Received 27 July 2015
Accepted 30 September 2015

Spot the phish by checking the pruned URL

Melanie Volkamer
Department of Computer Science, Technische Universität Darmstadt, Darmstadt, Germany

Karen Renaud
School of Computing Science, University of Glasgow, Glasgow, UK, and

Paul Gerber
Work and Engineering Psychology Research Group, Technische Universität Darmstadt, Darmstadt, Germany

Abstract

Purpose – Phishing is still a very popular and effective security threat, and it takes, on average, more than a day to detect new phish websites. Protection by purely technical means is hampered by this vulnerability window. During this window, users need to act to protect themselves. To support users in doing so, the paper aims to propose to first make users aware of the need to consult the address bar. Moreover, the authors propose to prune URL displayed in the address bar. The authors report on an evaluation of this proposal.

Design/methodology/approach – The paper opted for an online study with 411 participants, judging 16 websites – all with authentic design: half with legitimate and half with phish URLs. The authors applied four popular widely used types of URL manipulation techniques. The authors conducted a within-subject and between-subject study with participants randomly assigned to one of two groups (domain highlighting or pruning). The authors then tested both proposals using a repeated-measures multivariate analysis of variance.

Findings – The analysis shows a significant improvement in terms of phish detection after providing the hint to check the address bar. Furthermore, the analysis shows a significant improvement in terms of phish detection after the hint to check the address bar for uninitiated participants in the pruning group, as compared to those in the highlighting group.

Research limitations/implications – Because of the chosen research approach, the research results may lack generalisability. Therefore, researchers are encouraged to test the proposed propositions further.

Practical implications – This paper confirms the efficacy of URL pruning and of prompting users to consult the address bar for phish detection.

Originality/value – This paper introduces a classification for URL manipulation techniques used by phishers. We also provide evidence that drawing people's attention to the address bar makes them more likely to spot phish websites, but does not impair their ability to identify authentic websites.

Keywords Information security, Individual behaviour

Paper type Research paper

Information & Computer Security
Vol. 24 No. 4, 2016
pp. 372-385
© Emerald Group Publishing Limited
2056-4961
DOI 10.1108/ICS-07-2015-0032

1. Introduction

Phishing messages offer a bait embedded in alluring text which entices the recipient to click. If they do click, it redirects them to a *doppelgänger* website. It is not possible to prevent phishing messages by using technical means, as cyber criminals innovate and change techniques continuously (Trend Micro Incorporated, 2016). Blacklists, one of the most common techniques, struggle to stay current, as the number of new phishing attacks per day escalates with each passing year (APWG Internet Policy Committee, 2014). Moreover, in 2013, the Anti-Phishing Working Group reported that it took, on average, 28.75 h to detect new phish websites (APWG Internet Policy Committee, 2013).

To reduce the success of phishing attacks, the following requirements need to be met:

- Users should consult the address bar to check the URL.
- Such checks should deliver value: it should be possible for users to detect phish URLs and confirm the legitimacy of authentic web pages.

Both requirements are challenging to achieve. In the first place, many users do not seem to know that they ought to validate the integrity of the URL before divulging personal details, as confirmed by several studies (Egelman *et al.*, 2008; Wu *et al.*, 2006). The fact that phish websites still use arbitrary URLs or IP addresses (APWG Internet Policy Committee, 2013) suggests that phishers know that many users will not consult the address bar to check the URL.

Ensuring that the checking process delivers value is also challenging. It is important, to maximise phish detection, for users to realise that the domain name is the signal that reveals the legitimacy of the URL. Everything else is essentially noise. The fact that phishers often add a great deal of noise to their URLs [56.76 per cent of attacks in 2014 (APWG Internet Policy Committee, 2014)] means that they exploit the fact that many users do not understand that URLs contain both signal and noise. For example, phishers use www.amazon.com.books-online.de, and users might believe that it is a legitimate Amazon URL, because it contains the correct brand name, particularly if they are accustomed to reading from left to right. One other reason for users failing to detect that the URL is phish might be that the checking process is too cursory, failing to detect small manipulations. In this case, the signal is faulty but is so similar to the legitimate signal that they fail to detect the difference. This might be especially effective (for the phisher) if a great deal of noise is added so as to obscure the change to the signal. For example, phishers replace the “m” in Amazon with “rn” and add noise like “secure” (www.secure.arnazon.com/secure-login.html).

To support people in identifying the signal, many Web browsers highlight the domain part of the URL, attempting to focus the user’s attention on the actual signal. However, its introduction only slightly improved phish detection, as shown in (Lin *et al.*, 2011). The failure to deliver a large advantage could be due to the two requirements for phish detection not being met. Users might not have consulted the address bar at all or be unaware of the fact that the highlighted part is the signal when it comes to phish detection. To address the first requirement, we could ensure that users consult the address bar by making them aware of the need to do so. To ensure that the second one is satisfied to a maximum extent, we propose removing the noise, essentially pruning the URL: https://books-online.de and https://arnazon.com in the above examples.

