Protecting People from Phishing: The Design and Evaluation of an Embedded Training Email System

Ponnurangam Kumaraguru, Yong Rhee, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, Elizabeth Nunge

Carnegie Mellon University

ponguru@cs.cmu.edu, ywrhee@cmu.edu, acquisti@andrew.cmu.edu
lorrie@cmu.edu, jasonh@cs.cmu.edu, enunge@andrew.cmu.edu

ABSTRACT

Phishing attacks, in which criminals lure Internet users to websites that impersonate legitimate sites, are occurring with increasing frequency and are causing considerable harm to victims. In this paper we describe the design and evaluation of an embedded training email system that teaches people about phishing during their normal use of email. We conducted lab experiments contrasting the effectiveness of standard security notices about phishing with two embedded training designs we developed. We found that embedded training works better than the current practice of sending security notices. We also derived sound design principles for embedded training systems.

Author Keywords

Embedded training, phishing, email, usable privacy and security, situated learning

ACM Classification Keywords

D.4.6 Security and protection, H.1.2 User / Machine systems, H.5.2 User interfaces

INTRODUCTION

A *semantic attack* is a computer-based attack that exploits human vulnerabilities. Rather than taking advantage of system vulnerabilities, semantic attacks take advantage of the way humans interact with computers or interpret messages [33], exploiting the difference between the system model and the users' mental model [27].

Recently we have seen a dramatic increase in semantic attacks known as "phishing," in which victims get conned by spoofed emails and fraudulent websites. Victims perceive that these emails are associated with a trusted brand, while in reality they are the work of con artists interested in identity theft [22, 27, 30]. These increasingly sophisticated attacks not only spoof email and websites, but can also spoof parts of a user's web browser, for example to hide warnings and URL information [9]. User studies have shown that a large number of people fall for these phishing attacks, even when the participants are made aware that their ability to identify phishing attacks is being tested [7].

Phishing attacks are initiated through several vectors, the most popular of which is currently email [20, 30]. Phishing emails deploy a variety of tactics to trick people into giving up personal information: for instance, urging people to verify their account information, or asking people to take part in a survey in which they must provide their bank account number to be compensated. The increasing sophistication of these attacks makes them hard to distinguish from legitimate emails, and reduces the trust users afford to genuine websites [9].

Previous anti-phishing research has focused either on algorithms for detecting phishing attacks in web browsers [17, 34] or on evaluating the user interfaces of anti-phishing web browser toolbars [38]. However, there has been little work on preventing users from falling for phishing email messages [30].

Our work focuses on teaching people about the risks of phishing and training them to identify and avoid phishing attacks in email. Towards this goal, we are developing an *embedded training* approach that teaches people how to protect themselves from phishing during their regular use of email. Our approach consists of periodically sending users fake phishing emails that are actually from our system rather than from a scammer. If a person falls for our fake email and clicks on a link, we display an intervention that provides immediate feedback about what happened and what simple actionable steps users could take to protect themselves.

In this paper, we describe the design and evaluation of two interventions for our embedded training system, one that provides a warning as well as actionable items using text and graphics, and the other that uses a comic strip format to convey the same information. We also present the results of a user study that compares the effectiveness of typical email security notices sent out by e-commerce companies to alert their customers about phishing to the effectiveness of our two designs. Our evaluation suggests that typical email security notices are ineffective, while our embedded training designs are effective. Based on our results, we outline some design

CHI 2007, April 28–May 3, 2007, San Jose, California, USA.

principles for embedded training email notices that can be implemented easily today.

RELATED WORK

A variety of strategies to protect people from phishing have been proposed in the literature and implemented. These strategies fall into three major categories: silently eliminating the threat, warning users about the threat, and training users not to fall for attacks.

Silently Eliminating the Threat

The strategy of silently eliminating the threat provides protection without requiring any awareness or action on the part of users. This includes finding phishing sites and shutting them down, as well as detecting and deleting phishing emails automatically [17, 34]. If phishing threats could be completely eliminated using these methods, there would be no need for other protection strategies. However, existing tools are unable to detect phishing emails with one hundred percent accuracy, and phishing websites stay online long enough to snare unsuspecting victims. According to the Anti-Phishing Working Group (APWG), phishing sites stay online on average for 4.8 days [3].

Warning Users

A number of tools have been developed to warn users that the website they are visiting is likely fraudulent, either by providing explicit warnings or by providing interfaces that help people notice that they may be on a phishing website. Ye and Sean [39] and Dhamija and Tygar [8] have developed prototype "trusted paths" for the Mozilla web browser that are designed to assist users in verifying that their browser has made a secure connection to a trusted site. More common are web browser toolbars that provide extra cues—such as a red or green light indicating overall safety—to inform users that they may be at risk [12, 28, 35, 36]. However, there are three weaknesses with this approach. First, it requires people to install special software (although newer versions of web browsers have such software included). Second, user studies have shown that users often do not understand or act on the cues provided by toolbars [27, 38]. Third, a recent study shows that some anti-phishing toolbars are not very accurate, and even the best toolbars may miss over 20% of phishing websites [40].

Training Users

As the phishing threat currently cannot be eliminated entirely through automated tools or law enforcement action and users fail to heed toolbar warnings, we believe it is necessary to train users about phishing attacks and how to avoid them. The training strategy is complementary to the first two strategies and should be pursued in parallel with them.

