2007

Behavioral Response to Phishing Risk

Julie S. Downs
Carnegie Mellon University

Mandy Holbrook
Carnegie Mellon University

Lorrie F. Cranor
Carnegie Mellon University

Behavioral Response to Phishing Risk

Julie S. Downs
Carnegie Mellon University
Social & Decision Sciences
Pittsburgh, PA 15213
1-412-268-1862
downs@cmu.edu

Mandy Holbrook
Carnegie Mellon University
Social & Decision Sciences
Pittsburgh, PA 15213
1-412-268-3249
holbrook@andrew.cmu.edu

Lorrie Faith Cranor
Carnegie Mellon University
Computer Science & EPP
Pittsburgh, PA 15213
1-412-268-7534
lorrie@cmu.edu

ABSTRACT

Tools that aim to combat phishing attacks must take into account how and why people fall for them in order to be effective. This study reports a pilot survey of 232 computer users to reveal predictors of falling for phishing emails, as well as trusting legitimate emails. Previous work suggests that people may be vulnerable to phishing schemes because their awareness of the risks is not linked to perceived vulnerability or to useful strategies in identifying phishing emails. In this survey, we explore what factors are associated with falling for phishing attacks in a role-play exercise. Our data suggest that deeper understanding of the web environment, such as being able to correctly interpret URLs and understanding what a lock signifies, is associated with less vulnerability to phishing attacks. Perceived severity of the consequences does not predict behavior. These results suggest that educational efforts should aim to increase users' intuitive understanding, rather than merely warning them about risks.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: *Psychology;* H.1.2 [**User/Machine Systems**]: *Software psychology;* K.4.4 [**Electronic Commerce**]: *Security*

General Terms

Security, Human Factors

Keywords

Phishing, Survey

1. INTRODUCTION

Phishing is a type of *semantic attack* [1] in which attackers attempt to exploit the naiveté of some Internet users rather than exploiting bugs in computer software. Phishing attacks get more sophisticated over time as attackers learn what techniques are most effective and alter their strategies accordingly. Those working to stop phishing have less information than the attackers about how users respond to various types of attacks. However, knowledge of users' behavioral response is useful for developing techniques to educate users about phishing, for developing toolbars and other software designed to provide phishing-related warning indicators that users will actually pay attention to, and perhaps even for developing automated detection systems. In this paper we present the results of a preliminary survey designed to measure the behavioral response to phishing across a large population of Internet users.

Our present study is a pilot survey for a planned large-scale phishing survey. Although the pilot used a convenience sample that was limited in size and diversity, it already provides a number of interesting insights into behavioral response to phishing risk and allows us to examine relationships between demographic, experience, and behavioral factors that have not been previously studied. Most interestingly, we found that knowledge and experience predict behavioral responses to phishing attacks in ways that support the idea that better understanding can help to thwart such attacks.

2. METHODS

2.1 Participant Recruitment

Members of the Carnegie Mellon University community who registered for the Cyber Security Summit on campus were invited by email to participate in an on-line survey. Attendees at the summit included a diverse group of faculty, staff and students, including people who were concerned about computer security as well as those who participated to make amends for violating computing policy such as exceeding the allotted bandwidth. Each participant was offered a chance to win one of several prizes, including an LCD TV or an iPod.

2.2 Procedure

If participants were interested in participating in the survey, they visited the URL given to them in the invitation email. Two hundred and thirty-two participants completed the survey, with 180 completing half of the items (counterbalanced) prior to the Summit and half following, and the remainder completed all items following the Summit. No differences were observed between responses prior and following the Summit, so all data were collapsed. At the end of the survey, university administrators held the drawing and awarded the prizes.

2.3 Materials

The content of the survey was derived from findings from an open-ended interview study. The survey consisted of several sections: an email role play where respondents responded to images of emails and web sites, a URL evaluation section where respondents identified features of URLs, a section asking how respondents would react to different warning messages, a knowledge section where respondents interpreted the meaning of lock icons and jargon words, past experience with web sites, and ratings of potential negative consequences of phishing. All images

APWG eCrime Researchers Summit, October 4-5, 2007, Pittsburgh, PA, USA.

were modeled on an Internet Explorer browser in a Microsoft Windows operating system, to be familiar to the greatest number of participants.

2.3.1 Email and web role play

Participants were asked to role play a person named Pat Jones, who worked at a company called Cognix. They were shown images of five emails from Pat's inbox, each with a brief context for them to use in deciding how they would respond if they were Pat. Each email displayed one key URL link, and was shown with the mouse pointer positioned over the link and the actual URL displayed in the status bar. Attention was not directed to the status bar, but the information was there for those who chose to look. The relevant features of the emails are summarized in Table 1.