We conducted an online study with 411 participants to test the effectiveness of focusing people’s attention on the address bar and pruning the URL. We concluded that combining the approaches delivers a large improvement in phish detection (Figure 1).

Consequently, we recommend that Web browsers provide a pruned URL and implement mechanisms to draw the user’s attention to the URL to maximise the likelihood of phish detection.

2. Related work

Our review of the literature suggests two ways of reducing the success rates of phishing attacks: prevention and mitigation.

Prevention can be uniform or tailored. The former would remove all links from electronic messages by technical means. As the inclusion of links is so prevalent in common message usage, this is unlikely to be feasible. The slightly less authoritarian option is blacklisting. Phish websites could be detected and blacklisted so that browsers are alerted. Several ways of detecting phishing messages and websites have been investigated (Bar-Yossef *et al.*, 2009; Marchal *et al.*, 2012; Maurer and Herzner, 2012; Prakash *et al.*, 2010). Browsers could then block the sites (post-click) or display warnings so that the user is alerted to the strong possibility that the link is suspect. This solution seems viable but faces some challenges. First, such warnings are usually active to ensure that users pay attention by forcing them to dismiss the message. Active warnings can become annoying, and anything that annoys can hardly be considered helpful or effective. The problem is that such automated warning systems sometimes get things wrong, and this risks users gradually losing trust in them. Active warnings, then, also risk habituating users, as do other security warnings (Li and Helenius, 2007; Maurer *et al.*, 2011; Wu *et al.*, 2006; Zhang *et al.*, 2007), leading to users ignoring them over time (Akhawe and Felt, 2013; Egelman *et al.*, 2008). Second, according to PhishTank, phish websites are only detected after a significant amount of time (APWG Internet Policy Committee, 2013). In effect, users who rely on blacklists are still vulnerable during this pre-discovery window. Mitigation approaches are a necessity during this gap, as a complementary mechanism.

Mitigation can be achieved via awareness and educational drives, combined with different ways of simplifying the check. A number of researchers have focused on training users to spot phishing attacks (Alnajim and Munro, 2009; Canova *et al.*, 2014; Jansson and von Solms, 2011; Kirlappos and Sasse, 2012; Kumaraguru *et al.*, 2007; Kumaraguru *et al.*, 2007, Sheng *et al.*, 2007). Educational approaches deliver value, but there is an issue of people being willing to spend uninterrupted time engaged in educational programmes, which many may be unwilling to do, particularly if they are unaware that they are at risk. Awareness drives take time to raise awareness in the general population and need to be sustained (Parija *et al.* 2014). Moreover, they are not targeted towards those at risk: they raise awareness in a context-neutral fashion, not directly when people are engaged in the potentially risky activity. Although the “teachable moment” idea proposed by Kumaraguru *et al.* (2009) addresses the context neutrality of the education by delivering the message as and when people access a phish

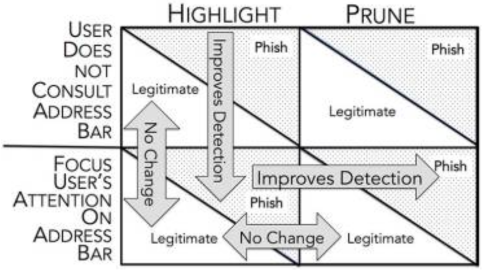


Figure 1.
Consequences of focusing users’ attention on the address bar and URL pruning

URL, it is still time-consuming, and the benefit of the educational message may be limited because of the innate complexity of the phish detection process.

There are proposals to reduce the complexity of the phish detection process by simplifying URLs. Three approaches are worth mentioning here:

(1) Domain highlighting highlights the domain name and is currently commonly implemented in many current Web browsers. Empirical tests have shown that it is not as effective as hoped (Lin *et al.*, 2011).

(2) Highlighting the address bar in the case of extended validation certificates is also unsatisfactory, because many authentic pages do not have extended validation certificates, so the user cannot rely on this as an infallible indicator of integrity.

(3) iOS in Safari removes the path. Chrome, too, implements a similar scheme called origin chip that removes the path of the URL. Both browsers potentially improve the situation, but sub-domains used by phishers might still confuse users. We have not been able to find any empirical evaluation of either of these browser-specific schemes.

URL pruning was proposed by Renkema-Padmos *et al.* (2014); its effectiveness evaluation is a contribution of this paper.

3. Experimental design

Our goal is to evaluate the impact of focusing people's attention and of URL pruning on phish detection. More precisely, we wanted to test the following hypotheses:

H1a. Users are significantly more likely to detect phish URLs if their attention is deliberately drawn to the Web browser's address bar which highlights the domain in the URL (i.e. status quo).

H1b. Users are equally likely to identify authentic URLs if their attention is deliberately drawn to the to the Web browser's address bar which highlights the domain in the URL (i.e. status quo).

H2a. Users detect significantly more phish URLs if their attention is drawn to the address bar which displays a pruned URL, as compared to drawing their attention to the address bar displaying the entire URL with the domain highlighted.

H2b. Users are equally likely to identify authentic URLs if their attention is deliberately drawn to the address bar which displays a pruned URL, as compared to drawing their attention to the address bar displaying the entire URL with the domain highlighted.