There are many approaches to training and educating users about phishing. The most basic approach is to post articles about phishing on websites, as has been done by government organizations [14, 15], non-profits [4] and businesses [11, 26]. A more interactive approach is to provide web-based tests that allow users assess their own knowledge of phishing. For example, Mail Frontier [23] has set up a website containing screenshots of potential phishing emails. Users are scored based on how well they can identify which emails are legitimate and which are not. Phishing education can also be conducted in a classroom setting, as has been done by Robila and Ragucci [31].

The idea of sending fake phishing emails to test users' vulnerability has been explored by several groups. Typically, at the end of such studies, all users are given additional materials to teach them about phishing attacks. This approach has been used with Indiana University students [18] and West Point cadets [16], as well as with employees at a New York state office [29]. Both the West Point and the New York state researchers conducted the study in two phases. In the first phase, participants did not have any prior preparation or training about phishing before being tested for their ability to detect phishing attacks. In the second phase, participants were given training materials and lectures about phishing before being tested again. Both studies showed an improvement in the participants' ability to identify phishing emails.[1]

Our work differs in that we are focused on the design and evaluation of email interventions to understand what kinds of designs are more effective in teaching people about phishing and actually protecting them in practice. For example, our studies suggest that the standard practice of sending out security notices is not an effective intervention. Furthermore, our work evaluates how well people can generalize what we teach them to other kinds of related attacks. The previous studies either tested participants only once [18] or tested participants on a single kind of attack on their intranet [16, 29]. Our work aims to teach people what cues to look for to make better decisions in more general cases. For example, rather than just teaching people not to fall for PayPal phishing attacks, we want people to learn how to identify phishing attacks in general.

DESIGN OF TRAINING EMAILS

In this section we describe our rationale for email intervention, the evolution of the design of our embedded training system, the results of an early version of that design, some design goals we derived from evaluating the early design and from related work, and the design of our current interventions.

[1] Although questions have been raised about the ethics of such deceptive approaches to educating users and studying the effectiveness of phishing attacks, the general consensus among the phishing research community seems to be that such studies are ethical when conducted with the approval of the appropriate institutional review boards [19]. This issue has been discussed at research conferences, for example, at a SOUPS 2005 panel "When User Studies Attack: Evaluating Security By Intentionally Attacking Users."

Figure 1: Early iteration of an intervention. When people clicked on a link in one of our training emails, it would bring up a screenshot (above) of the web browser showing that the URL they thought they clicked on was not the same as the URL that the web browser would go to (web browser status bar).

Our embedded training system works roughly as follows. People are periodically sent training emails, perhaps from their system administrator or from a training company. These training emails look just like phishing emails, urging people to go to some website and log in. If people fall for the training email and click on a link in that email, we provide an intervention that explains that they are at risk for phishing attacks and gives some tips for protecting themselves.

Rationale for Email Intervention

There are two primary intervention points for an anti-phishing training system: email and web. We chose to focus on email for three reasons. First, email is the main vector for delivering phishing messages to users. If we can prevent people from trusting phishing emails, it is likely they will not reach the vast majority of phishing websites. Second, anti-phishing websites [11, 15] require end-users to proactively visit them, limiting the number of people who will actually see these websites. In contrast, our approach brings information to end users and teaches them over time to differentiate between legitimate and illegitimate emails. Third, end users must already have some knowledge about phishing or other kinds of scams to seek out educational websites. In contrast, our approach (if distributed with standard email clients or sent by companies) works for experts as well as novices who are unaware of phishing, by educating end-users immediately after they have made a mistake.

We are also developing a game-based approach to training people to identify phishing websites. The email-based approach presented in this paper and the game-based training were designed to be complementary. Our email approach is designed to provide training in the course of normal email usage. If users are interested in learning more, they can then play our game to gain a more thorough understanding of phishing attacks and ways of identifying phishing websites.

Early Designs

We started our design with paper prototypes and refined our ideas using HTML prototypes. The version used in our user studies is implemented with HTML and JavaScript.

To gain insight into the design space we created and evaluated several prototypes of our embedded training system. One early design consideration was whether to show interventions immediately after a person had clicked on a training email or after they had tried to log into the website. Our paper prototypes strongly suggested that showing an intervention after a person had clicked on a link was better, since people who were shown interventions after logging in were confused as to why they were seeing warning messages about the risks of clicking on email links. We believe this is due to a gap between cause (clicking on a link) and effect (seeing a warning message about email after logging in).

To get a better feel for how well our ideas would work in practice, we created an HTML mockup in Squirrel Mail [37], a web-based email service. People who used our system encountered our training emails interspersed with regular email messages. If they clicked on a link in one of our training emails, they were taken to a separate web page and shown one of two interventions. The first intervention (see Figure 1) showed a screenshot of the email within the web browser itself, pointing out that the link they clicked on was not the same as the link they would actually go to as shown in the status bar. The second intervention was similar, but told people more directly that the link they clicked on did not take them to the website they intended by showing the brand name itself (in this case, "This is not eBay"). Both interventions also provided text at the top of the image describing why the participants were seeing such a page and informing them that they were at risk of falling for phishing attacks.

We did a pilot evaluation of our design with ten participants, using a variation of the protocol developed by Downs et al [10]. We asked our participants to role play as an employee at a company and to handle the email in the employee's mailbox the way they normally would. The employee's mailbox contained nineteen email messages, including a few phishing emails and two training emails.

