You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings

Serge Egelman
Carnegie Mellon University
egelman@cs.cmu.edu

Lorrie Faith Cranor
Carnegie Mellon University
lorrie@cs.cmu.edu

Jason Hong
Carnegie Mellon University
jasonh@cs.cmu.edu

ABSTRACT

Many popular web browsers now include active phishing warnings since research has shown that passive warnings are often ignored. In this laboratory study we examine the effectiveness of these warnings and examine if, how, and why they fail users. We simulated a spear phishing attack to expose users to browser warnings. We found that 97% of our sixty participants fell for at least one of the phishing messages that we sent them. However, we also found that when presented with the active warnings, 79% of participants heeded them, which was not the case for the passive warning that we tested—where only one participant heeded the warnings. Using a model from the warning sciences we analyzed how users perceive warning messages and offer suggestions for creating more effective phishing warnings.

Author Keywords

Phishing, warning messages, mental models, usable privacy and security

ACM Classification Keywords

H.1.2 User/Machine Systems, H.5.2 User Interfaces, D.4.6 Security and Protection

INTRODUCTION

Online security indicators have historically failed users because users do not understand or believe them. The prevalence of phishing, a scam to collect personal information by mimicking trusted websites, has prompted the design of many new online security indicators. Because phishing is a semantic attack that relies on confusing people, it is difficult to automatically detect these attacks with complete accuracy. Thus, anti-phishing tools use warnings to alert users to potential phishing sites, rather than outright blocking them.

The question remains, do anti-phishing warnings actually help users? Up until recently these tools have relied on passive indicators to alert users. A passive indicator indicates a potential danger by changing colors, providing textual information, or by other means without interrupting the user's task. However, research has shown that passive indicators are failing users because users often fail to notice them or do not trust them [23].

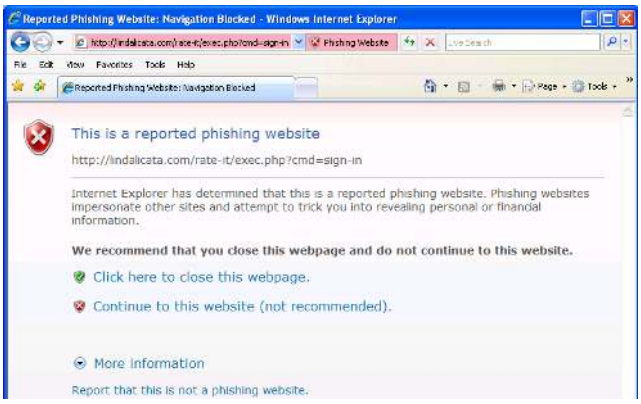

Figure 1. The active Internet Explorer 7.0 phishing warning.

Figure 2. The passive Internet Explorer 7.0 phishing warning.

The newest web browsers now include active warnings, which force users to notice the warnings by interrupting them. Microsoft's Internet Explorer 7 includes both active and passive phishing warnings (Figures 1 and 2, respectively). When IE7 encounters a confirmed phishing website, the browser will display an active warning message giving the user the option of closing the window (recommended) or displaying the website (not recommended). This warning is a full screen error, which turns the URL bar red if the user chooses to display the website (Figure 1). The passive indicator, a popup

CHI 2008, April 5 - 10, 2008, Florence, Italy.

Figure 3. The active Firefox 2.0 phishing warning.

dialog box, is displayed to the user when the browser believes a website is suspicious (Figure 2), but that website has not been verified as being a phishing website (i.e. it does not appear on a blacklist). We consider this warning to be more passive because it does not give the user any choices, nor does it make any recommendations.

Firefox 2.0 also includes an active phishing warning, which was part of the Google Toolbar extension for previous versions of Firefox. When a user encounters a confirmed phishing website, a non-interactive dimmed version of the website is displayed with an overlayed dialog box. The user is given a choice between continuing to the site or leaving. The user may also click the red 'X' in the corner of the warning, which has the same effect as continuing to the website (Figure 3).

In this study we compared the effectiveness of active and passive phishing warnings by analyzing them using a warning analysis methodology used by researchers in the warning sciences field, called the Communication-Human Information Processing Model (C-HIP) model [21].

This paper makes three contributions. First, it presents the results of a study evaluating the effectiveness of active security indicators in current web browsers. Second, it presents an analysis of the results using a model from the warning sciences. Third, it presents recommendations for improving these security indicators such that fewer users fall victim to online fraud.

We first frame our study within the context of previous phishing and warning research, and then describe the methodology behind our study. We then discuss the results of our user study and how effective we determined each warning message to be. Finally, we make recommendations based on these results for designing more effective security indicators.

BACKGROUND

In this section we describe previous work related to users' susceptibility to phishing, warning indicators used in web browsers, and user perceptions of warning messages.

Phishing Susceptibility

Despite growing efforts to educate users and create better detection tools, users are still very susceptible to phishing attacks. Unfortunately, due to the nature of the attacks, it is very difficult to estimate the number of people who actually fall victim. A 2006 report by Gartner estimated the costs at $1,244 per victim, an increase over the $257 they cited in a 2004 report [11]. In 2007 Moore and Clayton estimated the number of phishing victims by examining web server logs. They estimated that 311,449 people fall for phishing scams annually, costing around 350 million dollars [15]. Another study in 2007 by Florencio and Herley estimated that roughly 0.4% of the population falls for phishing attacks annually [9].

