# Teaching Johnny Not to Fall for Phish

Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti,
Lorrie Faith Cranor, Jason Hong

February 8, 2007
CMU-CyLab-07-003

CyLab
Carnegie Mellon University
Pittsburgh, PA 15213

# Teaching Johnny Not to Fall for Phish

Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti,
Lorrie Faith Cranor, Jason Hong
Carnegie Mellon University
ponguru@cs.cmu.edu, shengx@cmu.edu, acquisti@andrew.cmu.edu
lorrie@cmu.edu, jasonh@cs.cmu.edu

## ABSTRACT

Phishing attacks exploit users' inability to distinguish legitimate websites from fake ones. Strategies for combating phishing include: prevention and detection of phishing scams, tools to help users identify phishing web sites, and training users not to fall for phish. While a great deal of effort has been devoted to the first two approaches, little research has been done in the area of training users. Some research even suggests that users cannot be educated. However, previous studies have not evaluated the quality of the training materials used in their user studies or considered ways of designing more effective training materials. In this paper we present the results of a user study we conducted to test the effectiveness of existing online training materials that teach people how to protect themselves from phishing attacks. We found that these training materials are surprisingly effective when users actually read them. We then analyze the training materials using principles from learning sciences, and provide some suggestions on how to improve training materials based on those principles.

## 1. INTRODUCTION

Phishing attacks exploit users' inability to distinguish legitimate company websites from fake ones. Phishers send out spoofed emails that look as if they were sent by trusted companies. These emails lead to spoofed websites that are similar or virtually identical to legitimate websites, and lure people into disclosing sensitive information. Phishers use that information for criminal purposes, such as identity theft [26], [29].

People are vulnerable to phishing attacks because spoofed websites look very similar to legitimate websites. Dhamija et al. showed that people have trouble identifying phishing sites even in tests in which they have been alerted about the possibility of such attacks [10]. Furthermore, when phishers personalize their emails, they can further increase the likelihood that the attack will be successful [20], [25].

Researchers have developed several technical approaches to countering phishing attacks, including toolbars, email filters, and verified sender addresses [15]. However, these approaches are not foolproof. In a recent study of 10 anti-phishing tools, only one tool was able to correctly identify over 90% of phishing websites, and that tool also incorrectly identified 42% of legitimate websites as fraudulent [49]. Furthermore, while automated phishing detection is improving, phishers are adapting their attack techniques to improve their chances of success. Finally, contextual information known to the recipient may be needed to determine whether some email messages are legitimate. For example, the recipient may know whether a message comes from a business where they have made a purchase or whether email purportedly from a friend is written in their friend's writing style. Automated detection systems should be used as a first line of defense against phishing. However, since they are unlikely to be perfect, such systems should be complemented with training to improve the ability of users to recognize fraudulent email and websites.

Some experts have cast doubts on whether user training can be an effective way of preventing users from falling for phishing attacks. Jakob Nielsen, a web usability guru, has argued that educating users about security does not work [35]. Two recent papers tested user education in the form of Phishing IQ tests [3] and documentation for the extended validation feature in IE7 [23] and concluded that user education is ineffective. One security expert was recently quoted in a press report as saying, "User education is a complete waste of time. It is about as much use as nailing jelly to a wall," [16]. In short, the conventional wisdom seems to be that training users not to fall for phishing attacks is pointless.

This paper makes two research contributions. The first is a user study that demonstrates that existing anti-phishing educational materials are surprisingly effective if people actually read them. Our participants spent at most 15 minutes reading anti-phishing educational materials and then demonstrated significant improvements in their ability to recognize fraudulent websites when compared to a control group. The second contribution is an analysis of existing anti-phishing educational material. Based on the results of our user study and principles from the learning sciences, we present some suggestions for improving the content and presentation of training materials.

The remainder of this paper is organized as follows. We present related work in Section 2. In Section 3, we present the design and results of the user study we conducted to evaluate the effectiveness of the existing online training materials. In Section 4, we present our analyses of training materials using learning science principles. In Section 5, we present the lessons learned from the user study with some general suggestions to improve the training materials. We conclude in Section 6 with a discussion and directions for future work.

## 2. RELATED WORK

The volume of phishing attacks is increasing. According to the Anti-Phishing Working Group (APWG), the number of unique phishing websites reported in August 2006 was 10,091, compared to 7,197 in December 2005 [6]. Gartner estimates the total financial loss in 2006 due to phishing to be $2.8 billion [31]. Not only do victims lose their money and identities, but they also undergo significant emotional stress [26]. Many solutions to this problem have been proposed. We classified them into three categories: (1) preventing and detecting phishing scams; (2) tools to help users identify phishing web sites; and (3) user training.

### 2.1 Preventing and Detecting Phishing Scams

One way to combat phishing scams is to prevent spoofed emails and web pages from reaching end users. This can be achieved in a number of ways: (1) implementing filters to detect and delete emails automatically at the server [19], [40]; (2) finding and shutting down suspicious websites that have domain names similar to trusted brands; (3) installing toolbars to detect phishing websites (described in more detail in the next subsection); and (4) using domain keys and Sender Policy Framework (SPF) to verify the DNS domain of the email server and to reject forged addresses in the SMTP mail from address respectively [11], [39].

