Computers in Human Behavior 69 (2017) 421–436

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

Fishing for phishers. Improving Internet users' sensitivity to visual deception cues to prevent electronic fraud

María M. Moreno-Fernández [a], Fernando Blanco [a], Pablo Garaizar [b], Helena Matute [a, *]

[a] Faculty of Psychology and Education, University of Deusto, Spain
[b] Faculty of Engineering, University of Deusto, Spain

ARTICLE INFO

Article history:
Received 4 March 2016
Received in revised form
12 December 2016
Accepted 18 December 2016
Available online 19 December 2016

Keywords:
Phishing
Internet security
Easy-to-hard effect
Human-computer interaction
Discrimination learning
Visual discrimination

ABSTRACT

Phishing is a form of electronic fraud in which attackers attempt to steal sensitive information by posing as a legitimate entity. To maintain the attack unnoticed, phishers typically use fake sites that accurately mimic real ones. However, there are usually subtle visual discrepancies between these spoof sites and their legitimate counterparts that may help Internet users to identify their deceptive nature. Among all the potential visual cues, we choose to focus on typography, because it is often hard for phishers to use exactly the same font as in the original website. Thus, Experiment 1 assessed the effectiveness of visual discrimination training to help people detect typographical discrepancies between fake and legitimate websites. Results showed higher sensitivity to differences when undergraduate students were previously trained with easier versions of the discrimination task (i.e., involving more noticeable differences in typography) than when they were trained with the difficult target discrimination from the start (*easy-to-hard effect*). These results were replicated with a broader and more representative sample of anonymous Internet users in Experiment 2. Implications for the design of strategies to prevent electronic fraud are discussed.

1. Introduction

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, 2016, p. 2). Although phishers can use different strategies to reach their goals, in a typical scenario they pose as a reliable entity (e.g., trustworthy companies, acquaintances or even public bodies) and use e-mails as lures for driving Internet users to fraudulent websites. Deceitful websites are specifically designed to resemble the legitimate version, causing users to remain unaware of the fraud, and increasing their probability of being tricked.

Although not new, phishing has become an increasing threat for cyber-security. According to the Anti-Phishing Working Group (the leading international consortium of business, regulators, and agencies that monitor phishing attacks worldwide and attempt to coordinate responses to such attacks), the number of unique phishing websites detected during the 1st quarter of 2016 increased by 250% compared to the last quarter of 2015 (Anti-Phishing Working Group, 2016). Moreover, during the first semester of 2016 alone, a total of 102,573 submissions of suspected phishing attacks were verified as real (valid phishes) by the PhishTank community (OpenDNS, 2016). Given the scope of this threat and its consequences, a growing body of research has begun to explore how to prevent Internet users from being phished.

An extensive number of anti-phishing strategies have been developed, covering all stages of the phishing attack process, and using complementary approaches that range from technical to legal interventions (for a review, see Jakobsson & Myers, 2007; Mohammad, Thabtah, & McCluskey, 2015; Purkait, 2012). For example, a phishing attack can be detected at a very early stage (before it actually starts) by monitoring the registration of potential spoof domains, or by controlling unusual patterns of access to the legitimate website. The rationale of the latter approach rests on the necessity of phishers to repeatedly access the legitimate website to download and copy relevant contents to create the illegitimate version. The specific analysis of IP addresses associated with these unusual download activities may help to detect and react against an

* Corresponding author. Faculty of Psychology and Education, University of Deusto, Avda. Universidades 24, 48007 Bilbao, Spain.
E-mail address: matute@deusto.es (H. Matute).

http://dx.doi.org/10.1016/j.chb.2016.12.044

imminent phishing attack (Emigh, 2007). However, although predicting and blocking phishing activities at this early stage should be the optimal solution, it is not always possible.

When phishers succeed in launching the attack, the ideal strategy should be to prevent users from being exposed to the subsequent threat. With this aim in mind, a number of automatic detection strategies have been developed, ranging from blacklists of phishing domains (such as the Anti-Phishing Working Group blacklist, or the Google Safebrowsing service), to heuristic-based methods that can recognize phishing websites by analyzing their visual features (e.g., Liu, Deng, Huang, & Fu, 2006; Liu, Guanglin, Liu, Zhang, & Xiaotie, 2005; Maurer & Herzner, 2012; Medvet, Kirda, & Kruegel, 2008; Zhang, Liu, Chow, & Liu, 2011). But once again, although these technical approaches can be regarded as a good first line of defense against phishing, to date there is no strategy can completely prevent phishing attacks. Therefore, training users to detect fake websites and to protect themselves is currently a key component in cyber-security.