Phishing works because users are willing to trust websites that appear to be designed well. In a 2001 study on website credibility, Fogg et al. found that the "look and feel" of a website is often most important for gaining a user's trust [10]. A 2006 phishing study by Dhamija et al. found that 90% of the participants were fooled by phishing websites. The researchers concluded that current security indicators (i.e. the lock icon, status bar, and address bar) are ineffective because 23% of the participants failed to notice them or because they did not understand what they meant [7]. In a similar study, Downs et al. showed participants eight emails, three of which were phishing. They found that the number of participants who expressed suspicion varied for each email; 47% expressed suspicion over a phishing message from Amazon, whereas 74% expressed suspicion over a phishing message from Citibank. Those who had interacted with certain companies in the past were significantly more likely to fall for phishing messages claiming to be from these companies. Participants were also likely to ignore or misunderstand web browser security cues [8].

Phishing Indicators

New research has focused on creating new anti-phishing indicators because existing security indicators have failed. The Passpet system, created by Yee et al. in 2006, uses indicators so that users know they are at a previously-trusted website. Users can store an animal icon within the web browser for each trusted site with which they interact. The system will only send a password when the user recognizes that the animal icons match. Preliminary user testing suggests that this system is easy for users to use [25]. Other proposals have also been put forth to modify browser chrome to help users detect phishing websites. In one system, "synchronized random dynamic boundaries," by Ye and Smith, the browser chrome is modified to blink at a random rate. If the blink rate matches a trusted window's blink rate, the user knows that the window in question has not been spoofed [24]. A similar solution using a trusted window was also proposed by Dhamija and Tygar in 2005. In their system the chrome of the browser window contains a colored pattern that must be matched with the trusted window. The user knows to recognize the trusted window because it contains a personal image that the user selected during the initial configuration [6]. Since all of these proposals require the use of complicated third-party tools, it's unclear how many users will actually benefit from them. These proposals have only undergone minimal user testing in unrealistic environments. User testing should be performed under real world conditions before any new security indicator is recommended.

The SiteKey system was introduced in 2005 to simplify authentication by not forcing the user to install additional software. SiteKey uses a system of visual authentication images that are selected by the user at the time of enrollment. When the user enters his or her username, the image is displayed. If the user recognizes the image as the original shared secret, it is safe to enter the password [2]. However, a recent study found that 92% of participants still logged in to the website using their own credentials when the correct image was not present [19]. However, this sample may have been drawn from a biased population since others refused to participate, citing privacy and security concerns.

Some argue that the use of *extended validation* (EV) certificates may help users detect phishing websites. An EV certificate differs from a standard SSL certificate because the website owner must undergo background checks. A regular certificate only tells a user that the certificate was granted by a particular issuing authority, whereas an EV certificate also says that it belongs to a legally recognized company [4]. The newest version of Microsoft's Internet Explorer supports EV certificates, coloring the URL bar green and displaying the name of the company. However, a recent study found that EV certificates did not make users less likely to fall for phishing attacks. The study also found that after reading a help file, users were less suspicious of fraudulent websites that did not yield warning indicators [13].

Many web browser extensions for phishing detection currently exist. Unfortunately, a recent study on anti-phishing toolbar accuracy found that these tools fail to identify a substantial proportion of phishing websites [26]. A 2006 study by Wu et al. found that the usability of these tools is also lacking because many of them use passive indicators. Many users fail to notice the indicators, while others often do not trust them because they think the sites look trustworthy [23].

A MODEL FOR WARNINGS

In this paper we will analyze our user study results using a model from the warnings sciences. Computer scientists can benefit from studies in this field. Many studies have examined "hazard matching" and "arousal strength." Hazard matching is defined as accurately using warning messages to convey risks—if a warning does not adequately convey risk, the user may not take heed of the warning. Arousal strength is defined as the perceived urgency of the warning [12].

To date, few studies have been conducted to evaluate the arousal strength of software warnings. In one study of warning messages used in Microsoft Windows, researchers found that using different combinations of icons and text greatly affected participants' risk perceptions. Participants were shown a series of dialog boxes with differing text and icons, and were instructed to estimate the severity of the warnings using a 10-point Likert scale. The choice of icons and words greatly affected how each participant ranked the severity. The researchers also examined the extent to which individuals will continue to pay attention to a warning after seeing it multiple times ("habituation"). They found that users dismissed the warnings without reading them after they had seen them multiple times. This behavior continued even when using a similar but different warning in a different situation. The only way of recapturing the user's attention was to increase the arousal strength of the warning [1].

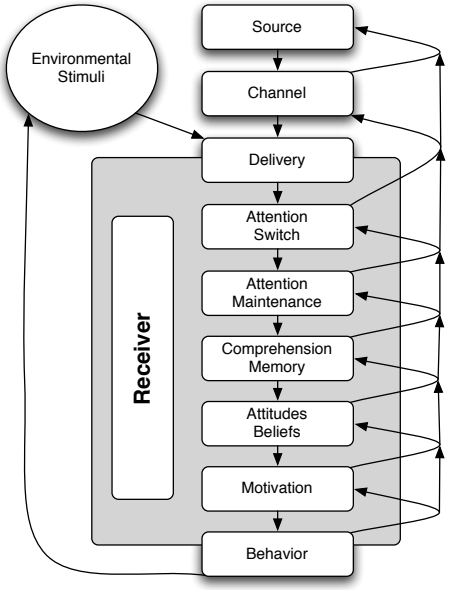

Figure 4. Diagram of the different phases of the C-HIP model [21].