Given current Internet technology and regulatory status, phishing attacks cannot be prevented completely. For example, filters are clearly not 100% effective, since phishing emails still routinely reach the inbox of many users. In addition, false positives are a serious concern for email filters. Due to cross-border jurisdictional problems, it is often difficult to shutdown phishing websites quickly: according to the Anti-Phishing Working Group (APWG), phishing sites stay online on average for 4.5 days [6]. For the domain keys solution to be successful, the adoption rate among organizations needs to be high. In short, techniques for preventing and detecting phishing scams are not foolproof. Consequently, we believe that users should be trained to identify phishing emails. This paper presents a study that helps us understand how effective current training materials are, and takes us a step closer to our goal of developing effective anti-phishing training materials.

### 2.2 Tools to Help Users Identify Phishing Web Sites

Dozens of tools are available that provide visual indicators to help users identify potential phishing scams. For example, some anti-phishing toolbars display colored icons to indicate the degree of danger of a website, while others provide risk ratings, information about the age and physical location of a web site, and other information designed to inform users about potentially fraudulent sites. Some of the toolbars available are Account Guard [1], EarthLink [13], Google Toolbar [21], Netcraft [36], SpoofGuard [42], SpoofStick [41], and Zillabar [50]. In addition, anti-phishing tools are now built into the Microsoft Internet Explorer, Firefox, and Netscape Navigator web browsers.

Toolbars can be effective because they present potentially relevant aspects of the underlying system model to users (i.e. hidden state such as the age of the website). Having a clearer model of the current state of things can help reduce misconceptions about what the system is doing and help users make better decisions. However, studies have shown that users often do not understand or act on the cues provided by toolbars [33], [46]. In addition, 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 [49]. Other tools, such as PassPet and WebWallet, try to engage users by requiring them to interact actively with the tool before giving out sensitive information [44], [45], [48]. However, even these solutions ultimately rely on the users' ability to make the right decision.

Ye et al. [47] and Dhamija and Tygar [9] have developed "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. Herzberg and Gbara have developed TrustBar, a browser add-on that uses logos and warnings to help users distinguish trusted and untrusted websites [22]. More user studies are needed to assess the effectiveness of these approaches.

In all of the above systems, users are still involved in the decision-making process. These tools aid users in making a decision, but they do not make the decision for users. Studies have shown that users frequently disregard the information presented by anti-phishing tools, often due to inaccurate beliefs about the nature of phishing attacks [46]. This suggests a need to raise users' awareness about phishing and to train users on how to avoid falling for these attacks.

## 2.3 User Training

A few approaches have focused on educating and training users about phishing. The most basic approach is to provide online information regarding phishing. This has been done by government organizations [18], non-profit organizations [5] and businesses [14]. Another approach allows users to take tests on phishing websites and emails. For example, Mail Frontier [30] 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. One study examined the effectiveness of such phishing tests at educating users, and concluded that they are not effective [3]. In another study, Robila et al. trained students in a class room setting, and demonstrated that class discussion and exercises made students more aware of phishing and better at recognizing phishing attacks [38].

Researchers have also tried a *contextual training* approach in which users are sent phishing emails to probe their vulnerability. At the end of the study, users are typically given additional materials informing them about phishing attacks in general. This approach has been used at Indiana University in studies conducted on students about contextual attacks making use of personal information (also known as *spear-phishing*) [25]; at West Point [20], [23]; and at a New York State Office [37].

In another paper, we presented the design and evaluation of an email-based approach to train people to avoid phishing attacks [28]. We called this approach *embedded training*, in that it trains people during their regular use of email. As in previous studies, we sent our subjects phishing emails, and then presented an intervention warning people who had fallen for our messages. Our study was conducted in a laboratory and interventions were presented immediately when users clicked on a phishing link in the email, rather than at the end of the study. Our goal was to evaluate how effective various intervention designs were and how well people could transfer knowledge from one situation to another. We created several designs based on learning sciences, and found that our interventions were more effective than standard security notices.

While previous work evaluated the effectiveness of phishing tests, classroom instruction, and email-based training, the work presented in this current paper examines the effectiveness of existing web-based training materials.

# 3. USER STUDY

The goal of our study is to determine the effectiveness of available web-based anti-phishing training materials. In this section we present the study design, participant details, and results.

## 3.1 Study Design

We based the design of our user study on Dhamija et al.'s study of users' ability to identify phishing websites [10]. Users were given the following scenario: "You have received an email message that asks you to click on one of its links. Imagine that you have clicked on the link to see if it is a legitimate website or a spoofed website." We then presented users with twenty websites and asked them to state whether a website was legitimate or phishing, and to tell us how confident they were in their judgments (on a scale of 1 to 5, where 1 means not confident at all, and 5 means very confident).