Table 1. Relevant features of five emails and corresponding web sites from email and web role play

Email	Legitimacy	Relevant features of email and sites
Cognix	real	•regarding work details •link in email: www.cognix.com •URL in status bar: http://www.cognix.com
NASA	real	•sender is known person •addressed to user •link in email: "this" •URL: antwrp.gsfc.nasa.gov/apod /astropix.html
eBay	real	•registered name "Pat Jones" displayed •link in email: "PAY [Click to confirm…]" •URL: https://payments.ebay.com/ws/eBay ISAPI.dll?item=6600378513
PayPal	phishing	•urgent request •lock image in body of web page •link: "Click here to activate your account" •URL: http://payaccount.me.uk/cgi-bin/ webscr.htm?cmd=_login-run
laptop	spear phishing	•generic message about eBay item •link: www.set-ltd.net •URL: www.set-ltd.net

Respondents were given seven options to indicate how they would respond: reply by email, contact the sender by phone or in person, delete the email, save the email, click on the link, copy and paste the URL, or type the URL into a browser window. They were permitted to check multiple responses (for example, if they would reply and then delete the email, they would check both options). They were also given the option not to answer the question (which fewer than 2% of respondents chose), and the option to indicate an "other" response where they could offer more sophisticated responses. We chose to limit the available options to relatively simple strategies in order to avoid teaching participants about new strategies. Between 5-15% of respondents gave an "other" response to each email. We were interested in particular in whether people said that they would click on the link in the email.

Those who indicated that they would click on a link in the website were shown an image of the web page that they would see by clicking on that link. In cases where the link displayed in the email was the same as the URL it referenced, those who indicated that they would copy and paste or type in the link were also shown the relevant page. On this page they were asked what they would do at the website. They were given six options: click on one or more links on the page, enter the information requested by the page, bookmark the page, save or archive the page, visit a related page, or leave the website. As before, they were also given the option to decline to answer or to provide another response.

The survey was presented as one of computer use, rather than computer security. Prior to the role play, respondents were asked whether they had ever had a misunderstanding due to communicating with email. The first email was intended to familiarize them with the procedure and create the impression that the study was about mundane aspects of email, such as etiquette, rather than security. It was a message from "Jean Smith" <jsmith@cognix.com> addressed to marketing team members, advising that a 3pm meeting had been moved to 3:30. The mouse pointer was positioned over a link displayed as www.cognix.com in the email and http://www.cognix.com in the status bar. The context provided noted that Pat's calendar showed a meeting at 3pm labeled "Cognix meeting in room 390." Respondents who indicated that they would click on the link, copy and paste it, or type it into a browser were then sent to a page with a website for the Cognix company. This page did not ask for any information to be entered, so the option of entering information was not included.

The second email image listed the sender as "Andrew Williams" <andrewwilliams@cognix.com>. The message read, "Hey Pat, I saw this and thought of you." The mouse pointer was positioned over the word "this" and displayed http://antwrp.gsfc.nasa.gov/ apod/astropix.html in the status bar. The content provided noted that Pat's address book has an entry for a co-worker named 'Andrew Williams'. Respondents who indicated that they would visit the URL in some regard were taken to NASA's Astronomy Picture of the Day. This web page did not ask for any information to be entered.

The third email was a legitimate message from eBay stating that Pat had won a bid on a Dell Latitude Laptop. The context provided noted Pat had an account with eBay. By choosing any of the URL options, they were taken to eBay's "Congratulations you've won" page. The web page included links for signing in and details about the winning bid and the item won.

The fourth email was a spoof PayPal message appearing to be from "service@paypal.com"<service@paypal.com>. The message was addressed generically as a typical phishing message might be, 'Dear PayPal User' and the context said that Pat had used PayPal in the past. The mouse over the hyperlink showed the spoof URL, http://payaccount.me.uk/cgi-bin/webscr.htm?cmd=_login-run in the status bar. The URL did not appear in the email text, rather the words, "Click here to activate your account". If the participant chose to click on the link, then they would be taken to a fake website resembling PayPal's login screen, which asked for their email address and password for their PayPal account. This web site had broken images, a lock image inside the web page and a foreign URL ending with ".uk".

In the last email, the context noted that Pat recently put an old laptop on eBay to try to sell it. This type of more targeted attack, where the content of the message is tailored to the recipient, is referred to as a "spear phishing" email. It read,

> *"Hello I am very interested in your unit. I would like to know your best price to buy it now and if you ship international. I am in London UK right now. I would also want to know the condition of the unit. I can pay with an escrow service that can handle the transaction. I am already registered with www.set-ltd.net and my user name is the same with my email address.*

> *Please let me know I am very interested. Thank you in advance. Gianluca."*

By clicking on the URL in the email, www.set-ltd.net, the participant was taken to a web site that spoofed an escrow service and asked them to enter their bank account number and route information.