Wogalter proposed the Communication-Human Information Processing Model (C-HIP) for structuring warning research, as shown in Figure 4. He suggests that C-HIP be used to identify reasons that a particular warning is ineffective [21]. The C-HIP model begins with a source delivering a warning through a channel to a receiver, who receives it along with other environmental stimuli that may distract from the message. The receiver goes through five information processing steps, which ultimately determine whether the warning results in any change in behavior.

We can ask the following questions to examine the different steps in Wogalter's model [5]:

1. *Attention Switch and Maintenance* — Do users notice the indicators?
2. *Comprehension/Memory* — Do users know what the indicators mean?
3. *Comprehension/Memory* — Do users know what they are supposed to do when they see the indicators?
4. *Attitudes/Beliefs* — Do they believe the indicators?
5. *Motivation* — Are they motivated to take the recommended actions?
6. *Behavior* — Will they actually perform those actions?
7. *Environmental Stimuli* — How do the indicators interact with other indicators and other stimuli?

Observing users as they complete a task while thinking aloud provides insights into most of the above questions. Alternatively, users can complete tasks and then fill out post-task questionnaires or participate in interviews, although these require users to remember why they did something and report it afterwards, and users sometimes say what they think

the researcher wants to hear. In our study of active phishing indicators we performed a think-alound experiment followed by a post-task questionnaire. We then used the C-HIP model to analyze our data.

METHODOLOGY

In this study participants made online purchases and then were told to check their email, whereupon they encountered phishing messages we had sent them. We observed participants visit the URLs in these phishing messages, at which point the participants were exposed to the web browser warning messages. We took note of whether participants read these warnings and how they chose to proceed. Finally, participants were given an exit survey.

The primary purpose of this study was to examine the effectiveness of phishing warnings found in current web browsers. These warnings serve as the last line of defense against a user divulging his or her sensitive information to a con artist. In other words, prior to these warnings being displayed, it is likely that users believe they are visiting legitimate websites. Thus, we needed users to fall for the phishing messages we sent them during our study so that they would be in a similar state of mind when they encountered the warnings. At the same time, we needed our attack to be plausible. Thus, we simulated a spear phishing attack. Spear phishing "involves personalized emails or emails sent to a specifically targeted group, such as employees of a particular organization" [8]. For instance, a phisher might send a message to email addresses at *aol.com* announcing account changes impacting AOL users. Since all the recipients are AOL users, this scam may have increased credibility because users believe it to be relevant to them. In our study, if participants did not believe our phishing messages to be credible, they would be less likely to follow the links and thus would not see the browser warning messages.

We framed our study as an "online shopping study"—items were purchased online, and then we sent the participants phishing messages claiming to be from those shopping websites. Participants were told that we were examining how they interact with shopping websites and that they needed to think aloud during their purchases. After the first purchase was made, participants checked their email to confirm that the order was going to be shipped, thereby encountering the first phishing message. Once the participants were confident that the first purchase had been completed, instructions were provided for the second purchase. This purchase was then made using a different website, and a different phishing message was sent. Participants in the experimental conditions were given an exit survey before leaving. In this section we will provide the details of our recruitment process and the study design.

Recruitment

This study was designed as a between-subjects study, with four different conditions using the Internet Explorer 7.0 and Firefox 2.0 web browsers: participants were shown either the Firefox warning (Figure 3), the active IE warning (Figure 1), the passive IE warning (Figure 2), or no warning at all. As of June 2007, users of Internet Explorer and Firefox comprised 58.5% and 34.0% of all Internet users, respectively [18]. Additionally, both browsers have automatic update features. Thus, it is only a matter of time before most users will be using the newest versions of these browsers which contain these phishing warnings. We began recruiting participants in May of 2007.

We did not tell participants that we were studying online security because we wanted to simulate a natural environment by not priming them to security concerns. We recruited participants from all over Pittsburgh in order to make our results generalizable. We attached flyers to telephone posts, bus stops, and community bulletin boards. We also posted online to Craigslist and a CMU website for recruiting study participants. We constructed a screening survey to screen out technically savvy individuals, users of certain web browsers, participants in previous phishing studies, and users of certain email providers. We also used this survey to glean some basic demographic information from participants, such as age, gender, occupation, prior online shopping experience, etc.

Participants who contacted us after seeing a recruitment flyer were directed to our online screening survey. Since we were examining the newest versions of Firefox (2.0) and IE (7.0) to include the active warnings, we made sure that all participants in the experimental conditions already used one of these browser versions. Thus the screening survey included a question about current browser version (with graphics depicting how to determine the version) to screen out users of other web browsers.

Since our lab has conducted previous studies on phishing, we were concerned about the potential for priming of prior participants. Thus we disqualified anyone who had previously participated in a phishing-related study. We were also concerned that savvy users would not believe the emails, and thus not be exposed to the warnings. We asked four questions to gauge each participant's experience:

- Have you ever designed a website?
- Have you ever registered a domain name?
- Have you ever used SSH?
- Have you ever configured a firewall?

In our pilot we discovered that participants who answered yes to all four questions were just as likely to believe the phishing emails as all other participants. Thus, we decided not to disqualify participants based on these questions.

We tried to make our scenarios as realistic as possible by requiring participants to use their own email accounts and financial information for the purchases. The screening survey explicitly asked whether or not they could check their email using a web browser on a foreign computer. We also asked them to enter their email addresses so that we could contact them as well as to determine which email provider they were using. We initially found that some of the larger free email providers were detecting our phishing messages and filtering them out. We minimized this problem by implementing DKIM and SPF on our outgoing mail server to help recipient mail servers verify the message sender.[1,2]

[1] http://www.dkim.org/

[2] http://www.openspf.org/

Of the 282 individuals who completed our screening survey, only 70 qualified and showed up. Despite our efforts to screen out individuals who used email providers that were likely to filter out our messages, we still found that we could not collect data from ten participants because they did not receive either of our phishing messages. These ten participants were not included in our results.