We used 20 websites for the study. Ten of them were phishing sites from the APWG database. The other ten were legitimate websites from popular financial institutions and online merchants, as well as random websites. We divided up the twenty websites into two groups (A and B), with 5 phishing sites and 5 legitimate sites in each group. In our test, participants were asked to view one group of sites (pre test), take a fifteen minute break to complete a task prescribed by the conditions below, and then view the second group of websites (post test). We randomized the order of pre test and post test, so that half the users used Group A in the pre test, and half used group B in the pre test. The list of websites used is shown in Table 2. We hosted the phishing websites on the local computer by modifying the host DNS file. Thus, our participants were not actually at risk and we were able to show them phishing sites even after they had been taken down. Before the study, we mentioned to participants that they can use any necessary means they want to determine the websites' legitimacy other than calling the institution. We also let participants use a separate web browser if they wanted, without prompting them about how or why this might be useful. Some participants used this other web browser to access a search engine to help determine whether a web site was legitimate or not. We used Camtasia Studio [7] to record our participants' computer screens and spoken comments during the study.

Using a between-subjects design, we had two conditions:

- **Control condition**: In this condition, participants were asked to play a simple computer game (such as solitaire or minesweeper) between the pre and post test.
- **Training material condition**: During the break between pre and post test, participants were asked to read what we judged to be the best web-based educational material on phishing currently available. The rest of the setup was identical to that used for the control condition.

## 3.2 Training Materials

We compiled and evaluated a list of 24 online anti-phishing training materials to select the materials for our study. Our final selections were eBay's tutorial on spoofed emails [14], Microsoft's Security tutorial on Phishing [32], and Phishing E-card from the U.S. Federal Trade Commission, [17]. We also included a URL tutorial from MySecureCyberspace [34], which is a portal for educating people about security risks and countermeasures on the Internet. In Table 1, we present information about the format of the instruction, length of the instruction in words, length of the instruction in printed pages, number of graphic examples, and what concepts they try to teach. All the training materials that we used for the study had some form of link to other resources about phishing and security in general.

Almost all the training materials started with some basic definition of phishing. An example definition is "Claiming to be sent by well-known companies, these emails ask consumers to reply with personal information, such as their credit card number, social security number or account password" [14]. All the materials presented a variation of this definition. Almost all the materials initially also provided definitions of “spoof emails” and then connected them to phishing emails.

These training materials also highlighted some characteristics of phishing emails and provided suggestions for how to avoid falling for such scams. Table 1 presents the characteristics of the emails and the suggestions. Almost all the materials mention some version of “organizations do not request personal information through emails.” Finally, these materials also presented information about what to do after falling for phishing emails. These suggestions included: reporting or forwarding the phishing email to spoof@ebay.com, and reporting them to the FTC.

**Table 1: Information about the training materials.**

| Source | Content format | Length in words | # of printed pages | # of graphic examples | Cues to look for in the email | Suggestions to avoid falling for scams |
|---|---|---|---|---|---|---|
| Microsoft | Webpage | 737 | 3 | 2 | - Urging urgent action<br>- Non-personalized greeting<br>- Requesting personal information through email | - Mouse over the link to see what web site it really goes to |
| eBay | Webpage | 1276 | 5 | 8 | - Urging urgent action<br>- Non-personalized greeting<br>- Requesting personal information through email<br>- Sender email address<br>- Links in the email<br>- Legitimate eBay address versus fake eBay address | - Open a new browser to type in the URL<br>- Never click on the link in an email<br>- How to identify legitimate eBay address |
| FTC Phishing E-card | Video | N/A | N/A | N/A | - Requesting personal information through email | - Do not provide personal information requested through email |
| URL tutorial from My Secure Cyberspace | Webpage | 236 | 1 | 0 | N / A | N / A |

## 3.3 Participants

We recruited 14 participants for each condition, for a total of 28 people. To recruit participants, we posted flyers around our campus, and posted recruitment messages on university bulletin boards and on craigslist.com.

We screened participants with respect to their knowledge of computers in general, aiming to recruit only participants who could be considered “non-experts.” We recruited users who answered “no” to two or more of the following screening questions: 1) whether they had ever changed preferences or settings in their web browser, 2) whether they had ever created a web page, and 3) whether they had ever helped someone fix a computer problem. These questions have served as good filters to recruit non-experts in other phishing-related studies [12], [27].

Our participants had the following demographics:

- Gender: 33% percent of the participants were male, and 67% percent of the participants were female.

- Age: 93% of the participants were 18-34 years old, 7% were 35-44 years old.
- Education Level: 9% of the participants had high school or less education, 48% of the participants were college undergraduates, 22% were college graduates, and 22% had graduate degrees.
- Race: 56% of the participants were Asian, 37% were white, 4% were African American, and 7% declined to answer.
- Years on the Internet: 15% of the participants have been using the Internet less than 5 years, 70% have been using the Internet between 6 and 10 years, and 15% have been using it between 11 and 15 years.
- Number of hours spent online per week: 4% of the participants use the Internet less than 5 hours each week, 19% use it 6-15 hours each week, 50% use it 16-30 hours each week, 15% use it 31-50 hours each week, and 12% use it more than 51 hours each week.

## 3.4 Results

In this section, we present the result of our study. We find that subjects in the training condition demonstrated significant improvements in their ability to recognize fraudulent websites.