3.1 Selection of websites and URL manipulation techniques

We selected 16 websites based on the degree of popularity (based on Alexa). We captured screenshots of the login pages. We selected the most widely used browser in Germany, i.e. Firefox[1]. To prepare the screenshots, the latest version of Firefox was installed, without any add-ons other than Java and Shockwave Flash, on a 64 Bit Windows 7 operating system. The two add-ons were essential to prevent corresponding warning messages from appearing.

Half of the screenshots were altered to be phish websites. The GNU Image Manipulation Program GIMP was used to prepare the phish page screenshots. As we used Socisurvey, we were limited to 150 KB per image, this could only be achieved by using JPEG compression on the images. This undeniably impacted the quality of the images. The image size was set to 1,024 × 768, as this was likely to fit into the browser window without too much scrolling[2]. As we wanted to test the impact of focusing users' attention on the address bar, and of URL pruning, we only modified the URL. Thus, phish websites could only reliably be detected by checking the URL (carefully). Note that we decided to use the HTTPS indicator for both phish and non-phish displays, because we wanted to investigate the impact of URL pruning, and not the impact of HTTPS being absent or present.

Next, we considered which URL manipulation techniques to apply. Researchers have proposed different URL manipulation classifications (Garera *et al.*, 2007; Lin *et al.*, 2011), with yet others appearing in public media[3]. We adapted the classification proposed by Lin *et al.* (2011), with each type's anticipated success depending on how well users understand URLs and the thoroughness of their URL checking. The categories are:

- *Obfuscate*: The phish URL is composed of an arbitrary name or IP address. The brand name of the authentic website does not appear: for example, www.slsdz.com and http://123.32.22.123 instead of, for example, www.amazon.com. This type is the most obvious and can probably be detected by a cursory check to see whether the brand name appears.
- *Mislead*: The phish URL embeds the authentic name somewhere to allay suspicions. For example, www.amazon.slz.com and www.new-books-online.com/amazon.com instead of www.amazon.com. This type cannot be detected by a cursory check for the presence of the brand name in the URL but this relies on the user knowing which part of the URL to check for the brand name.
- *Mangle*: The phish URL includes letter substitutions, different letter ordering or misspelling. Examples are paypa1 instead of paypal, googel instead of google or amazzon instead of amazon. This type could deceive users who know which is the relevant part of the URL but do not perform a thorough check and consequently do not detect minor differences.
- *Camouflage*: The domain name of the phish URL contains the brand name together with an extension or a different top-level domain. Examples are amazon-canda.ca, login-amazon.de, online-banking-bank-of-america.com or amazon.key. This type could deceive users who know which is the relevant part of the URL but who possess incomplete knowledge of all legitimate domain name(s) for the authentic website.

The selected phish websites and corresponding applied URL manipulation techniques are provided in Table I.

3.2 Procedure

We conducted a within-subject study (for *H1a* and *H1b*) and a between subject study (for *H2a* and *H2b*), with participants randomly assigned to one of two groups:

(1) *Highlighting*: The group sees the URLs with the domain highlighted.

(2) *Pruning*: The group sees the pruned URL.

The study comprised five phases which are detailed in the following paragraphs[4]:

Brand	URL (abbreviated with "…")
Facebook (Obfuscate)	L: www.facebook.com/login M: https://192.168.111.112/login
Twitter (Mislead)	L: https://twitter.com/ M: https://twitter.webmessenger.com/
Amazon (Mislead)	L: www.amazon.de/ap/signin/… M: www.amazon.de.bestellungabschliessen-buecheronline.de/ap/signin/…
Flickr (Mislead)	L: https://login.yahoo.com/config/login?.src=flickr.signin… M: https://login.xpla.net/config/login?.src=flickr.signin…
Postbank (Camouflage)	L: https://banking.postbank.de/rai/login M: https://banking.postbank-online-banking.de
Hotmail (Camouflage)	L: https://login.live.com/login.srf?wa=wsignin1.0&rpsnv=… M: https://login.hotmailsecure.com/login.srf?wa=wsignin1.0&rpsnv=…
PayPal (Mangle)	L: www.paypal.de/webapps/mpp/privatkunden M: www.paypa1.de/webapps/mpp/privatkunden
Lovefilm (Mangle)	L: www.lovefilm.de/visitor/login_lf.html M: www.lovefi1m.de/visitor/login_lf.html