Based on the browser versions that they indicated in the screening survey, participants were placed in one of the four conditions. The average age of participants was 28 ($\sigma = 10.58$), and there was no significant difference between the groups in terms of age or gender. The Firefox condition consisted of 20 users of Firefox 2.0, while the other two experimental conditions consisted of users of Internet Explorer 7 (20 participants in the active IE condition and 10 participants in the passive IE condition). The ten participants in the control group all used an older version of one of the two browsers. The control group was used to determine whether or not participants were willing to enter information into our phishing websites in the absence of any warning messages. This told us whether the warning was affecting phishing susceptibility or if it could be attributed to some other factor. The group sizes were chosen based on a power analysis performed prior to recruitment.

We were initially concerned that the self-selected nature of the groups (based on web browser preference) may have biased our study. However, we found no statistical differences between the average number of hours participants in each group claimed to use the Internet, nor with regard to the average number of email messages participants claimed to receive. In each of the active warning groups, exactly seven participants answered "no" to all of the questions used to gauge technical prowess. Thus, there were equal numbers of novices in each group.

Scenarios

We decided to spoof Amazon and eBay since they were the most commonly phished non-bank websites [17]. Thus, regardless of familiarity with the real websites, it is likely that participants have previously encountered phishing messages claiming to be from these websites. Our spoofed websites consisted of login forms for usernames and passwords. To make these websites look authentic, we registered two domain names: *ebay-login.net* and *amazonaccounts.net*. The websites were designed to mimic the login pages of the original websites. We created two spoof URLs at each domain name.

We took steps to ensure our phishing websites triggered the warnings in each web browser. Firefox downloads its locally stored blacklist from Google, so we modified it locally to include our URLs [16]. Microsoft agreed to add our spoof URLs to their remote blacklists, causing those URLs to trigger the IE phishing warnings.

We copied two common phishing emails spoofing Amazon and eBay and changed the content to fit our study. The message claiming to be from Amazon was sent out in plain text and informed the recipient that the order was delayed and would be cancelled unless the recipient clicked the included URL. The message claiming to be from eBay was in HTML and informed the recipient that all international orders needed to be confirmed by visiting a URL contained within the message. Both messages contained random order numbers to help convince the recipients of their legitimacy, though no information specific to the recipients was included in these messages in order to make our attacks realistic. The scenario was such that it would have been entirely possible for a person to have just completed a purchase from one of these websites and then received a generic phishing message spoofing that same website. It is also possible for a phisher to monitor wireless Internet traffic and conduct a similar phishing attack after detecting a purchase. We believe that the coincidental nature of this attack was the reason why many more participants fell for our attacks than what has been found in similar studies [8, 7, 19, 14, 20]. Previous phishing studies have spoofed companies with whom victims had relationships. However we are unaware of any user studies that have used phishing messages timed to coincide with a transaction with the spoofed brand.

Participants arrived at our laboratory and were told that they would be purchasing two items online from Amazon and eBay. We randomized the order in which the purchases were made. We also informed participants that we were recording them, so they needed to think aloud about everything they were doing. Participants did the study individually with the experimenter sitting behind them in the laboratory.

We were concerned that if we allowed participants to purchase whatever they wanted, they might take too long to decide, and that other factors might confound our results. We also wanted participants to focus on buying cheap items so that we could reimburse them for both purchases while still giving them enough additional money for their time. We limited the scope of the purchases by asking them to purchase a box of paper clips from Amazon, which cost roughly $0.50, plus around $6 in shipping (the exact prices changed with each order since all participants did not purchase the same paperclips). We asked participants to make their eBay purchases from a cheap electronics store based in Hong Kong that sold a variety of items for around $5-$10, including shipping. Participants were compensated $35 for their time and the purchases, which were made using their personal credit cards.

After each purchase, participants received a sheet of five questions relating to shopping. These questions were part of an unrelated study on shopping behaviors, but helped our study by convincing participants that this was indeed a shopping study. While the participant answered these questions, the experimenter sent them a phishing message. We constructed a web interface for the study, so that the experimenter only needed to enter an email address, the brand to spoof, and the experimental condition.

After the written questions were completed, the experimenter told the participant to "check your email to make sure that the order is confirmed and ready to ship so we can move on." When participants checked their email, they encountered legitimate messages relating to their orders as well as a phishing message. After examining (and reacting) to all of

Condition Name	Size	Clicked	Phished
Firefox	20	20 (100%)	0 (0%)
Active IE	20	19 (95%)	9 (45%)
Passive IE	10	10 (100%)	9 (90%)
Control	10	9 (90%)	9 (90%)

Table 1. An overview depicting the number of participants in each condition, the number who clicked at least one phishing URL, and the number who entered personal information on at least one phishing website. For instance, nine of the control group participants clicked at least one phishing URL. Of these, all nine participants entered personal information on at least one of the phishing websites.

the messages, participants received a set of instructions for the second purchase. After participants checked their email after the second purchase (thereby encountering the second phishing message), an exit survey was administered. This online exit survey contained questions about participants' reactions to the warning messages. The experimenter observed participants fill this out and asked followup questions if any of the responses were too terse or did not seem to follow the behaviors exhibited during the experiment. Those in the control group were not asked to complete an exit survey as they had not seen any warnings. Participants took an average of forty minutes to complete all the tasks and were given $35 in cash before leaving.