### *3.4.1 Participants Score and Behavior*

We found no significant correlation between the participants' score and gender (spearman rho = 0.31, n = 28, p = 0.309), age (spearman rho = 0.253, n = 28, p = 0.404), education (spearman rho = 0.20, n = 28, p = 0.51), race (spearman rho = 0.329, n = 28, p = 0.272), years on the Internet (spearman rho = 0.11, n = 28, p = 0.72), or number of hours spent online per week (spearman rho = -0.116, n = 28, p =0.706). Other studies have also found no correlation between these demographics and susceptibility to phishing [10], [12].

### *3.4.2 Effectiveness of Training*

We use two metrics to measure the effectiveness of training: the number of false positives and the number of false negatives. A false positive takes place when a legitimate site is mistakenly judged as a phishing site. A false negative takes place when a phishing site is incorrectly judged to be a legitimate site.

False negatives are usually worse than false positives in phishing, because the consequence of mistaking a legitimate site to be phishing is a matter of inconvenience, whereas the consequence of mistaking a phishing site to be real can lead to identity theft.

In our analysis, the false positive and false negative rates are calculated as:

$$\text{False Positive Rate} = \frac{\text{number of false positives}}{\text{number of legitimate sites}}$$

$$\text{False Negative Rate} = \frac{\text{number of false negatives}}{\text{number of phishing sites}}$$

We found that for the training group, there is a significant reduction in the false negative rate after the training — from 0.40 to 0.11 (paired t-test: µ1=0.40, µ2=0.11, p = 0.01, DF = 13). There is no statistically significant change in the false negative rate for the control group (paired t-test: µ1=0.47, µ2=0.43, p=0.29, DF=13). Figure 1 shows the comparison of false negatives in both the conditions in pre and post-test evaluation.

We also tabulated the training group's performance by website. In Table 2 we show, for each website, the percentage of correct answers before training and after training. The data shows that users made better decisions on eleven of the twenty sites, did not improve on four sites, and performed worse on five of them. While the false positive rate remained virtually unchanged for the control group, it increased from 0.31 to 0.41 in the training group. However, this increase is not statistically significant. (paired t-test: µ1=0.31, µ2=0.41,p=0.12, DF = 13). We explain the reason for the increase in false positives in detail in Section 3.4.3.

**Table 2: Percentage of correct answers for the training group before and after training**

| Website | Real / Spoof | Description | Pre Training % correct (avg conf) | Post Training %correct (avg conf) | Change |
|---|---|---|---|---|---|
| Paypal | Spoof | Fake URL bar displaying the real paypal URL; not requesting much information | 14% (4.0) | 71% (4.4) | +57% |
| PNC Bank | Spoof | Bank account update; pop-up window over the real PNC Bank web site; security lock; requesting credit card number | 57 (3.7) | 100 (4.1) | +43% |
| Citicards | Spoof | Citicard account update; lock on the page; requesting a lot of information | 42 (4.3) | 85 (4.5) | +43% |
| Royal Bank of Canada | Spoof | Sign in online banking page; layered information request; URL has no resemblance with the bank. | 42 (3.3) | 85 (4.8) | + 43% |
| HSBC | Spoof | Internet banking login page; layered information request; IP address URL | 50 (4.0) | 85 (4.8) | + 35% |
| Chase Student | Real | Primitive looking page with few graphics and links | 28 (4.5) | 50 (4.3) | +22% |
| Paypal | Real | Paypal login page | 85 (4.5) | 100 (4.5) | +15% |
| Barclays | Spoof | Faked Barclays login page; layered information request; IP address URL | 85 (4.1) | 100 (4.4) | +15% |
| AOL | Spoof | AOL account update, deceptive domain myaol.com | 85 (4.0) | 100 (4.7) | +15% |
| Halifax Bank | Spoof | Halifax bank login page; deceptive domain halifax-cnline.co.uk. | 85 (4.6) | 100 (4.4) | +15% |
| eBay | Real | eBay register page; requesting lots of information | 28 (5.0) | 42 (4.6) | +14% |
| Etrade | Real | Etrade home page | 100 (4.1) | 100 (4.2) | 0% |
| eBay | Spoof | Faked eBay login page; IP address URL | 85 (4.8) | 85 (4.8) | 0% |
| Wellsfargo bank | Spoof | Faked Wellsfargo home page; layered information request; sub domain deception with URL online.wellsfargo.wfosec.net | 71 (4.0) | 71 (3.8) | 0% |
| Desjardins | Real | Account login page; unfamiliar foreign bank | 57 (3.0) | 57 (3.5) | 0% |
| Card Financials Online | Real | Card Financial Online (part of MBNA); domain name has nothing to do with MBNA. | 42 (4.3) | 28 (3.5) | -14% |
| Bank of America | Real | Bank of America home page; URL: onlineast.bankofamerica.com | 83 (4.2) | 57 (3.7) | -26% |
| Chase online | Real | Online banking login page; URL: chaseonline.chase.com | 100 (4.5) | 71 (2.8) | -29% |
| Citibank | Real | Citibank login Page; URL: web-da.us.citibank.com | 71 (4.0) | 42 (4.0) | -29% |
| US Bank | Real | Online banking login page; URL: www4.usbank.com | 100 (4.2) | 57 (4.2) | -43% |