After completing the role play, respondents were asked to indicate whether Pat's emails were similar to the kind of email that they received, and whether they responded differently than they would have in their own email. The response format was an open-ended text box where respondents could provide simple yes or no answers, or provide a more detailed explanation if they wished.

2.3.2 URL evaluation

Respondents were then asked about "web site addresses, also called URLs." They were shown a series of four URLs (shown in Table 2) and asked what they could tell about the web site, based just on the URL. Participants were asked not to open the URL, but if they clicked on it they received the message, "Please answer the question just from looking at the URL. You do not need to open the web site."

Respondents were given eight options and could check more than one option if they wanted. Options of what they could determine by just the URL included:

- I can tell what company the site belongs to.
- I can tell something about where this website is hosted.
- I can tell that this site is secure.
- I can tell that this site is NOT secure.
- I can be pretty sure that this site is trustworthy.
- I can be pretty sure that this site is NOT trustworthy.
- It's not possible to tell any of these things just from the URL.
- I don't know how to tell any of these things just from the URL.

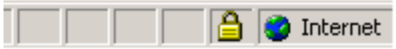
Participants could also decline to answer or list other, more specific details.

Table 2. URLs evaluated by participants

URLS evaluated
http://cgi.ebay.com/ws/eBayISAPI.dll?ViewItem&item=660037851
http://antwrp.gsfc.nasa.gov/apod/astropix.html
http://www.payaccount.me.uk/cgi-bin/webscr.htm?cmd=_login-run
http://www.ebay.me.uk/cgi-bin /webscr.htm?cmd=_login-run

2.3.3 Knowledge of icons and jargon

Participants were shown an image of the lock icon found within the chrome area of the browser and asked if they had seen "this lock image" before:



Respondents were asked four questions about what this lock image meant about a web site:

- that you need a key or a password to enter the site
- that a website is trustworthy
- that any information you enter will be sent securely
- that any information being displayed will be sent securely

The next question showed three other lock images, none of which were in the chrome area of a browser, and asked the participants which of those lock images represented the same meanings as the lock they had initially seen, with the option "none of these" included.

A series of knowledge questions asked participants to choose the best definition for four computer related terms: cookie, spyware, virus and phishing Participants were given the same list of eight possible definitions to choose from for each definition, as well as options to indicate familiarity with the word or lack thereof. Each term had one correct answer on the list. These options included:

- Something that protects your computer from unauthorized communication outside the network
- Something that watches your computer and sends that information over the Internet (*spyware*)
- Something websites put on your computer so you don't have to type in the same information the next time you visit (*cookie*)
- Something put on your computer without your permission, that changes the way your computer works (*virus*)
- Email trying to trick you into giving your sensitive information to thieves (*phishing*)
- Email trying to sell you something
- Other software that can protect your computer
- Other software that can hurt your computer
- I have seen this word before but I don't know what it means for computers
- I have never seen this word before
- Decline to answer
- Other (please specify)

2.3.4 Past web experience and consequences

Respondents were asked about their previous on-line experience. They indicated whether they had ever purchased anything on the web before, whether they had an active account with PayPal, and whether they had ever used eBay to either purchase or sell anything.

Participants were asked about a few security measures they may have implemented in the past. They indicated whether they had ever installed antivirus software on their computer or had ever adjusted their security preferences in their web browser.

Participants were asked if they had experienced any of a list of possible negative consequences of having information stolen, such as having their credit card number or bank account information stolen, their social security number stolen or someone trying to steal their identity, or whether they were aware that any of their information had ever been stolen or compromised in some way by entering it into a web site.

2.3.5 Negative consequences

Participants were asked to rate a series of possible outcomes from computer malice. To avoid ceiling effects, the first question asked about extremely negative outcomes, and the remaining questions asked about a range of negative consequences. Respondents used a 7-point scale to indicate how bad it would be if:

- all of your money and belongings were stolen and you had no insurance or way to get any of it back
- someone stole your credit card number or made bad charges on your credit card
- someone got your bank account number and PIN

- someone stole your social security number or your identity
- your computer automatically sent bad software to everyone in your online address book
- someone you didn't know could see everything that you typed on your computer
- your computer started to crash several times a day

3. RESULTS

3.1 Past Web Behavior

Most participants (85%) were PC users, and most of the rest (12%) were Mac users. Almost all (98%) respondents had made a purchase on the web at some point in the past. Most had also engaged to some degree with altering the protections on their computer, such as installing anti-virus software (93%) or adjusting their security preferences (79%), although the latter could be making the security settings tighter or more lax.

A minority had experienced some negative consequences of the type threatened by phishing, although not necessarily as a result of phishing. They reported having had a credit card stolen (21%), having had information stolen or compromised (14%), and a small number reported having had their social security number stolen or had someone try to steal their identity (3%).