Table I. Legitimate (L) and manipulated (M) URLs, including type of manipulation

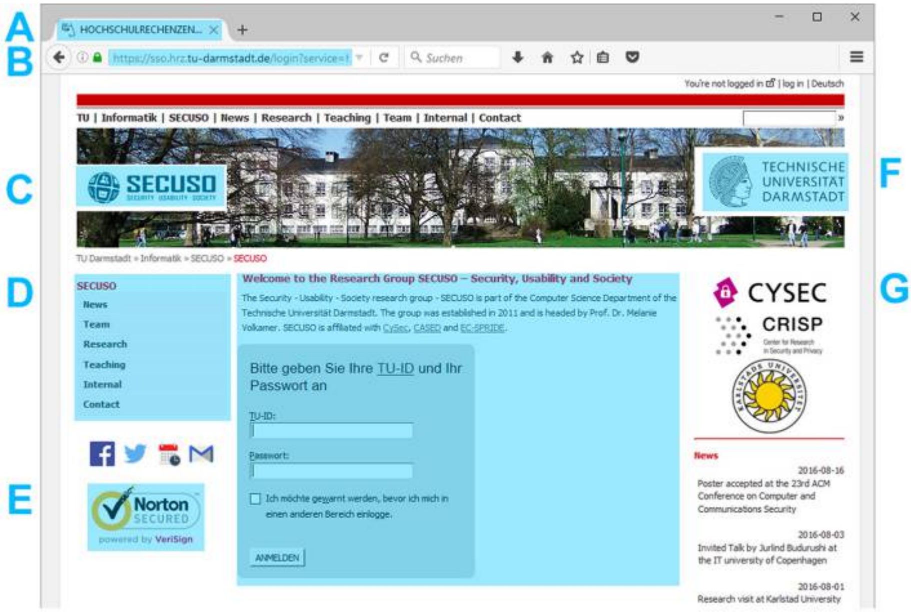
3.2.1 Phase 0: welcome. General information was provided, including the goal of the study (visual security aids provided by Web browsers), number of phases, the estimated duration, the prize and data protection. We did not focus their attention on URLs, domain highlighting or potential phishing attacks. We explained that it was important not to seek assistance. We did not elaborate by mentioning Google, as this could have been counter-productive, as it might have given them the idea of searching.

3.2.2 Phase 1: judging screenshots. Participants were presented with images of 16 websites (8 authentic and 8 phish), each on a different page and in random order. Participants were asked whether they would enter their password assuming they had an account[5]: "No", "Probably not", "Do not know", "Probably" and "Yes". The phase concluded with two questions, each on a separate page. Then, there was an open question asking what they based their decisions on. Afterwards, a picture of a website with different areas labelled (see Figure 2) was displayed, and they were asked to identify which one area they based their decisions on.

3.2.3 Phase 2: judging screenshots after directing participants to focus on the URL. We explained that the images would be shown again, and they should again decide whether to login. Participants were directed to focus on the address bar in making their decision. Screenshots were, again, displayed in a random order. Previous choices were not shown, and it was impossible to go back. This time, there were three exit questions: First, we asked whether participants felt that the instruction to focus on the address bar had helped them to make better decisions ("Not at all", "Not really", "Neutral", "Perhaps" and "Yes"). Afterwards, we asked the open question again regarding what they based their decision on. Furthermore, the picture of different areas of the website was replaced by different areas of the URL (Figure 3).

3.2.4 Phase 3: demographics. We requested information about age and gender.

3.2.5 Phase 4: debrief. During this phase, the real purpose of the study was explained, and participants were referred to a page providing more information about how to protect themselves against phishing attacks. They could, during this phase, delete their



HOCHSCHULRECHENZEN...

https://sso.hrz.tu-darmstadt.de/login?service=I

Suchen

You're not logged in | log in | Deutsch

TU | Informatik | SECUSO | News | Research | Teaching | Team | Internal | Contact

SECUSO SECURITY · USABILITY · SOCIETY

TECHNISCHE UNIVERSITÄT DARMSTADT

TU Darmstadt » Informatik » SECUSO » SECUSO

SECUSO
News
Team
Research
Teaching
Internal
Contact

Welcome to the Research Group SECUSO – Security, Usability and Society

The Security - Usability - Society research group - SECUSO is part of the Computer Science Department of the Technische Universität Darmstadt. The group was established in 2011 and is headed by Prof. Dr. Melanie Volkamer. SECUSO is affiliated with CySec, CASED and EC-SPRIDE.

Bitte geben Sie Ihre TU-ID und Ihr Passwort an

TU-ID:

Passwort:

□ Ich möchte gewarnt werden, bevor ich mich in einen anderen Bereich einlogge.

ANMELDEN

Norton SECURED powered by VeriSign

CYSEC

CRISP Center for Research in Security and Privacy

KARLSTADS UNIVERSITET

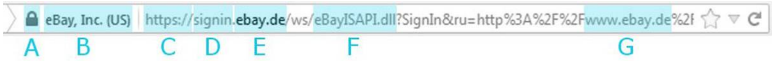
News

2016-08-16
Poster accepted at the 23rd ACM Conference on Computer and Communications Security

2016-08-03
Invited Talk by Jurlind Budurushi at the IT university of Copenhagen

2016-08-01
Research visit at Karlstad University

Figure 2.
Figure similar to the one displayed together with the question: "What did you base your decisions on?"



eBay, Inc. (US) https://signin.ebay.de/ws/eBayISAPI.dll?SignIn&ru=http%3A%2F%2Fwww.ebay.de%2F

Figure 3.
Figure that was displayed together with the question: "On which part of the URL are decisions based?"

own answers. They were invited to participate in the prize draw. Participants could also provide comments about the study and register to receive information about the results.