We were initially concerned that since participants did not explicitly want the items, the results might be skewed in favor of participants acting more cautious. However, we believe their desire to complete the study negated this. Thus, the desire to buy the items to complete the study was likely just as strong as if the participant were at home purchasing a desired item. Additionally, we do not believe that the cost of the items played any role since an attacker could use the stolen account credentials to make any number of larger purchases.

RESULTS AND ANALYSIS

Overall we found that participants were highly susceptible to our spear phishing attack. However, users of the active phishing warnings were largely protected, since 79% chose to heed them. We found a significant difference between the active IE and Firefox warnings ($p < 0.0004$ for Fisher's exact test) as well as no significant difference between the passive IE warning and the control group (i.e. significantly more users were helped by the active Firefox warning than the active IE warning, while the passive IE warning is no different than not displaying any warning). We also found significant differences between the active IE warning and the control group ($p < 0.01$) demonstrating that the active IE warning is still significantly better than not displaying any warning. Table 1 depicts these results. In this section we examine how participants reacted to the initial phishing messages, and then we use the C-HIP model to analyze why certain warnings performed better than others.

Phishing Susceptibility

Our simulated spear phishing attack was highly effective: of the 106 phishing messages that reached participants' inboxes, participants clicked the URLs of 94 of them (89%). While all participants made purchases from both Amazon and eBay, not every participant received both of our phishing messages due to email filtering. Only two participants (3%) did not attempt to visit any of the phishing URLs. Of the 46 participants who received both phishing messages, 43 clicked the Amazon link and 37 clicked the eBay link. However this difference was not statistically significant, nor were there any significant correlations based on which phishing message was viewed first. It should also be noted that every participant in the control group who followed a link from an email message also submitted information to the phishing websites (Table 1). Thus, in the absence of security indicators, it is likely that this type of phishing attack could have a success rate of at least 89%.

With regard to the technical questions mentioned in the *Recruitment* section, we noticed a negative trend between technical experience and obeying the warnings among Internet Explorer users (i.e. users with more technical experience were more likely to ignore the warnings). With Firefox, technical experience played no role: all users obeyed the warnings regardless of their technical experience.

We did not actually collect any information entered into the phishing websites. Instead the experimenter observed each participant and noted when they submitted information. Thus we cannot conclusively say whether all participants entered their correct information. However, the experimenter did note that all usernames were entered correctly, and no participants denied entering their correct information when asked in the exit survey.

We found that participants had very inaccurate mental models of phishing. Both of our phishing messages contained language that said the orders would be cancelled if they did not visit the URLs. Thirty-two percent of the participants who heeded the warnings and left the phishing websites believed that their orders would be cancelled as a result—they believed that the emails were really sent from eBay and Amazon. We asked 25 of the participants how they believed the fraudulent URLs came to them, and only three recognized that the emails had been sent by someone not affiliated with either eBay or Amazon (we added this question halfway through the study). Thus, there seems to be some cognitive dissonance between recognizing a fraudulent website and the fraudulent email that spread it. This raises grave concerns about Internet users' susceptibility to phishing. Highly targeted phishing attacks will continue to be very effective as long as users do not understand how easy it is to forge email. At the same time, effective browser warnings may mitigate the need for user education, as we will now show.

Attention Switch and Maintenance

The first stage in the C-HIP model is "attention switch." If a warning is unable to capture the user's attention, the warning will not be noticed and thus be rendered useless. Unlike the passive indicators examined by Wu et al. [23], the active warnings in Firefox and Internet Explorer get the user's attention by interrupting their task—the user is forced to choose one of the options presented by the warning.

Condition Name	Sample Size	Saw Warning	Read Warning	Recognized Warning	Understood Meaning	Understood Choices
Firefox	20	20	13	4	17	19
Active IE	20	19	10	10	10	12
Passive IE	10	8	3	5	3	5

Table 2. This table depicts the number of participants in each experimental condition, the number who saw at least one warning, the number who completely read at least one warning, the number who recognized the warnings, the number who correctly understood the warnings, and the number who understood the choices that the warnings presented.

This was not the case with the passive warning in IE (Figure 2). This warning is a single dialog box with only the option to dismiss it. We observed that it could take up to five seconds for this warning to appear. If a user starts typing during this period, the user's keystrokes will inadvertently dismiss the warning. Six of the ten participants in this condition never noticed the warning because their focus was on either the keyboard or the input box. Two of these participants had this happen on both phishing websites, so they had no idea they were ever exposed to any warnings. We found no statistical significance between this condition and the control group. Thus, this type of warning is effectively useless.

Effective warnings must also cause attention maintenance—they must grab the users' attention long enough for them to attempt comprehension. We examined the number of participants who read the warnings (as determined by self-reporting and confirmed by the observations of the experimenter) in order to determine their effectiveness at attention maintenance. Table 2 shows the number of warnings read and the number of participants who claimed to have seen the warnings prior to this study, for each experimental condition.

Not counting the two participants who failed to notice the warnings entirely, and the participant in the active IE condition who did not click on the URLs, we found that twenty-six of the remaining forty-seven (55%) claimed to have completely read at least one of the warnings that were displayed. When asked, twenty-two of these twenty-six (85%) said they decided to read the warning because it appeared to warn about some sort of negative consequences.

Upon seeing the warnings, two participants in the active IE condition immediately closed the window. They went back to the emails and clicked the links, were presented with the same warnings, and then closed the windows again. They repeated this process four or five times before giving up, though never bothered to read the warnings. Both said that the websites were not working. Despite not reading or understanding the warnings, both were protected because the warnings "failed safely." Thus, if users do not read or understand the warnings, the warnings can still be designed such that the user is likely to take the recommended action.