3.2 Validity of Role Play

One strength of the methodology used in this study is the embedded role play, which goes beyond mere self-report measures of behavior. Although not identical to real behavior in the wild, this tool is valuable to the extent that is provides a close approximation of real world patterns of behavior.

3.2.1 Face validity

In response to the open-ended question about whether Pat's emails were similar to the respondent's own and whether the respondent treated the role play any differently from his or her own email, over 95% of participants gave a meaningful response. A minority of respondents (25%) gave simple responses indicating that this experience was similar. The remainder gave explanations, many indicating the similarity, e.g., "I have received some emails like Pat's and I responded in the way I always do." Others described superficial differences between the role play emails and their own email, e.g., "this is how I would have responded, except I get a lot more spam." Others indicated differences between their own experience and the role play, with explanations of how they interpreted things that may have been different from their own experience, e.g.,

> *I would generally respond the same way as Pat would. I personally don't have an eBay or PayPal account but know many people that do, and generally believe that it's a large site, so it's safe and reliable.*

Although we cannot be certain that people respond identically to these role play emails as they would to real emails with consequences for their own information, it does appear that people took this task seriously and made a reasonable effort to respond as they would normally respond. Thus, patterns of responses to these emails should be roughly indicative of respondents' relative sophistication

3.2.2 Convergent and divergent validity

We assessed respondents' reported link clicks and other behaviors that would land them at the web site in emails (such as typing in the URL of the link) with related themes, controlling for their overall propensity to visit sites. We assessed overall visiting propensity by their response to the initial email, which was a simple email about changing a meeting that contained a link to the company website in the signature line.

Those who visited either of the two sites from valid links (for the NASA site and the eBay site) were significantly more likely to visit the other, r=.33, *p*<.01. Similarly, those who visited either of the two pages in financial emails (the eBay and PayPal sites) were more likely to visit the other, r=.36, *p*<.001. These results show convergent validity in the role play task. Importantly, however, the two sites without shared attributes (the NASA and PayPal sites) were not significantly correlated in frequency of visits, r=.15. This shows divergent validity, suggesting that this measure does not merely capture a willingness to visit links in emails, but rather a differentiation between different kinds of emails. A similar pattern, with slightly stronger correlations, was found when limiting the behavior to clicking on links.

The clicks and website visits for the spear phishing email showed no clear pattern in relation to the other emails, suggesting that this behavior is not very reliable. This indicates that the spear phishing email may be problematic in this test. However, the sensible pattern of partial correlations among the other three emails provides convergent and divergent validity to their use in the role play. Thus, we will use the behavior of clicking on the link in the phishing (PayPal) email, visiting that site by any means, and entering information on the phishing web site as proxies for susceptibility to phishing attacks.

3.3 Predictors of Behavior in Role Play

In this section we explore what factors are predictive of susceptibility to phishing, as measured by the behavior of clicking on the link in the phishing email in the role play, and of warranted email responses, as measured by the behavior of clicking on the links in the two legitimate emails. Overall rates at which respondents fell for the phishing scams do not represent meaningful data about behavior, as they are driven by the content of the stimuli used rather than a representative sample of phishing emails from the wild. Thus, only comparisons are reported, to indicate the degree to which different predictors were related to behavior.

3.3.1 Knowledge

Those who correctly answered the knowledge question about the definition of phishing were significantly less likely to fall for phishing emails, either by clicking on the phishing link (χ^2=10.90, *p*<.001), visiting the site (χ^2=4.62, *p*<.05), or entering information on the website (χ^2=9.79, *p*<.01), as shown in Figure 1.

However, knowledge about other computer risks and concepts was unrelated to clicking on the phishing link, whether about cookies (χ^2=0.3), spyware (χ^2=0.4), or viruses (χ^2=0.3). This pattern of results suggests that those who recognized the term "phishing" were also familiar enough with the concept to better protect themselves against the risk, but that general knowledge about other computer risks was not associated with better protection.

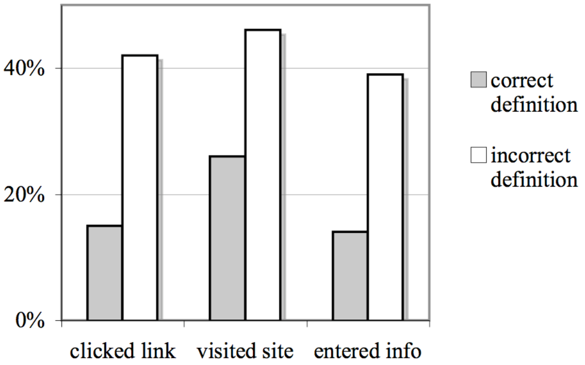


Figure 1. Behavior of participants who gave correct and incorrect definitions of phishing

Participants who correctly reported that none of the 3 non-chrome lock images meant the same thing as the standard lock image in chrome were less likely to fall for phishing emails, either by clicking on the link (χ^2=8.06, *p*<.01), visiting the phishing site (χ^2=10.39, *p*<.001), or entering their information (χ^2=10.25, *p*<.001), as shown in Figure 2.