Nine out of ten participants clicked on our first training message (essentially falling for our fake phishing email) and saw the information that we presented about phishing. However, almost all the users who viewed the training message were confused about what was happening. They did not understand why they were sent this email.

Furthermore, most of the participants who viewed the training message did not understand what it was trying to convey. A common response to the first intervention (Figure 1) was, "I don't know what it is trying to tell me." Some users understood the training message but were uncertain how to respond as the message did not suggest any specific actions to take. In debriefing sessions, participants reported that the second intervention was more useful than the first,

since they could understand that the website they were visiting was not part of eBay.

Another problem was that people were sometimes confused by the screenshot of the web browser. Many participants failed to notice the text at the top describing why they were seeing the warning, mostly because the browser screenshot was so large and visually dominating. A third problem was that people had to scroll to see the entire warning.

Nine users fell for our first phishing email (before any interventions), and seven users fell for the final phishing email (after both interventions), suggesting that this early design was not effective. Nearly all of the participants that clicked on a phishing link actually tried logging in, suggesting again that it would be better to intervene immediately after a person clicks on a link (since they are likely to fall for the phishing website) rather than after they try to log in.

In summary, the lessons from our early prototypes were:

- It is best to show interventions immediately after a person clicks on a link in a training email
- People expect to go to a website when they click on a link, so interventions need to make it extremely clear why they are not being taken to that website
- Interventions need to provide clear actionable items rather than general warnings about potential risks
- Text and images need to be simple and visually salient to convey the warning accurately and avoid confusion

Current Designs

Informed by our early designs, we created two new interventions: a text and graphics intervention and a comic strip intervention. The text and graphics intervention, shown in Figure 2, describes the risks of phishing, shows a small screenshot of the training email, points out cues that it is a phishing email, and outlines simple actions that users can take to protect themselves. The comic strip intervention, shown in Figure 3, conveys roughly the same information as the text and graphics intervention, but in a comic strip format. Our rationale here was that the first intervention had a great deal of text, which might cause people to just close the window without reading it. Comic strip stories are a highly approachable medium [6], so we decided to test the effectiveness of a comic strip approach to anti-phishing training.

To develop these two interventions we analyzed 25 online anti-phishing tutorials and selected guidelines that were frequently mentioned, simple enough for people to do, and effective. For example, some tutorials suggest using networking tools to analyze the age and owner of the domain. While effective, this is not an easy strategy for the large majority of people. The four suggestions we decided to teach people were:

- Never click on links in emails
- Initiate contact (i.e. manually type in URLs into the web browser)
- Call customer service
- Never give out personal information

The rationale for "Never click on links in emails" is that it is difficult for non-experts to determine whether links lead to legitimate web sites. Rather than attempting to teach people a complicated set of rules for differentiating between safe and unsafe links, we opted to teach them a simple rule, expecting that users would eventually work out their own adaptation of the rule.

The rationale for "Initiate contact" is that it is much safer for people to type in a web address into a web browser on their own or to use a bookmark, rather than trusting a link in an email.

For "Call customer service," the rationale is that many phishing attacks rely on scaring people into logging in to an account. Calling customer service is a fairly reliable way of determining if there really are any problems with one's account (assuming the phone number is obtained from a reliable source). We also believe that increasing the number of customer service calls will provide an incentive to companies to take stronger action against phishing, since such calls cost companies money. Although this seems like an extreme measure, it is also worth noting that no person in our studies actually called customer service. We argue that this is still a useful piece of advice given that it reminds people that there are offline ways to contact companies.

For "Never give out personal information", the rationale is that companies rarely ask for such information, and the large majority of such requests are phishing attacks.

However, learning science suggests that simply telling people to follow advice is insufficient. The literature indicates that it is better to present abstract information using concrete examples [1, 2, 5, 32]. In the text and graphics intervention, we chose to tie our advice to the email that led participants to the warning, by showing a small screenshot of that email and by showing a small screenshot of the web browser address bar. In the comic strip intervention, we chose to tie our advice to a short story explaining how scammers work and how the reader could do simple things to avoid phishing attacks.

Learning science also suggests that situated learning [2, 6, 13, 24, 25], where instructions are provided while people are solving a problem, is an effective teaching strategy. In the text and graphics intervention, we do this by showing all of the cues a person should look for on the left side of the warning and tie it immediately to simple steps that people can do to protect themselves. In the comic strip intervention, we take an alternative approach by situating people in a comic strip story that explains how scammers send phishing emails, how the reader can identify phishing cues, and what they can do if they suspect an email might be fraudulent. We decided to show the interventions immediately after a person clicks on a link in a training email. However, rather than taking people to a separate web page, we gray out our

training email and display a floating window on top. Our goal is to reduce confusion and let people know that they are still in the same place. Showing a floating window also brings the intervention closer to the center of the web browser content area, making it harder to miss important content. Both interventions include prominent titles and a cartoon image of a thief to help convey that participants are potentially at risk. We designed the interventions to be read without requiring any scrolling or clicking on additional links within the interventions. To view the latest designs please visit http://cups.cs.cmu.edu/trust/et_design.php.

EVALUATION

In this section, we present the design and results of a user study evaluating the effectiveness of our interventions compared to the current practice of sending out security notices. We conducted a laboratory study using three conditions, each of which had a different intervention. There were 10 participants in each condition for a total of 30 participants.