3.3 Ethics

Our university's ethical requirements with respect to respondent consent and data privacy were met. Participants first read an information page on which they were assured that their data would not be linked to their identity and that the responses would only be used for study purposes. Furthermore, the use of SoSciSurvey ensured that data were stored in Germany and were thus subject to German data protection law. We debriefed participants at the end of the study and provided them with links to websites providing information about phishing.

3.4 Recruitment and incentives

The recruitment of participants was carried out over various online channels to achieve as much diversity in our sample population as possible. In addition to flyers and postings, friends have also been directly requested to participate via e-mail. In the internet, announcements were made on Facebook, in different user groups and on a

Germany-wide forum for mountain bikers (approximately 9,600 registered users), as well as on a "Berlin" Blog (approximately 100 daily visitors). It was also posted on the intranet of a company with approximately 6,000 employees in various locations. Flyers were distributed at our university, in halls of residence and in supermarkets in several cities. The flyers were checked several times throughout the survey period and replaced as necessary, as far as they were publicly accessible. To encourage participation, we invited participants to enter their e-mail address into a prize draw for a €50 Amazon gift voucher upon completion of the study.

4. Results

A total of 411 participants completed the experiment (210 in the highlighting and 201 in the pruning group). Participants were recruited using snowball sampling. The average age was 34.7 years in both groups, with participants' age ranging from 16 to 72 years. Approximately, one-third were female.

4.1 Hypotheses testing

We were interested in whether instructing users to focus on the address bar would improve phish detection and whether pruning of the URL would deliver any further improvement. For the analyses, we only considered those that did not check the URL in Phase 1 to test the hypotheses. This was to ensure that we isolated the impact of the instruction to focus attention on the address bar. Those who already did so in Phase 1, unprompted, were thus removed so that their success/failure did not skew the findings. We refer to these participants as the "uninitiated".

To identify the uninitiated participants, we analysed the open text answers related to the question about what participants based their decisions on in Phase 1. We searched for words indicating that the participants looked at the URL: such as URL, address bar and correct/wrong address. Mention of HTTPS or the lock icon was not counted here. We also checked whether they had identified parts of the screen other than Part B in Figure 2 as the part of the screen they focused on when making their decision. We found that 46 per cent did not check the URL to judge whether a website was authentic or phish (for their demographics, see Table II). These, the uninitiated, are arguably most likely to be at risk.

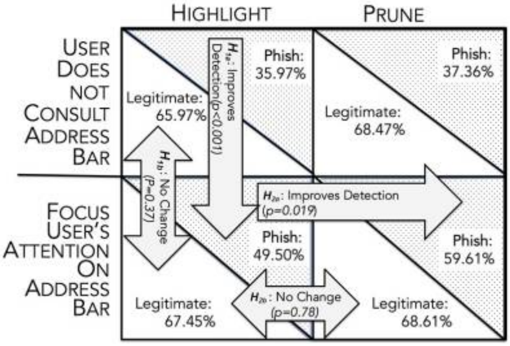
We next calculated the probabilities of a correct decision for uninitiated participants in both phases (i.e. without and with the hint to focus on the address bar) and for both types of web pages (authentic and phish). We then conducted two different analyses. To test the first two hypotheses, we conducted a repeated-measures multivariate analysis of variance (MANOVA) for uninitiated participants in the highlighting group. The within-subject factor (independent variable) was the phase (Phase 1 vs Phase 2, i.e. without and with the hint to consult the address bar) in which the two dependent variables were the probability of correct decision for authentic and for phish pages.

To test the third and fourth hypotheses, we conducted MANOVA for the uninitiated participants for both groups (highlight/prune). The independent variable was the experimental group (highlight vs prune), and dependent variables were the probability of a correct decision for authentic and phish pages in Phase 2 (with the hint to consult the address bar). Figure 4 gives an overview of the descriptive data which correspond to these analyses:

Table II.
Uninitiated age and gender (questions were optional)

Descriptor	Total	Highlighting	Pruning
No. of participants	411	210	201
Uninitiated (did not consult address bar)	189	101	88
Age provided	183	96	87
Average age	39.31	39.88	38.68
Median	37	38	35
Youngest	16	16	16
Oldest	72	72	71
Gender provided	187	100	87
Male	102	58	44
Female	84	42	42
Other	1	0	1

Figure 4.
Average percentages for detecting a phish and identifying a legitimate website based on different conditions, i.e. (not) focusing users' attention on the address bar and (not) pruning the URL and corresponding p-values



H1a. Our analysis confirms that users are significantly more likely to detect phish URLs if their attention is drawn to the Web browser's address bar which highlights the domain in the URL ($F = 25.70$; $p < 0.001$; $\eta^2 = 0.20$).

H1b. Our analysis shows that users are equally likely to identify authentic URLs if their attention is drawn to the Web browser's address bar which highlights the domain ($F = 0.82$; $p = 0.37$; $\eta^2 = 0.01$).