Nineteen participants claimed to have previously seen these particular warnings. A significantly higher proportion of participants in the active IE condition (50%) claimed to have recognized the warnings as compared to participants in the Firefox condition (20%; $p < 0.048$ for Fisher's exact test). Many of the participants who encountered the active IE warning said that they had previously seen the same warning on websites which they trusted, and thus they ignored it. It is likely that they did not read this phishing warning because IE uses a similar warning when it encounters an expired or self-signed SSL certificate. Therefore they did not notice that this was a slightly different and more serious warning.

We found a significant negative Pearson correlation between participants recognizing a warning message and their willingness to completely read it ($r = -0.309$, $p < 0.03$). This implies that if a warning is recognized, a user is significantly less likely to bother to read it completely (i.e. habituation). Thus, very serious warnings should be designed differently than less serious warnings in order to increase the likelihood that users will read them. This was also the basis for Brustoloni and Villamarín-Salomón's work on dynamic warnings [3].

Warning Comprehension

A well-designed warning must convey a sense of danger and present suggested actions. In this study we asked participants what they believed each warning meant. Twenty-seven of the 47 participants (57%) who saw at least one of the warnings correctly said they believed that they had something to do with giving information to fraudulent websites (Table 2). Of the 20 participants who did not understand the meaning of the warnings, one said that she did not see it long enough to have any idea, while the others had widely varying answers. Examples include: "someone got my password," "[it] was not very serious like most window[s] warning[s]," and "there was a lot of security because the items were cheap and because they were international."

Using Fisher's exact test, we found that those using Firefox understood the meaning of the warnings significantly more than those exposed to the active IE warnings ($p < 0.041$) and the passive IE warnings ($p < 0.005$), though we found no significant difference between the active and passive IE warnings. We found a significant Pearson correlation between completely reading a warning and understanding its meaning for the active IE warning ($r = 0.478$, $p < 0.039$), but not for Firefox. Since all but one Firefox user correctly understood what the warning wanted them to do, this implies that users did not need to completely read it to know the appropriate actions to take.

Overall, 31 of the 47 participants who noticed the warnings mentioned that they thought they were supposed to leave the website or refrain from entering personal information. Those who did not understand the warnings provided responses such as "panic and cancel my accounts," "confirm information about the orders," and "put in my account information so that they could track it and use it for themselves."

Attitudes and Beliefs

We asked participants how their attitudes and beliefs influenced their perceptions and found a highly significant correlation between trusting and obeying the warnings (i.e. users who did not trust the warnings were likely to ignore them; $r = 0.76$, $p < 0.0001$). More telling, all but three participants who ignored a warning said it was because they did not trust the warning. Two of the participants who ignored the warnings in the active IE group said they did so because they trusted them but thought the warnings were not very severe ("since it gave me the option of still proceeding to the website, I figured it couldn't be that bad"). The other participant who trusted the warning yet ignored it was in the passive IE group and blamed habituation ("my own PC constantly bombards me with similar messages"). All three of these participants questioned the likelihood of the risks, and thus were more interested in completing the primary task.

We found a significant correlation between recognizing and ignoring a warning ($r = 0.506$, $p < 0.0003$). This further implies that habituation was to blame when participants ignored warnings: they confused them with similar looking, but less serious warnings, and thus did not understand the level of risk that these warnings were trying to convey. This was only a problem for the warnings used by IE, as all the Firefox users obeyed the warnings (though only 20% claimed to have seen them before, compared to the 50% with IE). The IE users who ignored the warnings made comments such as:

- "Oh, I always ignore those"
- "Looked like warnings I see at work which I know to ignore"
- "Have seen this warning before and [it] was in all cases [a] false positive"
- "I've already seen such warnings pop up for some other CMU web pages as well"
- "I see them daily"
- "I thought that the warnings were some usual ones displayed by IE"

A warning should not require domain knowledge for a user to understand it. In order to examine whether prior knowledge of phishing impacted user attitudes towards the warnings, we asked them to define the term "phishing." Twenty-six of the forty-seven participants who noticed the warnings were able to correctly say they had something to do with using fraudulent websites to steal personal information. We calculated Pearson's correlation coefficient and found a significant correlation between knowing what phishing is and both reading ($r = 0.487$, $p < 0.0005$) and heeding ($r = 0.406$, $p < 0.005$) the warnings. Thus, if a user does not understand what phishing is, they are less likely to be concerned with the consequences, and thus less likely to pay attention to the warning.

Motivation and Warning Behaviors

Table 1 depicts the number of participants from each condition who fell for at least one phishing message. Some participants only clicked on one of the two phishing messages, and in other cases some participants only received one phishing message due to email filtering.

Overall we found that active phishing warnings were significantly more effective than passive warnings ($p < 0.0002$ for Fisher's exact test). We showed the passive Internet Explorer warning to ten different participants, but only one participant heeded it and closed the website, whereas the other times participants dismissed it and submitted personal information to the phishing websites (in two of these cases participants failed to notice the warnings altogether). We found that this passive warning did not perform significantly different than the control group ($p < 1.0$ for Fisher's exact test). The active IE warning was ignored by nine participants, while in the Firefox condition every participant heeded the warning and navigated away from the phishing websites. This was a highly significant difference ($p < 0.0004$, for Fisher's exact test), however the active IE warning still performed significantly better than the control condition ($p < 0.01$) and the passive IE warning ($p < 0.044$).