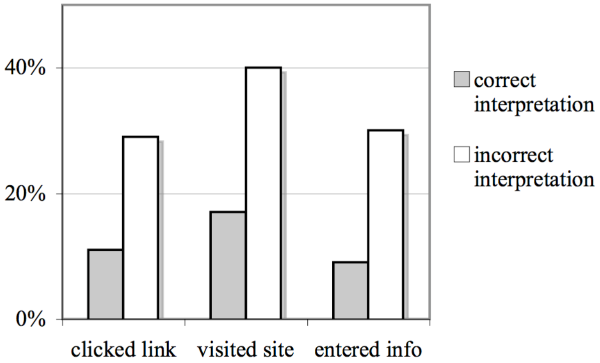


Figure 2. Behavior of participants who gave correct and incorrect interpretations of non-chrome lock images

This measure of knowledge was correlated with having the correct definition of phishing (*r*=.20, *p*<.01), so we conducted a logistic regression predicting each phishing behavior from both terms and found that even when controlling for knowledge of phishing (ß=-1.27, *p*<.01), knowledge of the lock images still predicted entering information on the phishing site (ß=-1.29, *p*<.05) with the same pattern and significance levels for clicking on the phishing link, and a slightly stronger effect of knowledge of the lock images predicting visiting the phishing site.

Those who had experience with commonly spoofed sites were less likely than others to click on the phishing links, including those with PayPal accounts (15% vs. 31%, χ^2=6.79, *p*<.05) and those who reported ever having used eBay in the past (16% vs. 33%, χ^2=7.13, *p*<.01). Similar results were found for visiting the phishing site and entering information there. These people have likely seen phishing attacks in the past that they have needed to evaluate as possibly relevant, given their own personal accounts. Whether this exposure merely raised their awareness or fooled them the first time, it seems to have educated them. This suggests that new attacks – whether just new to the individual or newly developed – are likely to be bigger threats.

3.3.2 Use of situational cues

There were various cues present that people could use to evaluate the emails in the role play, most notably the linked URL as shown in the status bar. We don't know how many people used that information, since we couldn't easily ask them about it without alerting them to its relevance. However, because we asked them about these URLS at a later point in the survey, we can explore the relationship between their assessments of the URLs and their behavior in the role play to determine whether their judgments were correlated with their behavior.

For the legitimate emails, those who believed that they could tell that the site was trustworthy were more likely to visit the site than others, both for the legitimate email from eBay (63% vs. 40%, χ^2=7.67, *p*<.01), and for the email linking to NASA's site (64% vs. 37%, χ^2=11.00, *p*<.001). Once at the eBay site, those thinking the URL was trustworthy were also more likely to say that they would enter the requested login information (33% vs. 16%, χ^2=6.43, *p*=.01).

For the phishing email, thinking that the linked site was *un*trustworthy was a significant predictor of behavior, with those thinking it was untrustworthy being less likely to click on the link than others (9% vs. 30%, χ^2=10.30, *p*<.001), and less likely to enter information on the page (8% vs. 29%, χ^2=11.45, *p*<.001). Additionally, those who recognized from the URL that the site was not secure were less likely to click on the link than others (12% vs. 33%, χ^2=11.74, *p*<.001), and less likely to enter information on the page (8% vs. 34%, χ^2=17.80, *p*<.001).

3.3.3 Perceptions of negative consequences

The consequences of having one's social security number stolen was rated as significantly worse than any other consequences, not significantly different from the range-defining question about having all of one's belongs stolen with no insurance to recover it.

Table 3 indicates how bad each consequence was seen, in descending order. Interestingly, having a credit card number stolen was perceived as the least serious outcome, significantly lower even that having a computer crash often.

Table 3. Perceived severity of negative consequences

Consequence	**Perceived Severity**
All belongings stolen	6.73$_a$
Social security number stolen, ID compromised	6.65 $_a$
Someone seeing what you type	6.43 $_b$
PIN number stolen	6.43 $_b$
Computer crashing often	6.21 $_c$
Malware sent to all addresses in address book	6.05 $_c$
Credit card number stolen	5.77 $_d$

Note: Means with different subscripts are significantly different from one another at the *p*<.05 level.