H2a. Our analysis shows that users detect significantly ore phish URLs if their attention is drawn to the address bar displaying a pruned URL, than a highlighted URL with the domain highlighted ($F = 5.56$; $p = 0.019$; $\eta^2 = 0.029$).

H2b. Our analysis shows that users are equally likely to identify authentic URLs if their attention is deliberately drawn to the address bar displaying a pruned URL, as when the URL is displayed with the domain highlighted ($F = 0.80$; $p = 0.78$; $\eta^2 = 0.00$).

4.2 Further results

Do people feel it helped that their attention was focused on the address bar? When asked about this, the average rating was 3.8 (on a scale from 1, "Not At All", to 5, "Absolutely"), and the median was 4 for those participants who did not check the URL in Phase 1. More precisely, 71 per cent stated that they either agreed or agreed absolutely.

Can people identify the relevant parts of a URL? We analysed whether people knew which parts of the URL were relevant, based on the applicable question in Phase 2 (Figure 3). Considering all participants, 36 per cent selected the domain name as the signal part of the URL when it comes to phish detection, and 32 per cent selected HTTPS-related parts, i.e. the lock icon (A) or the HTTPS (C). Considering only the uninitiated participants, 31 per cent selected the domain name and 41 per cent selected HTTPS-related parts.

Status quo for phish detection: in the highlighting group in Phase 1 (which represents the status quo), the phish detection rate was, on average, 53.61 per cent (averaged across both the uninitiated and the initiated participants).

5. Discussion

Our results show that we can improve phish detection by focusing the user's attention on the address bar. The uninitiated participants acknowledged the value of this instruction when asked. It seems wise to deliberately implement approaches to focus user attention on the address bar. We can improve it even further by pruning the URL so that when they check, the checking process is as simple as possible.

Despite the success of our two-pronged approach, we have to acknowledge that we only achieved a 60 per cent detection rate (improved from 36 per cent) for uninitiated participants. One reason might be that 51 per cent of the uninitiated participants based their decision on whether HTTPS was present, or not. This is likely to be a flawed strategy, as it is relatively commonplace for phishers to use HTTPS (Dhamija *et al.*, 2006). Therefore, we can conclude that although URL pruning is promising, the misplaced trust in the presence of HTTPS, in this context, needs to be addressed.

5.1 Approaches to focusing user attention on the URL

There are a number of different ways of focusing user attention on the URL:

- *Awareness campaigns*: Example is distributing posters at public places that show that phish protection requires careful examination of the URL.
- *Using visual means to draw the user's attention to the address bar*: This could be an arrow pointing towards the address bar or enlarging the address bar or highlighting it. We still believe, as argued in Section 2, that a fully active approach will be counter-productive, so what we are proposing is a semi-active approach. These visual alerts should appear occasionally so that their novelty makes them noticeable. Repeating these visual mechanisms from time to time is also important to maintain awareness and to remind people of the need to check.
- *Display a passive message*: A passive message can be displayed next to entry field when users focus on the corresponding entry field (e.g. a password field) for the first time (Maurer *et al.*, 2011).

5.2 Deploying URL pruning

The rules for changing the URL display are simple and can be easily implemented in most URL parsing and display libraries. Some users may use the URL as an aid in navigating the site, editing the URL directly instead of using links provided by the web page. To support this, the implementation of URL pruning would need to allow users to view the entire URL by clicking on the address bar (as mobile browsers do). Note, for deployment, it is important to be aware that URL pruning is tricky when there are addresses such as www.ucl.ac.uk. Pruning this URL to ac.uk would remove the signal, rather than only the noise. As highlighting confronts the same issue, both schemes could make use of advice lists to inform their pruning or highlighting. There are acknowledged challenges with keeping such lists current, but that is not the focus of this paper.

5.3 Addressing misplaced trust

We have to strive to refine and extend our two-pronged approach to improve detection rates. More precisely, it is necessary to explain the meaning of HTTPS being in place and signal nature of the domain. One idea is to add an explanation line below the actual address bar.

5.4 Limitations

Our results describe a best-case scenario. As security was the primary goal for participants, we asked them whether they would log in. One could argue that we should first solve the problem in a best-case scenario before trying to improve the situation in a more challenging context. Furthermore, we tested exclusively with HTTPS, and, as such, we cannot say anything about user behaviour with respect to HTTP either for authentic or phish URLs. Although various types of people participated, the sample is not entirely representative of the German population using the internet, as, for example, only 34 per cent of the participants were female (overall).

Participants could easily have used another browser window to verify the integrity of the URLs. Participants' ratings could well have been influenced by trust or distrust in the corresponding service or by prior knowledge that PayPal, for example, is often subject to frequent attacks.

6. Conclusion and future work

Although phishing has enjoyed media attention for some time now, phishers still catch people with their bait. We reported on the impact of a two-pronged intervention: instituted in an attempt to improve phish detection. The first is to draw a user's attention to the address bar, and the second is to simplify the URL by pruning it. We tested the efficacy of these approaches in a Web browser context in an online study.