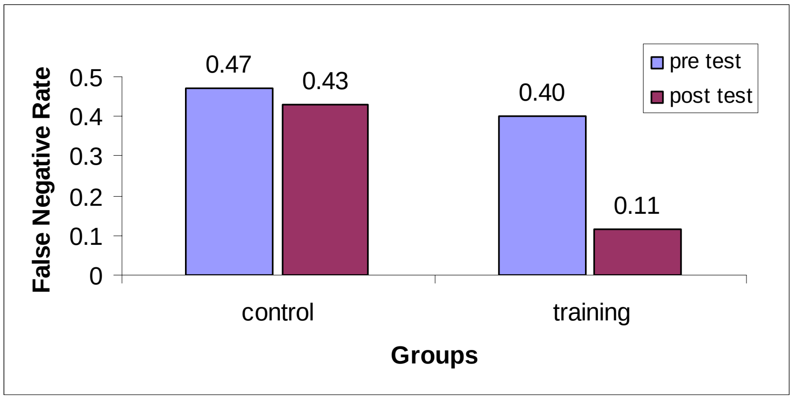


**Figure 1: False negative rates. N(control) = 14, N(training) = 14. We found no significant change in the false negative rate for the control group, but did find a statistically significant reduction in the false negative rate for the training group.**

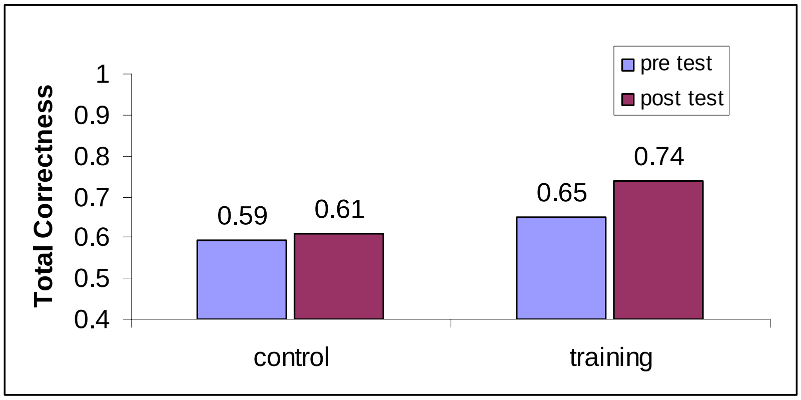


**Figure 2: Total correctness. N(control) = 14, N(training) = 14. We found no significant change in the total correctness for the control group, but did find a small statistically significant difference for the training group.**

Total correctness is defined as the ratio of the number of correctly identified websites to the total number of websites shown to the participants. Figure 2 shows the total correctness for the control and the training group. The total correctness for the control group changes from 0.59 in the pre test to 0.61in the post test, however, this change is not statistically significant. The total correctness of the training group changes from 0.65 to 0.74. This change is marginally statistically significant (p=0.11). The increase in false positive rate offsets the improvements from finding phishing websites. While the false positive rate remained virtually unchanged for the control group, it increased in the training group.

### *3.4.3 User Strategies for determining website legitimacy*

Our analysis shows that participants used a variety of strategies to determine website legitimacy and these strategies vary in their effectiveness. This is in alignment with other research. Previous user studies [9], [12] have discussed users' strategies for determining website legitimacy. Dhamija et al. categorized user strategies into five categories: relying on web contents only, relying on content and domain name only, relying on content with address and https, relying on all the above plus padlock icon, and relying on all of the above plus certificates. Downs et al. [12] discuss cues that users are sensitive to when judging the legitimacy of a site. Such cues include secure site lock icons, broken images on the webpage, unexpected or strange URLs, and the indication of an https connection.

During our user study, we asked users to think aloud the reasons for their decisions. We recorded these reasons and organized them into seven categories: design and content, URL, information requested by the website, consistency, search engine, prior knowledge, and security indicators. During our analysis we reviewed the recordings of each user's session and coded the strategies used at each site into one of these seven categories. As noted earlier in Section 3.1, we let participants in both conditions use a second web browser for whatever purpose they desired, and some used this extra web browser to load a search engine. Table 3 explains these strategies in detail and shows the percentage of web sites where they were used by the participants in the control group. The percentage was calculated across all websites and all participants. This percentage adds up more than 100% because a user could use multiple strategies together. To ensure they did not bias participants, study administrators only prompted participants to speak about their decisions if they did not do so (which usually only happened at the beginning of a study). At no point did the test administrators provide hints or ask participants to look at certain cues.