Qualitatively, we examined why participants were motivated to heed or ignore the warnings. A total of thirty-one participants chose to heed the warnings, and in twenty-three of these cases participants said that the warnings made them think about risks:

- "I didn't want to get burned"
- "...it is not necessary to run the risk of letting other potentially dangerous sites to get my information"
- "I chose to heed the warning since I don't like to gamble with the little money I have"
- "I felt it better to be safe than sorry"
- "I heeded the warning because it seemed less risky than ignoring it"

Participants who chose to submit information said that they did so because they were unaware of the risks (i.e. they did not read the warnings), were used to ignoring similarly designed warnings (i.e. habituation), or they did not understand the choices that the warnings presented.

Environmental Stimuli

In the passive IE condition, three of the participants who ignored the warnings said they did so because they incorrectly placed some degree of trust in the phishing website because of stimuli other than the warning messages. When asked why they chose to ignore the warnings, one participant said she had "confidence in the website." Another participant ignored the warning "because I trust the website that I am doing the online purchase at." These answers corroborate Fogg's work, showing that the look and feel of a website is often the biggest trust factor [10]. Participants who ignored the active IE warning provided similar answers, and also said that they ignored the warnings because they trusted the brands that the emails had spoofed.

We also found that when some participants saw the warnings, they examined other security context information before making a decision. One Firefox user reexamined the original phishing email and noticed the lack of any personalized information. She then decided to "back out and log in from the root domain to check." After seeing the warnings, ten other Firefox users also examined either the URL bar or the email headers. Some observations included: "The

URL did not match the usual eBay URL and so it could be fraudulent;" "I did look at the URL that I opened from the email, and the sender of the email, to confirm that they did look suspicious;" and "it made me look at the web address which was wrong." One participant in the passive IE condition and three in the active IE condition incorrectly used this information to fall for the phishing attacks. Some of the comments included: "The address in the browser was of amazonaccounts.com which is a genuine address" and "I looked at the URL and it looked okay."

Finally, at least four participants claimed that the timing of the phishing emails with the purchases contributed to them ignoring the warnings. It is unclear how susceptible these participants would have been to a broader phishing attack, rather than the targeted attack that we examined.

DISCUSSION

In this section we provide some recommendations for improving the design of phishing indicators based on the results of our study.

Interrupting the primary task — Phishing indicators need to be designed to interrupt the user's task. We found that the passive indicator, which did not interrupt the user's task, was not significantly different than not providing any warning. The active warnings were effective because they facilitated attention switch and maintenance.

Providing clear choices — Phishing indicators need to provide the user with clear options on how to proceed, rather than simply displaying a block of text. The users that noticed the passive Internet Explorer warning, read it but ignored it because they did not understand what they were supposed to do. They understood it had something to do with security, but they did not know how to proceed. In contrast, the active warnings presented choices and recommendations which were largely heeded. Wu found similar results with regard to providing users with clear choices [22].

Failing safely — Phishing indicators must be designed such that one can only proceed to the phishing website after reading the warning message. Users of the active Internet Explorer warning who did not read the warning or choices could only close the window to get rid of the message. This prevented them from accessing the page without reviewing the warning's recommendations. However, users of the passive Internet Explorer warning had the option of clicking the familiar 'X' in the corner to dismiss it without reading it, and accessing the page anyway.

Preventing habituation — Phishing indicators need to be distinguishable from less serious warnings and used only when there is a clear danger. Users ignored the passive indicators because they looked like many other warnings that users have ignored without consequences, thus they appear to be "crying wolf." Even the active Internet Explorer warning was not read in a few cases because users mistook it for other IE warnings. More people read the Firefox warnings because they are designed unlike any other warnings. Dynamic warning messages may help prevent habituation [3].

Altering the phishing website — Phishing indicators need to distort the look and feel of the website such that the user does not place trust in it. This can be accomplished by altering its look or simply not displaying it at all. The overall look and feel of a website is usually the primary factor when users make trust decisions [10]. When the website was displayed alongside the passive indicators, users ignored the warnings because they said that they trusted the look of the website.

CONCLUSION

This study has given us insights into creating effective security indicators within the context of phishing. Such indicators are clearly needed as 97% of participants believed the phishing emails enough to visit the URLs. Of the participants who saw the active warnings, 79% chose to heed them and close the phishing websites, whereas only 13% of those who saw the passive warnings obeyed them. Without the active warning indicators, it is likely that most participants would have entered personal information. However, the active indicators did not perform equally: the indicators used by Firefox performed significantly better than the active warnings used by IE, though both performed significantly better than the passive IE warnings (which was not significantly different from not showing any warnings in the control group).

As phishing attacks continue to evolve, it is likely that highly targeted attacks will become more prevalent. Future indicators within the phishing context need to be designed such that they interrupt the user's primary task, clearly convey the recommended actions to take, fail in a secure manner if the user does not understand or ignores them, draw trust away from the suspected phishing website, and prevent the user from becoming habituated.

ACKNOWLEDGMENTS

Thanks to the members of the Supporting Trust Decisions project for their feedback, and Matthew Williams for his assistance. This work was supported in part by the National Science Foundation under grant CCF-0524189. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the National Science Foundation or the U.S. government.