Our first finding was that that people could not be relied on to check the URLs: 46 per cent did not do so unprompted. Furthermore, merely focusing their attention on the address bar improved phish detection significantly. Our next finding is that focusing users' attention on the address bar, in combination with URL pruning, improved phish detection even more. Our study revealed that there is a widespread misplaced trust in HTTPS, which is problematical. We proposed some approaches to address this.

It is worth mentioning that it is probably impossible to improve detection to the extent that 100 per cent of phish websites will be detected. What we can do, however, is to keep attempting to improve the situation as much as possible. Pruning undeniably improves detection rates, and we plan to continue to find ways of improving detection further.

Many people, globally, are increasingly accessing the Web from their mobile devices. Pruning has the potential to deliver even greater benefit in this context, with limited screen space and the potential for phishers to exploit this fact by adding enough noise so that they cannot even see the actual domain. For example, they would see: www.mail.

login.amazon.de… while the actual URL is: www.mail.login.amazon.de.phishing-smartphone-users.com

As future work, we plan to carry out lab investigations and further field studies. We also plan to study the efficacy of URL pruning in other contexts, such as e-mails and mobile platforms.

This study has confirmed the fact that no one mechanism or tool will solve the phishing problem. It is more likely that a suite of tools has to be deployed to help the population at large to protect themselves from phish attacks. Blacklisting can reduce the success rates of identified phish websites. Banks can act to protect themselves by buying mangled domains, as eBay has done for paypa1[6]. Web surfers still have to protect themselves; they cannot assume that someone else is doing this on their behalf. We should implement some kind of mechanism to maximise the likelihood that they will check the URL and that they are not misled by HTTPS, and we should also prune the URL to ensure that that the checking task is as simple as possible.

Notes

1. www.browser-statistik.de/statistiken/ (accessed 27 July 2015).
2. 10 per cent of German users use this setting according to Kuhn (2013) and StatCounter (2013).
3. www.phishtank.com, www.securityskeptic.com/2014/08/is-it-a-Phish-5-visual-deceptions-in-paypal-Phishing-url-composition-.html (accessed 27 July 2015).
4. Questions and screenshots were translated from German for this paper.
5. For bank web pages, this question asked whether they would enter their account number and PIN.
6. https://who.is/whois/paypa1 (accessed 16 June 2015).

References

Akhawe, D. and Felt, A.P. (2013), "Alice in warningland: a large-scale field study of browser security warning effectiveness", *Proceedings of the 22nd USENIX Security Symposium, Washington, DC*, 14-16 August.

Alnajim, A. and Munro, M. (2009), "An anti-phishing approach that uses training intervention for phishing websites detection", *Sixth International Conference on Information Technology: New Generations ITNG (2009)*, Las Vegas, CA, 27-29 April, pp. 405-410.

APWG Internet Policy Committee (2013), "Global phishing survey: trends and domain name use in 2h2013", available at: http://docs.apwg.org/reports/APWG_GlobalPhishingSurvey_2H2013.pdf

APWG Internet Policy Committee (2014), "Phishing activoty trends report. 2nd quarter", available at: https://apwg.org/resources/apwg-reports/

Bar-Yossef, Z., Keidar, I. and Schonfeld, U. (2009), "Do not crawl in the DUST: different URLs with similar text", *ACM*, Vol. 3 No. 1, http://doi.acm.org/10.1145/1462148.1462151.

Canova, G., Volkamer, M., Bergmann, C. and Borza, R. (2014), "Nophish: an antiphishing education app", *Security and Trust Management*, Springer, Wroclaw, pp. 188-192.

Dhamija, R., Tygar, J.D. and Hearst, M. (2006), "Why phishing works", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', ACM, Quebec*, pp. 581-590.

Egelman, S., Cranor, L.F. and Hong, J. (2008), "You've been warned: an empirical study of the effectiveness of web browser phishing warnings", *CHI 2008 Conference on Human Factors in Computing Systems', Florence*, pp. 1065-1074.

Garera, S., Provos, N., Chew, M. and Rubin, A.D. (2007), "A framework for detection and measurement of phishing attacks", 2007 ACM Workshop on Recurring Malcode', Alexandria, pp. 1-8.

Jansson, K. and von Solms, R. (2011), "Simulating malicious emails to educate end users on-demand", 2011 3rd Symposium on Web Society (SWS), Port Elizabeth, 26-28 October, pp. 74-80.

Kirlappos, I. and Sasse, M.A. (2012), "Security education against phishing: a modest proposal for a major rethink", *IEEE Security and Privacy Magazine*, Vol. 10 No. 2, pp. 24-32.

Kuhn, B.-L. (2013), "Top-Statistik: aktuelle Bildschirmaufloesungen im Juli 2013 (Screenshot Resolution in July 2013)", available at: www.proteus-solutions.de/~Unternehmen/News-PermaLink:tM.F06!sM.PV00!Article.955799.asp

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L.F., Hong, J. and Nunge, E. (2007), "Protecting people from phishing: the design and evaluation of an embedded training email system", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', San Jose, CA, CHI '07, ACM, New York, NY*, pp. 905-914, http://doi.acm.org/10.1145/1240624.1240760.