These perceived negative consequences were generally unrelated to any of the behaviors relating to falling for the phishing email. The exception to this was a significant zero-order relationship between perceived severity of having a credit card number stolen and willingness to enter information onto the PayPal phishing site, although the pattern was the opposite of what might be expected (those who entered the information rated the perceived consequences of having it stolen as more negative). This relationship can be explained by a confounding variable: knowledge about phishing attacks. Those who knew about phishing attacks were less likely to enter their information (as reported above), and also rated the consequences of having a credit card number stolen as less negative. Entering both perceived consequences and knowledge of phishing into a logistic regression reveals that knowledge of phishing predicts not entering the information (ß=-1.31, *p*<.01), and perceived consequences do not significantly predict behavior (ß=.79, *p*=.07).

However, it is interesting to note that perceived consequences of the most severe outcome do predict refusal to visit legitimate sites and enter information there, perhaps reflecting false alarms. Higher ratings of the severity of having one's social security number stole were related to lower likelihood of visiting the legitimate sites, including the NASA site (ß=-.86, *p*<.05) and the Cognix company site (ß=-.58, *p*<.01), and that these relationships hold even after controlling for other factors such as whether the site is seen as trustworthy, knowledge about the lock images, or knowledge about spyware, etc.

4. FUTURE WORK

This preliminary survey contained only a small number of emails in the role play, and a limited number of correlates. However, given the evidence for good validity of the role play, we are currently conducting an extended survey with more emails, more covariates, and a larger and broader sample. This survey also includes a more sophisticated attempt at spear phishing and more examples of both legitimate and phishing emails. The larger survey will also explore in more depth the costs and benefits of avoiding phishing.

In future work, we plan to gather more direct evidence of the validity of the role play, by correlating performance with data from how people treat legitimate and phishing email messages in the wild. Having such a benchmark would add great credibility to this task as one that can be easily administered through on-line surveys as a proxy for phishing susceptibility. Such a tool would be valuable for evaluating educational efforts in the future.

5. DISCUSSION

5.1 Summary of Findings

The role play exercise appears to be a reliable measure of behavioral response to phishing attacks, and is easy to implement. Better knowledge of web environments predicts lower susceptibility to phishing attacks but not increased false positive responses to legitimate emails. Specifically, understanding of what phishing attacks are and ability to parse URLs accounts for a substantial amount of the variance. This suggests that education about how to interpret cues in browsers may have a role in helping people to avoid phishing attacks.

On the other hand, the ratings of consequences suggest that fear of credit card theft is not a great motivator for protecting one's information. This could be due to protections provided against credit card theft, such that individuals are often not liable for fraudulent purchases made. Furthermore, these perceived consequences are a good predictor of false alarms, but not of identifying or avoiding phishing attacks. Therefore, protections against phishing might not gain much traction from warnings about how easy it would be for a phisher to steal one's card.

5.2 Limitations

There are several limitations to the current study. First, the sample was drawn by convenience and is not expected to be representative of the larger population of email users. All participants were members of the Carnegie Mellon University community, including students, faculty, and staff. Furthermore, they came from two distinct groups within that community: those who self-selected into a group interested in learning about security on-line, and those who were making amends for past violations of university computer usage rules. All of these factors increase the likelihood that these individuals are more computer savvy than average.

A second limitation is in the small set of stimuli used for the role play. In order to keep this survey short enough to be administered as part of this event, we included only one legitimate financial email and one regular phishing email in the set. A further limitation is the failure of our spear phishing email to engage participants. Such a small number of individuals clicked on the link in the spear phishing email (6%) that this behavior could not be reliably correlated with other measures in the survey. Furthermore, the images used in the survey were based on Internet Explorer's interface on a PC, to be applicable to the greatest number of participants. However, participants who were more familiar with other systems may have been at a disadvantage by not having their normal cues available.

A third limitation of this study is the lack of direct consequences for behavior. Participants might be more willing to engage in risky behavior in this role play since they are immune to any negative outcomes that may ensue. Similarly, participants are not risking opportunity costs from being too conservative in their behavior. Given the comments provided, we believe that respondents took this task seriously and acted in ways similar to how they would treat their own email accounts, but without analogous incentives we cannot be confident of this. It has been shown that people behave slightly less cautiously in role-play situations compared to real-world settings [2], although the overall patterns of behavior are very similar. There is no indication that the predictors described here should differ in their relationship to role-played behavior compared to real-world behavior.

5.3 Relationship to Previous Work

A number of relatively small studies have been done previously to gain insights into users' behavioral response to phishing. These studies have provided much needed insights, but were not designed to examine the role of unprompted understanding of cues in behavioral responses to phishing attacks.

Dhamija, et al. [3] conducted a lab study in which they showed 20 web sites to each of 22 participants and asked them to determine which web sites were fraudulent. Although most participants made use of some of the available browser-cues, such as the address bar and the status bar, 23% looked for security indicators only within web site content. The authors concluded that browser security indicators are misunderstood or ignored frequently, and many users have never noticed them. They also concluded that users generally have great difficulty distinguishing legitimate web sites from spoofs [3].

Sheng, et al. [4] conducted a similar lab study in which they showed 42 participants 20 web sites and asked them to determine which were fraudulent. However, in this study participants took a break after reviewing half the sites. During the break, one group of participants played an anti-phishing game, one group read an anti-phishing tutorial, and one group played solitaire and did other unrelated activities. Similar to Dhamija, et al., the authors found that users have difficulty determining which sites are legitimate. However, after less than 15 minutes of training via the anti-phishing game, study participants improved their ability to distinguish legitimate and fraudulent sites considerably. The participants in the tutorial condition also improved, although not as much as those in the game condition [4].

The two studies mentioned above evaluated users' ability to identify fraudulent web sites without providing the context for visiting those sites. Since users often arrive at phishing web sites as a result of clicking on links in phishing email messages, it is useful to evaluate users' response to phishing email messages as well. Kumaraguru, et al. conducted a lab study in which 30 participants, divided into three conditions, were each asked to play the role of an administrative assistant working for a company and process the email messages in that person's inbox. The inbox email was a mix of legitimate and phishing messages, as well as two anti-phishing training messages. The training messages varied across the three conditions. Prior to reading the training messages, most participants clicked on the links in the phishing messages and proceeded to enter personal information into the phishing web sites. Participants' performance after reading the training messages varied by condition, with participants in one condition performing dramatically better following the training [5].

Downs, et al. [6] conducted a lab study in which they interviewed 20 non-expert computer users to understand their decision strategies when handling email. As part of the interview, participants were asked to role play and respond to a set of legitimate and fraudulent email messages in an office assistant's inbox. The authors found that many participants were not very familiar with phishing, or had limited knowledge of what cues to look for to distinguish legitimate messages from fraudulent ones. Even those participants who were somewhat knowledgeable, were not very good at extrapolating what they knew about phishing and applying it to unfamiliar attacks.

All four of these lab studies provide insights into when users will fall for phishing attacks, and two of them also provide insights into the effectiveness of various approaches to anti-phishing training. However, these and other recent phishing-related laboratory studies are not readily generalizable to a larger population and their authors were able to find few, if any, correlation between demographics, personal characteristics, and behaviors relevant to the studies.

Some larger field studies have provided larger data sets, but because they were less controlled than lab studies, only limited demographic and background information was available about participants, and behavioral information was typically collected based on only one phishing email per participant. For example, field studies have been conducted in which phishing messages were sent to students at the University of Indiana [7], West Point cadets [8], and New York State employees [9]. The Indiana study measured the relationship between susceptibility to a phishing attack and certain attributes of the participants, including class and major. However, our present study is the first relatively large-scale study we are aware of to find a correlation between phishing susceptibility and phishing-related knowledge and experience.

Our study helped us to validate and refine a methodology for gathering data about phishing susceptibility using an online survey instrument. While this approach has its limitations and does not offer as realistic an environment as a study in which participants are exposed to phishing messages in their own inboxes, it offers a number of benefits. This approach avoids the many difficulties associated with launching simulated phishing attacks. It also provides a better opportunity to collect data on participant's knowledge and experience, which can be used to inform the development of anti-phishing educational materials.

5.4 Implications for Development of Tools

A variety of tools and techniques are advancing the fight against phishing. However, phishing detection remains an arms race. Commercially available browser-based tools still miss a large fraction of phishing URLs [10]. There is also evidence that users may ignore toolbar warnings, even when they are accurate [11]. Email-based automated detection tools show promise [12], but have limitations as well. Efforts by frequently-phished brands to educate their customers via email appear to have had little success, as users seem to ignore such messages [5]. All of these tools and techniques stand to benefit from knowledge of users' behavioral response to phishing.

Understanding the behavioral response to phishing has the most direct implication for anti-phishing education. If we know what factors cause people to fall for phish we gain insights into what specific things to educate people about and to whom to target education. Of course, we would like to try to develop education that will not only teach people how to avoid known attacks but will also anticipate future ones. Recent work on an embedded training approach to anti-phishing education [5] as well as an anti-phishing game [4] have benefited from the knowledge gained from previous behavioral studies. For example, the interview results suggested that any training about URLs would need to provide some basic understanding about their construction before detailed explanations were offered. The current study indicates that new tools and cues should provide users with a more complete understanding of the underlying concepts that they reflect, rather than merely indicating safety or risk.

Developers of anti-phishing tools for end users can use insights into users' behavioral responses in the design of more effective user interface messages that users will be less likely to ignore. Merely alerting users to a risk that they may not fully understand is likely to lead to uninformed behavior, with users just looking for the right button to click in order to make the warning go away. Helping users to understand the nature of the threat will help them to better deal with it. This distinction is important to consider in development of automated detection systems, which are motivated by the goal of overriding the user's decisions with a smarter system. If such a system were perfect, it could be implemented under the radar. However, given that there will never be full protection against false negatives and false positives, users will always need the power to override the automated system, thus providing an opportunity for attacks. If the user has an intuitive understanding of what the threats are and how the automated system works, they will be far more likely to override it when appropriate, and less likely to be tricked into overriding it by increasingly sophisticated attacks.

6. CONCLUSIONS

These results suggest that future efforts to squelch phishing attacks should consider both sides of the user coin: educating users to interpret cues in the browser environment, and making these cues more intuitively understandable to relatively naïve users. No automated system will ever be foolproof, leaving the user as the perennial weak link in the system. When the user is asked to make judgments about system performance, warning messages would do well to provide them with enough understanding to make sensible judgments rather than merely asking for a simple response of "OK" or "Cancel." Additionally, threats about consequences appear to be more likely to raise the level of false positives without actually protecting people very well against phishing attacks. Further research is needed to determine whether particular education efforts aimed at the concepts identified in this study will result in better ability to detect and respond appropriately to phishing attacks.

7. ACKNOWLEDGMENTS

We gratefully acknowledge support from National Science Foundation grant number 0524189 entitled "Supporting Trust Decisions," and from the Army Research Office grant number DAAD19-02-1-0389 entitled "Perpetually Available and Secure Information Systems." We would also like to thank Alessandro Acquisti, Sven Dietrich, Jason Hong, Norman Sadeh, Serge Egelman, Ian Fette, Ponnurangam Kumaraguru, and Steve Sheng for their helpful insights in development of the survey tool.

8. REFERENCES

1. Schneier, B. 2000. Semantic Attacks: The Third Wave of Network Attacks. *Crypto-Gram Newsletter.* October 15, 2000, http://www.schneier.com/crypto-gram-0010.html

2. Schechter, S. E., Dhamija, R., Ozment, A., Fischer, I., 2007 The Emperor's New Security Indicators. *IEEE Symposium on Security and Privacy*, 20-23 May 2007.

3. Dhamija, R., Tygar, J. D., and Hearst, M. 2006. Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada, April 22 - 27, 2006). R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, Eds. CHI '06. ACM Press, New York, NY, 581-590. DOI= http://doi.acm.org/10.1145/1124772.1124861.

4. 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 Third Symposium on Usable Privacy and Security,* 2007.

5. 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. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA, April 28 - May 03, 2007). CHI '07. ACM Press, New York, NY, 905-914. DOI= http://doi.acm.org/10.1145/1240624.1240760

6. Downs, J. S., Holbrook, M. B., and Cranor, L. F. 2006. Decision strategies and susceptibility to phishing. In *Proceedings of the Second Symposium on Usable Privacy and Security* (Pittsburgh, Pennsylvania, July 12 - 14, 2006). SOUPS '06, vol. 149. ACM Press, New York, NY, 79-90. DOI= http://doi.acm.org/10.1145/1143120.1143131.

7. Jagatic, T., Johnson, N., Jakobsson, M., and Menczer. F. Social Phishing. To appear in *Communications of the ACM,* October, 2007.

8. Ferguson, A. J. 2005. Fostering E-Mail Security Awareness: The West Point Carronade. EDUCASE Quarterly. 2005, 1. Retrieved March 22, 2006, http://www.educause.edu/ir/library/pdf/eqm0517.pdf.

9. New York State Office of Cyber Security & Critical Infrastructure Coordination. 2005. Gone Phishing… A Briefing on the Anti-Phishing Exercise Initiative for New York State Government. Aggregate Exercise Results for public release.

10. Zhang, Y., S. Egelman, L. Cranor, and J. Hong. 2007. Phinding Phish: Evaluating Anti-Phishing Tools. In *Proceedings of the 14th Annual Network and Distributed System Security Symposium (NDSS 2007),* San Diego, CA, 28 February -2 March, 2007.

11. Wu, M., Miller, R. C., and Garfinkel, S. L. 2006. Do security toolbars actually prevent phishing attacks?. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006). R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, Eds. CHI '06. ACM Press, New York, NY, 601-610. DOI= http://doi.acm.org/10.1145/1124772.1124863

12. Fette, I., N. Sadeh and A. Tomasic. Learning to Detect Phishing Emails. June 2006. ISRI Technical report, CMU-ISRI-06-112 (To be presented at WWW 2007). http://reports-archive.adm.cs.cmu.edu/anon/isri2006/CMU-ISRI-06-112.pdf