REFERENCES

1. AMER, T. S., AND MARIS, J. B. Signal words and signal icons in application control and information technology exception messages – hazard matching and habituation effects. Tech. Rep. Working Paper Series–06-05, Northern Arizona University, Flagstaff, AZ, October 2006.
2. BANK OF AMERICA. How Bank of America SiteKey Works for Online Banking Security. http://www.bankofamerica.com/privacy/sitekey/, 2007.
3. BRUSTOLONI, J. C., AND VILLAMARÍN-SALOMÓN, R. Improving security decisions with polymorphic and audited dialogs. In *SOUPS '07: Proceedings of the 3rd symposium on Usable privacy and security* (New York, NY, USA, 2007), ACM Press, pp. 76–85.

4. CERTIFICATION AUTHORITY/BROWSER FORUM. Extended validation ssl certificates, Accessed: July 27, 2007. http://cabforum.org/.
5. CRANOR, L. F. What do they "indicate?": Evaluating security and privacy indicators. *Interactions 13*, 3 (2006), 45–47.
6. DHAMIJA, R., AND TYGAR, J. D. The battle against phishing: Dynamic security skins. In *Proceedings of the 2005 Symposium on Usable Privacy and Security* (New York, NY, USA, July 6-8 2005), ACM Press.
7. DHAMIJA, R., TYGAR, J. D., AND HEARST, M. Why phishing works. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems* (New York, NY, USA, 2006), ACM Press, pp. 581–590.
8. DOWNS, J. S., HOLBROOK, M., AND CRANOR, L. Decision Strategies and Susceptibility to Phishing. In *Proceedings of The 2006 Symposium on Usable Privacy and Security* (Pittsburgh, PA, July 12-14, 2006).
9. FLORENCIO, D., AND HERLEY, C. A large-scale study of web password habits. In *WWW '07: Proceedings of the 16th international conference on World Wide Web* (New York, NY, USA, 2007), ACM Press, pp. 657–666.
10. FOGG, B., MARSHALL, J., LARAKI, O., OSIPOVICH, A., VARMA, C., FANG, N., PAUL, J., RANGEKAR, A., SHON, J., SWANI, P., AND TREINEN, M. What Makes Web Sites Credible? A Report on a Large Quantitative Study. In *Proceedings of the ACM Computer-Human Interaction Conference* (Seattle, WA, March 31 - April 4, 2001), ACM.
11. GARTNER, INC. Gartner Says Number of Phishing E-Mails Sent to U.S. Adults Nearly Doubles in Just Two Years. http://www.gartner.com/it/page.jsp?id=498245, November 9 2006.
12. HELLIER, E., WRIGHT, D. B., EDWORTHY, J., AND NEWSTEAD, S. On the stability of the arousal strength of warning signal words. *Applied Cognitive Psychology 14* (2000), 577–592.
13. JACKSON, C., SIMON, D., TAN, D., AND BARTH, A. An evaluation of extended validation and picture-in-picture phishing atttacks. In *Proceedings of the 2007 Usable Security (USEC'07) Workshop* (February 2007). http://www.usablesecurity.org/papers/jackson.pdf.
14. KUMARAGURU, P., RHEE, Y., ACQUISTI, A., CRANOR, L. F., HONG, J., AND NUNGE, E. Protecting people from phishing: the design and evaluation of an embedded training email system. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2007), ACM Press, pp. 905–914.
15. MOORE, T., AND CLAYTON, R. An empirical analysis of the current state of phishing attack and defence. In *Proceedings of the 2007 Workshop on The Economics of Information Security (WEIS2007)* (May 2007). http://www.cl.cam.ac.uk/ twm29/weis07-phishing.pdf.
16. OBERHEIDE, J. Google safe browsing, November 6 2006. http://jon.oberheide.org/blog/2006/11/13/google-safe-browsing/.
17. OPENDNS. PhishTank Annual Report. http://www.phishtank.com/, October 2007.
18. REFSNES DATA. Browser statistics, Accessed: April 4, 2007. http://www.w3schools.com/browsers/browsers_stats.asp.
19. SCHECHTER, S. E., DHAMIJA, R., OZMENT, A., AND FISCHER, I. The emperor's new security indicators. In *Proceedings of the 2007 IEEE Symposium on Security and Privacy* (May 2007).
20. SHENG, S., MAGNIEN, B., KUMARAGURU, P., ACQUISTI, A., CRANOR, L., HONG, J., AND NUNGE, E. Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 2007 Symposium On Usable Privacy and Security* (Pittsburgh, PA, July 18-20, 2007), ACM Press.
21. WOGALTER, M. S. Communication-Human Information Processing (C-HIP) Model. In *Handbook of Warnings*, M. S. Wogalter, Ed. Lawrence Erlbaum Associates, 2006, pp. 51–61.
22. WU, M. *Fighting Phishing at the User Interface*. PhD thesis, Massachusetts Institute of Technology, August 2006.
23. WU, M., MILLER, R. C., AND GARFINKEL, S. L. Do Security Toolbars Actually Prevent Phishing Attacks? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Held in Montreal* (2006), ACM Press, pp. 601–610.
24. YE, Z. E., AND SMITH, S. Trusted paths for browsers. In *Proceedings of the 11th USENIX Security Symposium* (2002), pp. 263–279.
25. YEE, K.-P., AND SITAKER, K. Passpet: Convenient password management and phishing protection. In *SOUPS '06: Proceedings of the Second Symposium on Usable Privacy and Security* (New York, NY, USA, 2006), ACM Press, pp. 32–43.
26. ZHANG, Y., EGELMAN, S., CRANOR, L. F., AND HONG, J. Phinding phish: Evaluating anti-phishing tools. In *Proceedings of the 14th Annual Network & Distributed System Security Symposium (NDSS 2007)* (28th February - 2nd March, 2007). http://lorrie.cranor.org/pubs/toolbars.html.