1.1. Human behavior and user-oriented approaches

The critical role of human behavior in the success of phishing attacks has encouraged the development of strategies aimed at promoting safer decisions across all the stages of the phishing attack flow in which human performance is involved. Some of these approaches have focused on teaching users to identify deception cues in phishing attack vectors such as e-mails whilst providing, at the same time, security tips (e.g., Anti-Phishing Phyllis™, Wombat Security Technologies, 2016; or PhishGuru, Kumaraguru et al., 2007). However, in addition to emails, phishers may currently use a wide range of strategies to lure users (e.g., messages posted on social media, phone calls, or SMSs). Therefore, designing preventive strategies to help users at successive stages of the attack (that is, once the illegitimate site is visited) becomes essential.

Client-side strategies such as security indicators (toolbars, warnings, or browser indicators) have been developed to signal trustworthiness or to alert users about potentially dangerous sites. Recent research has shown that warnings can effectively reduce people's likelihood of disclosing sensitive information on legitimate websites, although this reduction depends on the warning word used and on the identity information targeted (see Carpenter, Zhu, & Kolimi, 2014). Unfortunately, research on phishing also highlights the limited effectiveness of security indicators because people do not use them as expected. For example, Dhamija, Tygar, and Hearst (2006) carried out a laboratory study to assess the ability of Internet users to detect fraudulent websites, as well as the strategies that they used for judging website legitimacy. Participants were asked to categorize websites as legitimate or not, rating their confidence in their responses, and explaining the reasons underlying their choice. Results showed that even in a non-natural environment where participants were warned and primed about the possibility of being fooled, they could not distinguish accurately between spoof and legitimate websites (40% of participants' choices were incorrect). But what is probably more surprising is that browsers' warning cues such as address bars, status bars, or security indicators (e.g., lock icons in the address bar), went unnoticed by many participants.

Alsharnouby, Alaca, and Chiasson (2015) replicated and extended previous results in a more recent study using eye tracking. The authors used a procedure similar to the one used by Dhamija et al. (2006) but, in addition to behavioral measures and participants' self-reports, they included eye-tracking measures to obtain additional information about the user's attention to security cues. Their results confirmed that participants were not able to reliably identify fraudulent sites, spending most of the time examining the content of the website and paying little attention to security indicators (for similar results, see also Aburrous, Hossain, Dahal, & Thabtah, 2010; Lin, Greenberg, Trotter, Ma, & Aycock, 2011; Whalen & Inkpen, 2005; Wu, Miller, & Garfinkel, 2006). These studies reveal the essential role of human behavior in phishing success, and they highlight the relevance of considering human vulnerabilities when designing preventive strategies.

One main aspect of this vulnerability is the users' knowledge about security and security indicators. Users may not have enough information about these technical resources. For example, Wogalter and Mayhorn (2008) asked a group of participants to rate the extent to which they would trust the information of a website based on trustworthiness signals (i.e., domain suffixes, organization domain names, and quality seals that actually can be used as indicators of website reliability). The authors found that the reported trust on the website contents was related to these three indicators, but, surprisingly, participants showed limited abilities to discriminate between real and fictitious quality seals and organizational domain names. The lack of human competence at this level has raised the interest in educational approaches.

Educational strategies are primarily concerned with teaching the general concepts of cyber-security and phishing by using exercises to reinforce concepts, or by employing specific guided training protocols. However, whilst recent research has pointed out the value of these educational interventions (Kumaraguru et al., 2009, 2007; Sheng et al., 2007), there are other factors that may hinder the use of security indicators even when users do have enough knowledge about them. One of these factors is directly related to users' motivations when using the Internet and the awareness of the possibility of being tricked.

When using the Internet, users are mainly dedicated to their primary goals, that is, browsing web pages, trying to find a product on an e-commerce site, or just replying to their e-mails. Security is rarely their main goal, and consequently it is usually set aside. This "*unmotivated user property*" (Whitten & Tygar, 1999), together with other limitations imposed by human cognitive capacities that might affect decision-making (see Jones, Towse, & Race, 2015 for a review), pose a great challenge for web security as they may restrict the use of security tools.

Phishing attacks commonly profit from human confidence and the cognitive limitations of Internet users (see Dhamija & Tygar, 2005). Thus, scammers usually promote trust beliefs and judgments about legitimacy by simply creating websites that look as similar as possible to the originals, a strategy that becomes effective because of peoples' tendency to overlook security warnings (as discussed above). In this situation, it is important to develop additional strategies that take into account the flaws in human cognition, and their potential interaction with the effectiveness of anti-phishing measures.

A potential option is to increase alertness by improving users' sensitivity to visual deception cues whenever subtle differences exist between an original website and a fake site. If this were possible, websites requiring higher security measures, such as banks or health companies, could train their users to increase their ability to discriminate the original website from potential fakes. Although there are other levels of inconsistency that users might be trained to detect besides perceptual discrepancies (for example, on a procedural level users may be trained to detect credential login inconsistencies); this paper will explore the former approach to help Internet users protect their security by taking advantage of well-known research principles of human visual discrimination learning.