**Table 3: Strategies used by the control group.**

| Strategies | Examples | % of web sites where strategy was used (across all participants in control group) |
|---|---|---|
| Design & Content | - The design of the websites is poor/ professional.<br>- The links (images) are functioning / broken.<br>- Existence of up-to-date contact information, copyright statement, privacy and security statements.<br>- There are security locks in the content, verisign symbol, TRUSTe logo | 42% |
| URL | - The URL contains numbers.<br>- The address looks suspicious. | 31% |
| Information requested | - Amount of information requested is too much / all right.<br>- The website is / not requesting sensitive information.<br>- It is all right / weird for website to request my information here. | 19% |
| Consistency | - All the links on one page are pointing to the same site.<br>- Logos and colors of different pages match. | 16% |
| Search engine | - Using a search engine to double check the legitimacy of the site. | 16% |
| Prior knowledge | - I have an account with the company, I know this company.<br>- I have seen the website / know the company.<br>- I have / know someone who is a victim of this site. | 6% |
| Security indicator | - The URL has https in them.<br>- There is secure browser pad lock. | 3% |

### *3.4.4 What users are learning and what they are not learning*

We compared the strategies that our participants used before and after the training (Table 4). Our results show that the participants in the training group relied on the design and content of a website as well as their prior knowledge less often after training than before training. Furthermore, they examined the URLs of the webpage and the amount of information requested more often during post training than pre training. Both of these results are encouraging, as they show that our participants learned to avoid poor strategies and started to adopt good strategies. Finally, we did not observe any significant changes in the control group.

**Table 4: Percentage change in strategies that participants used**

| Strategies | Training (change) | Control (change) |
|---|---|---|
| Design & content | -15% | -1% |
| Prior knowledge | -11% | -5% |
| URL | +23% | +2% |
| Information requested | +13% | -3% |

The training materials taught participants that phishing sites often request sensitive user information (such as credit card PIN numbers and social security numbers), whereas legitimate companies do not. After training, our participants paid more attention to what information the websites were requesting. This leads us to conclude that users are learning these techniques from the training materials.

As for URLs, the Microsoft and eBay training materials teach (1) the correct URL for their respective sites, and (2) some example URLs that phishers use to trick people. However, the training materials do not provide general information about identifying phishing URLs.

For identifying IP-address-based scams (which use IP addresses in the URL instead of a human-readable domain name), participants in the training group seemed to perform quite well, as only one user failed to recognize them (and failed twice on it). This participant's rationale was that "both of the two sites do not ask for much information." In contrast, in the control group, our participants failed to identify seven IP-address-based phishing sites.

Phishing sites also use deceptive URLs that are hard to detect. In Dhamija et al.'s study, 92% of the users fell for www.bankofthevvest.com (two *v*'s, instead of a *w*). In our study, none of the participants in the training group fell for the deceptive domain halifax-cnline.com (change of "o" to "c" in halifax-online.com) after training. Our participants noticed the typo immediately.

**Table 5: Reasons for post training failures. A primary cause was a misleading or confusing URL.**

| Website | Pre Training % correct (avg conf) | Post Training %correct (avg conf) | Change | Reasons for failure |
|---|---|---|---|---|
| MBNA business (real) | 42 (4.3) | 28 (3.5) | -14% | Domain name usecfo.com has nothing to do with MBNA. |
| Bank of America (real) | 83 (4.2) | 57 (3.7) | -26% | URL onlineast.bankofamerica.com, users were expecting www.bankofamerica.com |
| Chase online (real) | 100 (4.5) | 71 (2.8) | -29% | URL chaseonline.chase.com, user expecting www.chase.com |
| Citibank (real) | 71 (4.0) | 42 (4.0) | -29% | URL web-us.da.citibank.com, user expecting www.citibank.com |
| US Bank (real) | 100 (4.2) | 57 (4.2) | -43% | URL www4.usbank.com, user expecting www.usbank.com |

However, our participants had a hard time interpreting longer URLs, especially URLs using sub-domains. For example, many of the participants in the training condition labeled wellsfargo.com.wfcnet.net as legitimate because the word *wellsfargo.com* appeared in the name. Similarly, they labeled chaseonline.chase.com and web-da.citibank.com as phishing sites because they misunderstood the URL. Not understanding the URL was the primary cause of errors after training (see Table 5).

### *3.4.5 User Response to Training materials*

The amount of time that subjects spent on the training materials ranged from 4.30 to 11.00 minutes (mean = 6.99, s.d. = 2.34, var = 5.49). Among the participants tested in the training group, only three users clicked on some of the resource links to read more about phishing (two in the FTC materials and one in MySecureCyberspace). All of our participants in the training condition completely read through the Microsoft and FTC materials, while only one completely read through the eBay materials and four read through the MySecureCyberspace materials. Participants spent most of the time on the Microsoft and FTC materials and less than 3.5 minutes on the eBay tutorial. Some of the participants assumed that the eBay tutorial had only one page, while it actually had five. Except for one person, all others skimmed through the tutorial materials quickly. Participants generally responded positively to the education material. When asked to rate the materials in terms of the educational value and fun level, 93% of the participants rated the materials as very or extremely educational, while 29% rated the materials as very or extremely fun. Many of the participants highlighted the FTC E-card animation as the best among the materials.

To summarize, the ability to identify phishing websites improved due to training. Subjects learned that legitimate companies do not request sensitive information or login credentials through email. Users were able to unlearn some of their bad strategies and learn good strategies. However, they still were unable to properly parse longer URLs with sub-domains.