Kumaraguru, P., Cranor, L.F. and Mather, L. (2009), "Anti-phishing landing page: turning a 404 into a teachable moment for end users", *Sixth Conference on Email and Anti-Spam*, Mountain View, CA.

Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., Cranor, L.-F. and Hong, J. (2007), "Getting users to pay attention to anti-phishing education: evaluation of retention and transfer", *Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit', eCrime '07, Pittsburgh, PA*, ACM, *New York, NY*, pp. 70-81, http://doi.acm.org/10.1145/1299015.1299022

Li, L. and Helenius, M. (2007), "Usability evaluation of anti-phishing toolbars", *Journal in Computer Virology*, Vol. 3 No. 2, pp. 163-184, http://dx.doi.org/10.1007/s11416-007-0050-4.

Lin, E., Greenberg, S., Trotter, E., Ma, D. and Aycock, J. (2011), "Does domain highlighting help people identify phishing sites?", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', Vancouver, BC, CHI '11, ACM, New York, NY*, pp. 2075-2084.

Marchal, S., François, J. and Engel, T. (2012), "Proactive discovery of phishing re-lated domain names", *Research in Attacks, Intrusions, and Defenses'*, Springer, pp. 190-209.

Maurer, M.-E., De Luca, A. and Kempe, S. (2011), "Using data type based security alert dialogs to raise online security awareness", *Proceedings of the Seventh Symposium on Usable Privacy and Security', Pittsburgh, PA, 20-22* July, *ACM, New York, NY*, p. 2.

Maurer, M.-E. and Herzner, D. (2012), "Using visual website similarity for phishing detection and reporting", *Proceedings of the 2012 Annual Conference Extended Abstracts on Human Factors in Computing Systems', CHIEA '12, Austin, TX, 5-10* May, *ACM, New York, NY.*

Parija, D., Patra, T., Kumar, A., Swain, B., Satyanarayana, S., Sreenivas, A., Chadha, V., Moonan, P. and Oeltmann, J. (2014), "Impact of awareness drives and community-based active tuberculosis case finding in Odisha, India", *The International Journal of Tuberculosis and Lung Disease*, Vol. 18 No. 9, pp. 1105-1107.

Prakash, P., Kumar, M., Kompella, R.R. and Gupta, M. (2010), "Phishnet: predictive blacklisting to detect phishing attacks", *INFOCOM, 2010 Proceedings IEEE', IEEE Press, Piscataway, NJ*, pp. 346-350.

Renkema-Padmos, A., Volkamer, M. and Renaud, K. (2014), "Building castles in quick sand: blueprint for a crowdsourced study", CHI'14 Extended Abstracts on Human Factors in Computing Systems', ACM, pp. 643-652.

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L.F., Hong, J. and Nunge, E. (2007), "Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish", *Proceedings of the 3rd Symposium on Usable Privacy and Security', SOUPS '07, ACM, New York, NY*, pp. 88-99, http://doi.acm.org/10.1145/1280680.1280692.

StatCounter (2013), "Top 14 screen resolutions in Germany from July to Dec 2013", available at: http://gs.statcounter.com

Trend Micro Incorporated (2016), "Cybercriminals reinvent methods of malicious attacks", available at: www.crime-research.org/analytics/3451/

Wu, M., Miller, R.C. and Garfinkel, S.L. (2006), "Do security toolbars actually prevent phishing attacks?", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', CHI '06, Paris, 27 April 2 May, ACM, New York, NY*, pp. 601-610, available at: http://doi.acm.org/10.1145/1124772.1124863.

Zhang, Y., Egelman, S., Cranor, L.F. and Hong, J. (2007), "Phinding phish: evaluating anti-phishing tools", 14th Annual Network and Distributed System Security Symposiums, San Diego, CA.

About the authors

Dr Melanie Volkamer has been appointed Full Professor for Usable Privacy and Security at Karlstad University. She is also a Professor (Kooperationsprofessur) at the Department of Computer Science of Technische Universität Darmstadt (Germany) since August 2016. Before she was an Assistant Professor at TU Darmstadt, Professor Volkamer has been heading the research group "SECUSO – Security, Usability and Society" since 2011.

Karen Renaud is a Scottish Computing Scientist working on all aspects of Human-Centred Security and Privacy in the School of Computing Science at the University of Glasgow and is one of five UK Cyber Security Fulbright Awardees for 2016/2017. She is particularly interested in supporting innovation in security and privacy. Karen Renaud is the corresponding author and can be contacted at: karen.renaud@glasgow.ac.uk

Paul Gerber works as a Doctoral Researcher since 2012 after he had received his Diploma in Psychology at the Technische Universität Darmstadt. He works for the work and engineering psychology, as well as the SECUSO research group. In his work for SECUSO, he seeks for ways to improve user understanding for security and privacy issues connected with mobile applications. He also works with security experts to foster their abilities to accurately decide whether mobile applications fit to the security policies in their business. For the FAI, he is responsible for the eye-tracking laboratory and is doing his research in human-machine interaction. Currently, he works for the MoPPa project which deals with the so-called Privacy Paradox from an interdisciplinary point of view.