Participants

As this research is focused on educating novice users about phishing attacks, we recruited participants with little technical knowledge. We posted fliers around our university and local neighborhoods, and then screened users through an online survey. We recruited users who said they had done no more than one of the following: changed preferences or

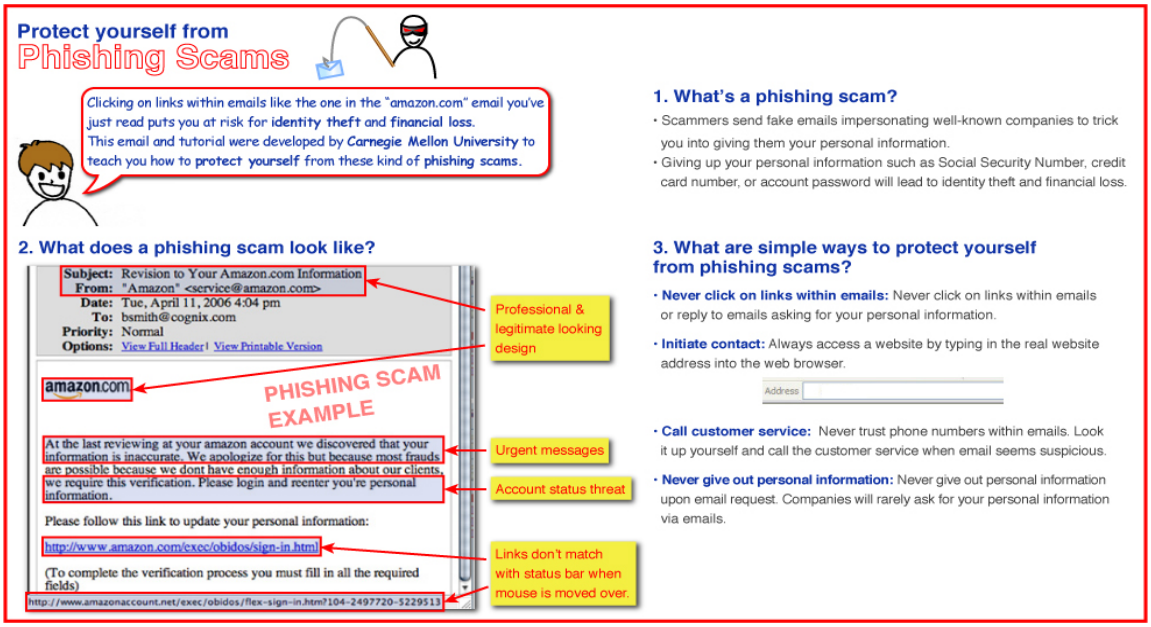

Figure 2. The text and graphics intervention includes text with an annotated image of the training email that led to this warning.

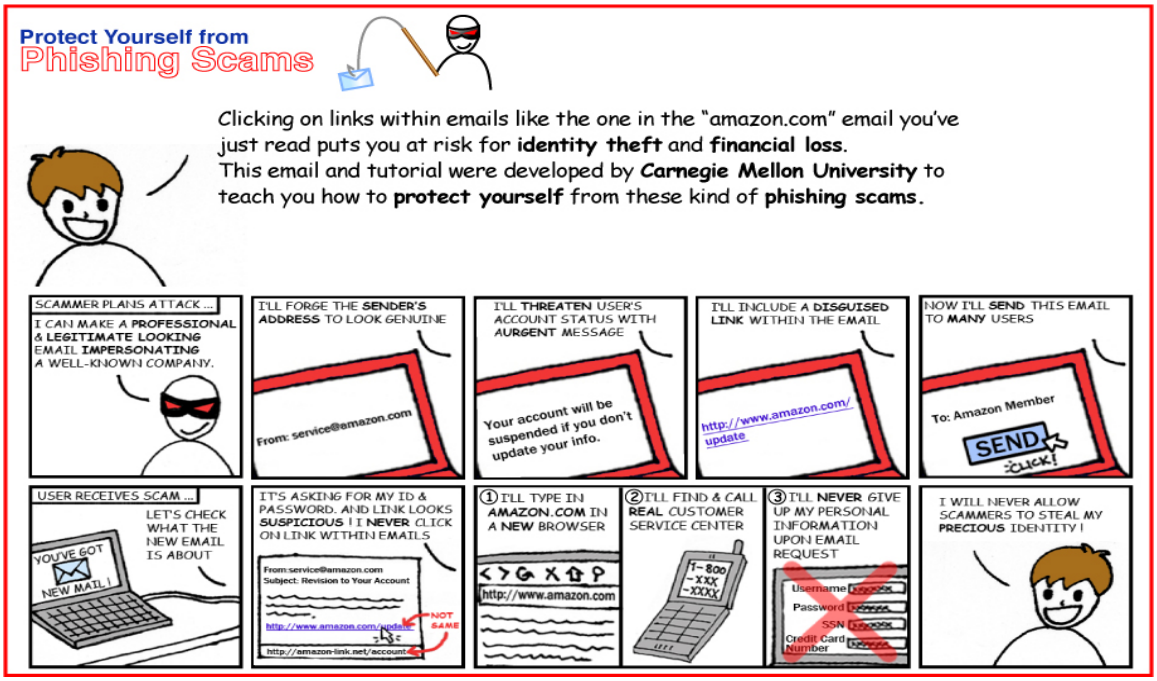

Figure 3. Comic strip intervention uses a comic strip to tell a story about how phishing works and how people can protect themselves.

settings in their web browser, created a web page, and helped someone fix a computer problem. This approach has served as a good filter to recruit non-experts in other studies [10, 21].

Each participant was randomly placed in one of three groups. The "notices" group was shown typical security notices, the "text/graphics" group was shown the text and graphics intervention displayed in Figure 2. The "comic" group was shown the comic strip intervention displayed in Figure 3. Table 1 shows the demographics of our participants.[2]

	Notices Group	Text/Graphics Group	Comic Group
Gender			
Male	50%	40%	20%
Female	50%	60%	80%
Computer			
PC	100%	100%	70%
Mac	0%	0%	30%
Browser			
IE	80%	60%	60%
Others	20%	40%	40%
Average emails per day	51.4	36.9	15
Average Age	31.2	27.5	21.1

Table 1: Demographics of the participants

Methodology

We used a 1.40GHz Compaq laptop running Microsoft Windows XP home edition to conduct the user studies. The participants used Internet Explorer 6.0 for accessing emails through Squirrel mail [37].

The user study consisted of a think-aloud session in which participants played the role of "Bobby Smith," an employee of Cognix Inc. who works in the marketing department. Participants were told that the study investigated "how people effectively manage and use emails." They were told that they should interact with their email the way they would normally do in their real life. If a participant was not familiar with Squirrel mail, we gave that participant a quick tutorial describing how to perform simple actions. We also mentioned that we would be able to answer questions about using Squirrel mail during the study, but we would not be able to help them make any decisions. We asked participants a few pre-study questions about their use of email to reinforce the idea that this was a study about use of email systems. We recorded the audio and screen interactions using Camtasia.

[2] One outlier in the notices group received 300 emails daily, but did not perform particularly better or worse than others in this group. We found no significant relationship between propensity to fall for phishing attacks before the intervention and any of the demographic information we collected. Other studies have also found no correlation between these demographics and susceptibility to phishing [7,10].

We gave participants an information sheet describing the scenario and asked them to read it aloud and ask clarification questions. The information sheet included the usernames and passwords for Bobby Smith's email account and accounts at Amazon, American Express, Citibank, eBay and PayPal. We also provided username and password information in a physical wallet that participants could use throughout the study.

Each participant was shown 19 email messages, arranged in a predefined order. Nine messages were legitimate email messages that Bobby Smith received from co-workers at Cognix, friends and family. These emails expected Bobby Smith to perform simple tasks such as replying. Two messages were simulated legitimate emails from organizations with which Bobby Smith had an account. The mailbox also contained two spam emails, four phishing emails, and two training emails (security notices or embedded training interventions). Table 2 shows the email distribution shown to the users. Of the four phishing emails only two of the emails were from organizations where Bobby Smith had an account. One of these phishing emails was placed before the first training email and the other was placed after the second training email.

1. Legitimate	6. Legitimate	11. Intervention	16. Phishing
2. Legitimate	7. Legitimate	12. Spam	17. Phishing
3. Phishing	8. Spam	13. Legitimate	18. Legitimate
4. Legitimate	9. Legitimate	14. Phishing	19. Legitimate
5. Intervention	10. Legitimate	15. Legitimate	

Table 2: Email arrangement in the study.

All the phishing, spam, and security notice emails that we used for this study were based on actual emails we had collected. We created exact replicas of the phishing websites on our local machine by running Apache and modifying the host files in Windows so that IE would display the URL of the actual phishing websites. All replicated phishing websites were completely functional and allowed people to submit information.

We used a completely functional Squirrel mail implementation for users to access Bobby Smith's email. We wrote a Perl script to push emails into the Squirrel mail server; and used this script to change the training emails for each group.

After participants finished going through Bobby Smith's emails, we asked them some post-study questions and we debriefed them. During the debriefing we asked them questions about their choices during the study. We also showed training messages belonging to a different group than the one they had been placed in for the study. For example, participants who viewed Figure 2 in their study were shown Figure 3 after the study and vice versa. They were then asked about their views of both designs.

RESULTS

In this section we present the results of our user study. In this paper we consider someone to have fallen for a phishing attack if they click on a link in a phishing email, regardless of

whether they go on to provide personal information. Although not everyone who clicks on a phishing link will go on to provide personal information to a website, in our study people who clicked on phishing links provided information 93% of the time. In addition, clicking on phishing links can be dangerous even if someone does not actually provide personal information to the site because some phishing sites can transmit malware to a user's computer.

Security Notices Intervention

There was no difference between the number of participants clicking on links in phishing emails before and after the two security notice messages. The first security notice users saw was a security message that eBay/PayPal sends to customers. The email was linked to a real website [11]. Only five (50%) users in this group clicked on the first security notice link in the email to learn more about phishing attacks. Among these five participants only two (40%) actually read through the content in the web pages, whereas the other three (60%) skimmed through the content and closed the window. Nine (90%) participants clicked on the second security notice; this security notice was sent from the system administrator of Cognix. During the post-study debriefing we asked whether the notices had been helpful. The participants who had seen the security notices said the information took too long to read and they were not sure what the messages were trying to convey. Nine participants (90%) fell for the phishing email before the security notice email and nine participants (90%) fell for the final phishing email. The mean percentage of participants falling for the three phishing emails presented after the security notices was 63%.

Text and Graphics Intervention

In this group eight participants (80%) fell for the first phishing email while all participants clicked on the training message link in the training email. Seven participants (70%) clicked on the second training message and seven participants (70%) fell for the final phishing email. The mean percentage of participants falling for the three phishing emails presented after the interventions was 30%. Many participants checked for whether they had an account with the financial institution before clicking on the link after going through the training message. Only one user (10%) clicked on the phishing message that was sent from Barclays Bank which they did not have an account with. When asked why he had done so, the user said, "just because it [the link] was there and I wanted to check what they show." Most participants liked the way the information was presented; a common comment was: "Having the image and the text with callouts was helpful." One user told us: "Giving the steps to follow to protect from phishing was helpful." Another said, "This is definitely useful and good stuff and will remember that [to look for URLs in the status bar]."

Comic Strip Intervention

Our results indicate that our comic strip intervention was the most effective in educating people about phishing attacks. All the participants in this group fell for the first phishing email and also clicked on the training message. Six participants (60%) clicked on the second training message and only three participants (30%) fell for the final phishing email. The mean percentage of participants falling for the three phishing emails presented after the interventions was 23%. Some participants said they preferred the comic to the text/graphics intervention because it engaged them with a story. However, other participants felt that the text/graphics version was more serious and professional. One user said, "The comic version is good for children but I would prefer text with the image."

Comparison

We can see a significant difference in the ability to recognize phishing emails between the notices group and the comic group. In the notices group nine participants (90%) fell for the final phishing email whereas in the comic group only 3 participants (30%) fell for this email (Chi-Sq = 23.062, DF = 1, P-Value = 0.001).

We also compared the effectiveness of security notices against the effectiveness of the text and graphic intervention. The number of participants falling for phishing attacks before and after training in the notices group was nine (90%), while the number of participants falling for phishing attacks in the text/graphics group was eight (80%) before training and seven (70%) after training. The difference between these two groups was not as significant (Chi-Sq = 0.364, DF = 1, P-Value = 0.546) as the difference between the notices and comic groups.

There was significant difference in effectiveness of the two embedded training interventions (Chi-Sq = 16.880, DF = 1, P-Value = 0.001). The mean scores across the three phishing emails after intervention was lowest for the comic group. Figure 4 presents a comparison of the three training methodologies for all the emails that had links in them.

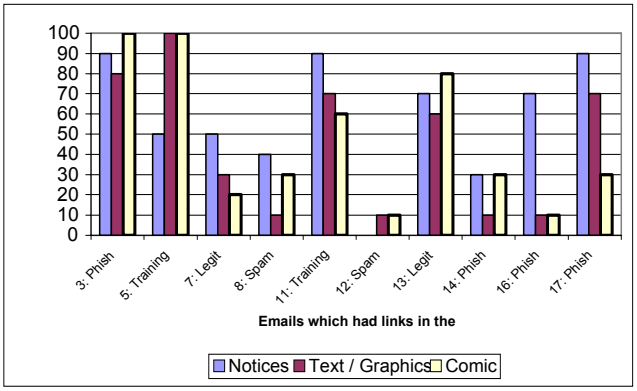

Figure 4: Comparison of different methods of training for each group for all the emails which had link in them. The number represents the location of the email in the email arrangement. Participants in the Comic strip group were able to identify phishing emails better than other two groups.

In our post-study questions we asked participants in the comic and text/graphics groups: "Which one [design] would you prefer and why would you prefer it?" Nine (45%) of the twenty participants preferred the comic version of the

information representation and eleven (55%) preferred the text with graphics version.

During the post-study session, we asked specific questions about the training methodology and about the awareness these methods raised about phishing. One of the questions was: "Did the method create awareness about phishing attacks?" Only two (20%) participants said the security notices method created awareness about phishing attacks, while in both the other groups all participants (100%) said the method created awareness about phishing attacks. We also asked participants: "Do you think this method will help you learn techniques to identify false websites and email?" None of the participants said the security notices would help them, while all of the participants in the other groups thought the embedded training messages would help them.

We also compared data for the individual performance of the participants before and after training. We observed that 9 out of 10 participants (90%) in the notices group clicked the first phishing email and out of these 8 participants (89%) clicked on the final phishing email. In the text/graphics group, 8 participants (80%) clicked on the first phishing email out of which 5 (63%) clicked on the final phishing email. In the comic group, 10 participants (100%) clicked on the first phishing email out of which 3 participants (30%) clicked on the final phishing email. We found that individual performance of participants is significantly different between the notices group and comic group (Chi-Sq = 18.245, DF = 1, P-Value = 0.001). Also there was significant difference between the performance of participants in the text/graphics group and the comic group (Chi-Sq = 7.222, DF = 1, P-Value = 0.007). There was no significant difference between the performance of participants in the notices group and the text/graphics group.

During the post-study session we also asked the participants: "On a scale of 1 to 7, where 1 is not at all confident and 7 is most confident, how confident were you while making decisions on clicking links and replying to emails?" In the notices group the values ranged from 4 to 7 (mean = 5.4, s.d. = 1.1, variance = 1.2), in the text/graphics group values ranged from 3 to 6 (mean = 4.6, s.d. = 0.9, variance = 0.8) and in the comic group values ranged from 3 to 7 (mean = 5.5., s.d. = 1.3, variance = 1.6). Participants in the three groups had similar levels of confidence in handling emails.

General observations

Participants seem to identify the Nigerian scam email (email number 12) easily. Only two of the thirty participants (6.7%) clicked on the link in this email. Only nine participants (30%) actually clicked on the link in the second phishing email, which was ostensibly from a company they did not have an account with. Among these nine participants, four (44.4%) realized that they did not have an account with the service once they clicked on the link, and so they closed the window immediately.

Twenty-four (80%) of all the participants were not familiar with the mouse-over technique to see the actual URL before clicking on the link. Most participants appreciated being taught a technique for identifying the actual link in the emails. One user said, "I did not know to look for links before [in email], I will do it now."

One user in the text/graphics group did not click on any links in the emails because of his personal experience where he had been a victim of identity theft. This user stated, "I was a victim of online credit card fraud, so from then on I decided not to click on links in the emails." No user in the study actually entered random information to test the phishing site's reaction. Two participants used search engines to help their decision about how to react to an email. One user Googled the phrase "Bank of Africa" from the Nigerian scam. Another user said, "I will ask one of my friends to help me make a decision here, she knows about these things better than me." We plan to further investigate the idea of training users to seek help from external and reliable sources to help them make better decisions.

Among the participants who did not understand the training messages we saw similar behavior as discussed by Dhamija et al. [7]. Novice users use misleading signals [21] to make their decisions. For example, one of the participants used the privacy report icon on the phishing website that we created to decide that the website was legitimate. When asked why he did that, he said: "I do that often to find whether the website is legitimate." Another participant mentioned that "the logo [Citibank] is real so the site must be legitimate." Another participant said, "I visited this website [PayPal] some days back. It looks the same as before, so it must be legitimate." A few other participants were satisfied that the website must be legitimate because it showed updated account information after they entered their personal information.

The repetitive training in a short time span was helpful for some participants. Some participants did not understand what was going on the first time the training information was presented, but read it carefully the second time.

DISCUSSION

As observed in other studies, we saw that novice users use misleading signals to make decisions. We believe that properly designed training messages and interventions can help novice users to detect and use meaningful signals.

Our results strongly suggest that security notices are not very effective in teaching people about phishing attacks. We believe this is because people are unclear as to why they are receiving such emails, and because it is difficult for them to relate to an abstract problem that they may not believe is likely to occur. In addition, some participants claimed that they knew about phishing and knew how to protect themselves, but fell for the phishing scams regardless. This also suggests that people may be overconfident about what they know, especially if they have seen such security notices in the past, and thus disregard them.

Our results also indicate that our comic strip intervention was most effective. The primary differences between our two

interventions is that the comic strip format has significantly less text and more graphics, and tells a story to convey its message. We believe that it is worth investigating further to tease out which of these factors are most important, and if other media—such as a short video of a story—might be even more effective.

Based on the results of our low-fidelity prototypes and user studies with our embedded training system, we present some design principles that can be applied to the design of training messages and anti-phishing interventions.

- Embed the training into users' regular activities so they do not have to go to a separate website to learn about phishing attacks.
- Make it clear why users are being warned—for example, what the risks are and what caused the warning.
- Do not delay the warnings; present them immediately after the user clicks on the link.
- Use training messages with the same content that users have just seen, as this helps them concretely relate to what is being discussed in the training message.
- Supplement training text with story-based graphics and annotations.
- Keep the training messages simple and short. One reason the security notices did not work well was too much text.
- Give clear actionable items that participants can easily do to protect themselves.

CONCLUSIONS AND FUTURE WORK

In this paper we have presented the design and evaluation of embedded training methods that teach people about phishing during their normal use of email. From a series of low-fidelity prototypes we drew design criteria that guided the designs of two interventions (see Figures 2 and 3). We conducted lab experiments contrasting the effectiveness of standard security notices about phishing with these two interventions.

Our results suggest that the current practice of sending out security notices is ineffective. Our results also indicate that both of our embedded training interventions helped teach people about phishing and to avoid phishing attacks, and that our comic strip format was the most effective intervention. Based on the results of our early prototypes and user studies, we also presented some design principles for teaching people about phishing. Our results can be put into immediate practice, as they can be implemented easily using current technologies.

We are currently designing a more interactive training system that can adapt to the skill level of participants. We also plan to deploy and evaluate our system with a wider audience.

ACKNOWLEDGEMENTS

This work was supported in part by National Science Foundation under grant CCF-0524189, and by the Army Research Office grant number DAAD19-02-1-0389. 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. The authors would like to thank all members of the Supporting Trust Decisions project for their feedback.

REFERENCES

1. Anderson, J. R., A. T. Corbett, K. Koedinger and R. Pelletier. 1995. Cognitive tutors: Lessons learned. *The Journal of Learning Sciences*, 4, pp. 167-207.
2. Anderson, J. R., M. R. Lynne and Herbert A. Simon. 1996. Situated Learning and Education. *Educational Researcher.* Vo. 25, No. 4, pp. 5 – 11.
3. Anti-Phishing Working Group. Phishing Activity Trends Report. 2006. http://www.antiphishing.org/reports/apwg_report_jan_2006.pdf.
4. Anti-Phishing Working group. http://www.antiphishing.org/. Retrieved on Sept 20, 2006.
5. Betrancourt, M. and A. Bisseret. 1998. Integrating textual and pictorial information via pop-up windows: an experimental study. *Behaviour and Information Technology*. Volume 17, Number 5, pp. 263-273(11).
6. Clark, R. C. and E. M. Richard. 2002. *E-Learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning.* Pfeiffer, San Francisco, USA.
7. 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.
8. Dhamija, R. and Tygar, J. D. 2005. The battle against phishing: Dynamic Security Skins. In *Proceedings of the 2005 Symposium on Usable Privacy and Security* (Pittsburgh, Pennsylvania, July 06 - 08, 2005). SOUPS '05, vol. 93. ACM Press, New York, NY, 77-88. DOI= http://doi.acm.org/10.1145/1073001.1073009.
9. Drake, C. E., J. J. Oliver and E. J. Koontz. MailFrontier. Anatomy of a Phishing Email. Retrieved Feb 27, 2006, http://www.mailfrontier.com/docs/MF_Phish_Anatomy.pdf.
10. 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.
11. eBay. Spoof Email Tutorial. Retrieved December 30, 2006. http://pages.ebay.com/education/spooftutorial/
12. eBay Toolbar. Retrieved December 30, 2006. http://pages.ebay.com/ebay_toolbar/
13. Erhel, S. and E. Jamet. 2006. Using pop-up windows to improve multimedia learning. *Journal of Computer Assisted Learning*, Volume 22, Number 2. pp. 137 - 147.
14. Federal Trade Commission. An E-Card for You game. Retrieved December 30, 2006. http://www.ftc.gov/bcp/conline/ecards/phishing/index.html.