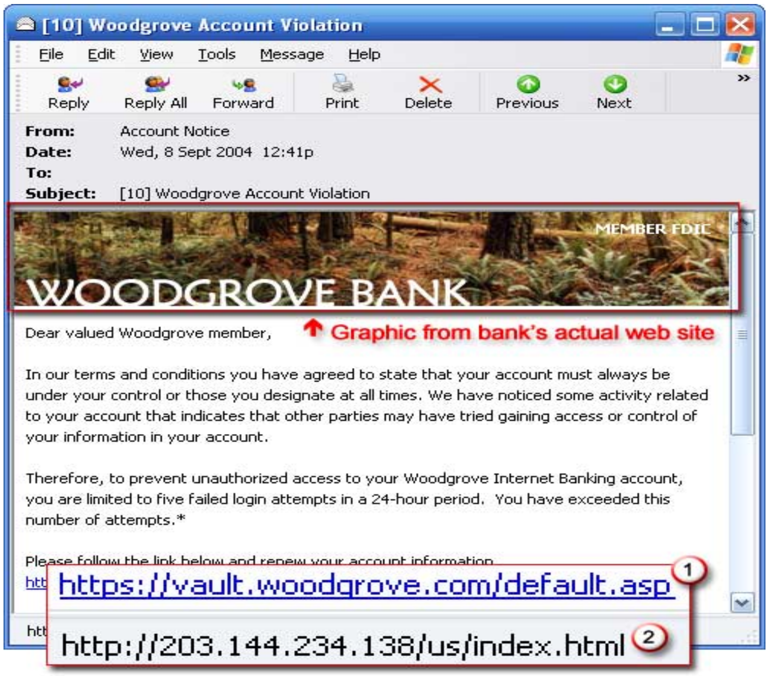
## 4. ANALYSIS OF EXISTING TRAINING MATERIALS

In the previous section we discussed the content and the effectiveness of existing online anti-phishing training materials. Although the training materials used turned out to be surprisingly effective, in this section we discuss their presentation style as well as strategies to make them even more effective through principles derived from the learning sciences literature.

**Table 6: Availability of principles in different training materials / mechanisms;**

**√ is available, X is not available, & is partially available**

| Principle | eBay | FTC | Microsoft | MySecure Cyberspace |
|---|---|---|---|---|
| Multimedia principle | √ | √ | √ | X |
| Contiguity | & | X | & | X |
| Personalization and story | X | √ | X | X |
| Simplicity | X | √ | & | X |
| Immediate feedbacks | X | X | X | X |



From: Account Notice
Date: Wed, 8 Sept 2004 12:41p
To:
Subject: [10] Woodgrove Account Violation

Dear valued Woodgrove member,

In our terms and conditions you have agreed to state that your account must always be under your control or those you designate at all times. We have noticed some activity related to your account that indicates that other parties may have tried gaining access or control of your information in your account.

Therefore, to prevent unauthorized access to your Woodgrove Internet Banking account, you are limited to five failed login attempts in a 24-hour period. You have exceeded this number of attempts.*

Please follow the link below and renew your account information.

https://vault.woodgrove.com/default.asp

http://203.144.234.138/us/index.html

**Figure 3: One of the training images from the online training materials [32]**

Learning science is the body of research that examines the foundations of how people develop knowledge and learn new skills. Learning science researchers have developed learning science instructional design principles that can be applied to impart effective learning [4], [8]. These principles have been applied and evaluated in the context of e-learning and intelligent tutors. Here we discuss some of the learning science instructional design principles that we used in evaluating the existing online training materials.

Table 6 presents our analysis of the training materials with respect to the learning science instructional design principles we examined. The rest of this section examines each of these principles in depth.

### 4.1 Multimedia Principle

This principle states that adding graphics to words can improve learning. In particular, explanative illustrations should be used to help people understand the material better, while the use of purely decorative illustrations should be minimized [8]. Table 6 presents our analysis of all the online training materials, showing that all of them except MySecureCyberspace used images along with text. In examining the training materials, we found that some illustrations were used more for decorative purposes than for explanative purposes. Another issue is that one set of training materials included a graphical example of a deceptive URL (Figure 3) but did not provide an explanation for the image in a caption or in the body of the text [32]. The multimedia principle would suggest designing training materials so that text and images are presented together, as discussed by Kumaraguru et al. [28].

### 4.2 Contiguity Principle

This principle states that placing corresponding words and graphics near each other can improve learning. Studies have shown that integrated text and graphics produce better learning than when they are separated [8]. One common violation that we found in the online training materials was that example images and corresponding text were located far apart from each other. In the majority of cases, the cause was long web pages that required scrolling. In a few instances, the cause was information being presented on different web pages. Our analysis suggests that eBay did the best job in terms of integrating text and graphics. However, Table 6 also shows that none of the existing online training materials apply this principle consistently.

## 4.3 Personalization and Story Based Instruction Principle

This principle states that using a conversational style can be more effective for learning than a formal style. Using characters and a story line can also improve learning [8]. Most of the online materials on phishing do not implement this principle. From Table 6, we can see that only the FTC has implemented this principle.

## 4.4 Simplicity

Keeping the instruction simple and short is an essential principle for designing training materials. Research has shown that people learn better when their working load memory is minimized [4]. Other studies have shown that length of the instruction is one of the reasons why people do not read the training materials that are available through security notices. This principle suggests that short training materials will be most effective [28].

## 4.5 Provide Immediate Feedback on Errors

This principle suggests that providing immediate feedback to users when they make an error can induce better learning [4]. Providing training materials immediately after users fall for phishing emails offers immediate feedback [28]. Most online materials do not make use of this principle: they are not designed to give feedback. Materials that include game or test components can provide immediate feedback.

# 5. DISCUSSION

In the previous sections we presented the results of a user study in which users spent an average of seven minutes reading existing web-based anti-phishing educational materials. Our results show that users demonstrated significant improvements in their ability to recognize fraudulent websites after reading the online training materials.

Our results appear to contradict the results of previous user studies [3], [23]. For example, Anandpara, et al. suggested that "education only increases awareness, but not real ability." However, we believe their results may be due to the poor quality of the training materials they used. Their study used only a one-page FTC phishing alert to train participants. This alert does not provide clear advice about identifying phishing emails and does not follow any of the learning science principles discussed in this paper.

Our results also appear to contradict the conventional wisdom that training users to avoid phishing attacks do not work. This wisdom is generally based on the assumptions that (1) computer security concepts are difficult to teach to non-experts; and (2) because security is a secondary task for users, they will not spend time reading training materials. Although we generally agree with these assumptions, we believe the obstacles they pose can be overcome. We believe that people can be taught to identify phishing scams without the need for them to understand complicated computer security concepts. We demonstrated that by teaching a few simple concepts to our user study participants, they were able to identify most of the phishing web sites. For example, participants learned that IP addresses in URLs and websites that request sensitive information out of context are generally indicative of phishing sites. Thus, the difficulty of teaching users complicated computer security concepts may not actually be an obstacle. The second obstacle may be more difficult to overcome outside of a situation where people are required to read training materials. However, as we have demonstrated that training materials can be effective if people do read them, it seems worthwhile to explore ways of getting people to read them For example, our group is developing an embedded training system [28] as well as a web-based anti-phishing game.

Further work is needed to determine the most effective way of delivering training materials so that people will read them, as well as ways to improve existing training materials to make them even more effective. Based on the results of our study, we propose three ways of improving existing training materials:

- **Teach users that taking the design and content of a website as a cue for determining its legitimacy is a bad strategy.** Phishers can fake the design and the content of websites easily, and our analysis shows that even after the training, users still use the design and the content of the webpage as one of the primary cues. The educational materials we examined do not teach users to avoid this strategy.
- **Focus on longer URLs, and some basics of domain name knowledge.** Our study results also show that users' lack of knowledge about URLs and domain names make them vulnerable to phishing sites whose sub-domain name match the real organization's domain. Furthermore, an increase in awareness without adequate knowledge increases the false positive rate. Therefore, we recommend teaching the basics of domain names and URLs.

Our discussion of instructional design principles from the learning sciences can help to design better training materials. From our user study we also found that the most effective training materials complied with most of the instructional design principles. We believe that better learning can occur when training materials relate to users' prior knowledge. We also found that users use counter productive strategies like examining the design and content of the website to make their decision, so training materials have to address these myths.

As in other user studies, there are some limitations in our study also. Our participants were more educated and younger than the general Internet user population, so the results may not be generalizable to other groups. In addition, our study evaluated participants' ability to identify phishing web sites without showing them the phishing email messages that would typically take someone to such web sites.

## 6. CONCLUSIONS AND FUTURE WORK

In this paper we have presented the results of a user study that evaluated the effectiveness of existing online anti-phishing training materials. We demonstrated that—contrary to popular wisdom—anti-phishing user education can be effective: users get significantly better at identifying phishing websites when they actually read training materials. We also showed the different strategies that users adopt to recognize phishing sites, and how those strategies evolve due to the training. We also presented an analysis of existing training materials using learning science principles, and derived recommendations to develop further training materials in the context of phishing.

We have not tested the relative importance of the learning science principles in the context of phishing education; we plan to do this as a future work. We also plan to test whether these principles can be generalized to educate users about other online security issues.

## ACKNOWLEDGMENTS

We gratefully acknowledge support from National Science Foundation grant number CCF-0524189 entitled "Supporting Trust Decisions." The authors would like to thank all members of the Supporting Trust Decisions project for their feedback. In particular we would like to thank Yong Woo Rhee and Elizabeth Nunge for conducting some of the user studies that were discussed in this paper. We would also like to thank CyLab, Carnegie Mellon University. The authors would also like to thank Dr. Anne Fay of Eberly Center for Teaching Excellence and Dr. Vincent Aleven for their advice on the learning science aspects of this study.

## REFERENCES

[1] Account Guard. Retrieved Nov 3, 2006, http://pages.ebay.com/ebay_toolbar/.

[2] Adams, A. and M. A. Sasse. 2005.Users are not the enemy: why users compromise security mechanisms and how to take remedial measures. In Lorrie Cranor and Simson Garfinkel (Eds.) *Security and Usability: Designing Secure Systems that People Can Use*. O'Reilly.

[3] Anandpara, V., Dingman, A., Jakobsson, M., Liu, D., and Roinestad, H. 2007. Phishing IQ tests measure fear, not ability. In Usable Security (USEC'07). http://usablesecurity.org/papers/anandpara.pdf.

[4] Anderson, J. R., A. T. