, S., Wardman, B., Warner, G., Cranor, L., Hong, J., Zhang, C., 2009. An empirical analysis of phishing blacklists. In: Proceedings of the Conference on Email and Anti-Spam (CEAS).
Sobey, J., Biddle, R., Oorschot, P., Patrick, A.S., 2008. Exploring user reactions to new browser cues for extended validation certificates. In: Proceedings of the 13th European Symposium on Research in Computer Security (ESORICS), Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, pp. 411–427.
Vishwanath, A., Herath, T., Chen, R., Wang, J., Rao, H.R., 2011. Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. Decis. Support Syst. 51, 576–586.
Whalen, T., Inkpen, K.M., 2005. Gathering evidence: use of visual security cues in web browsers. In: Proceedings of Graphics Interface (GI), Canadian Human–Computer Communications Society, Victoria, Canada, pp. 137–144.
Whittaker, C., Ryner, B., Nazif, M., 2010. Large-scale automatic classification of phishing pages. In: Network and Distributed System Security Symposium (NDSS).
Whitten, A., Tygar, J.D., 1999. Why Johnny can't encrypt: a usability evaluation of PGP 5.0. In: Proceedings of the 8th USENIX Security Symposium, USENIX Association, Berkeley, CA, USA.
Workman, M., 2008. Wisecrackers: a theory-grounded investigation of phishing and pretext social engineering threats to information security. J. Am. Soc. Inf. Sci. Technol. 59, 662–674.
Wright, R.T., Marett, K., 2010. The influence of experiential and dispositional factors in phishing: an empirical investigation of the deceived. J. Manag. Inf. Syst. 27, 273–303.
Wu, M., Miller, R.C., Garfinkel, S.L., 2006. Do security toolbars actually prevent phishing attacks? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI), ACM, Montreal, Canada, pp. 601–610.
Yahoo! Inc, August 2014. Personalized Sign-in Seals. ⟨https://protect.login.yahoo.com/⟩.
Yee, K.P., Sitaker, K., 2006. Passpet: convenient password management and phishing protection. In: Proceedings of the Symposium on Usable Privacy and Security (SOUPS), ACM, Pittsburgh, USA, pp. 32–43.