15. Federal Trade Commission. Phishing Alerts. Retrieved December 30, 2006. http://www.ftc.gov/bcp/conline/pubs/alerts/phishingalrt.htm
16. Ferguson, A. J. 2005. Fostering E-Mail Security Awareness: The West Point Carronade. *EDUCASE Quarterly*. http://www.educause.edu/ir/library/pdf/eqm0517.pdf.
17. Fette, I., N. Sadeh and A. Tomasic. Learning to Detect Phishing Emails. June 2006. ISRI Technical report, CMU-ISRI-06-112. http://reports-archive.adm.cs.cmu.edu/anon/isri2006/CMU-ISRI-06-112.pdf.
18. Jagatic, T.,N. Johnson, M. Jakobsson and F. Menczer. Social Phishing. To appear in the *Communications of the ACM*. Retrieved March 7, 2006, http://www.indiana.edu/~phishing/social-network-experiment/phishing-preprint.pdf.
19. Jakobsson, M. and Ratkiewicz, J. 2006. Designing ethical phishing experiments: a study of (ROT13) rOnl query features. In *Proceedings of the 15th international Conference on World Wide Web (*Edinburgh, Scotland, May 23 - 26, 2006). WWW '06. ACM Press, New York, NY, 513-522. DOI= http://doi.acm.org/10.1145/1135777.1135853
20. James, L. 2005. *Phishing Exposed*. Syngress, Canada.
21. Kumaraguru, P., A. Acquisti and L. Cranor. 2006. Trust modeling for online transactions: A phishing scenario. *Proceedings of Privacy Security Trust*, Oct 30 - Nov 1, 2006, Ontario, Canada
22. Lininger, R. and R. Dean. 2005. *Phishing: Cutting the Identity Theft Line.* Wiley, publishing Inc. Indianapolis, Indiana, USA.
23. Mail Frontier. Phishing IQ. http://survey.mailfrontier.com/survey/quiztest.html. Retrieved Sept 20, 2006
24. Mayer, R.E. *Multimedia Learning.* 2001. New York Cambridge University Press.
25. Mayer, R.E. and R. B. Anderson. 1991 Animations Need Narrations: An Experimental Test of a Dual Coding Hypothesis. *Journal of Educational Psychology*. Volume 83, Number 4. pp. 484 – 490.
26. Microsoft. Consumer Awareness Page on Phishing. Retrieved September 10, 2006. http://www.microsoft.com/athome/security/email/phishing.mspx .
27. Miller, R. C. and M. Wu. 2005. Fighting Phishing at the User Interface, In Lorrie Cranor and Simson Garfinkel (Eds.) *Security and Usability: Designing Secure Systems that People Can Use.* O'Reilly.
28. Netcraft. Retrieved September 10, 2006. http://news.netcraft.com/
29. 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.
30. Richmond, R. Hackers set up attacks on home PCs, financial firms: study. Retrieved September 25, 2006. http://www.marketwatch.com/News/Story/Story.aspx?dist=newsfinder&siteid=google&guid=%7B92615073-95B6-452E-A3B9-569BEACF91E8%7D&keyword=
31. Robila, S. A., J. James and W. Ragucci. 2006. Don't be a phish: steps in user education. ITICSE '06: *Proceedings of the 11th annual SIGCSE conference on Innovation and technology in computer science education*. pp 237-241. New York, NY, USA.
32. Schmeck, R. R. (Ed) 1988. Learning styles and strategies. New York: Plenum Press.
33. Schneier, B. 2000. Semantic Attacks: The Third Wave of Network Attacks. Crypto-Gram Newsletter. Retrieved Sep 2, 2006, http://www.schneier.com/crypto-gram-0010.html#1.
34. SpamAssasin. Retrieved September 10, 2006. http://spamassassin.apache.org/
35. SpoofGuard. Retrieved September 10, 2006, http://crypto.stanford.edu/SpoofGuard/
36. SpoofStick. Retrieved September 10, 2006. http://www.spoofstick.com/
37. SquirrelMail. Retrieved September 10, 2006. http://www.squirrelmail.org/
38. 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.
39. Ye, Z. and Sean S. Trusted Paths for Browsers. 2002. Proceedings of the 11th USENIX Security Symposium. pp. 263 - 279. USENIX Association. Berkeley, CA, USA.
40. 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.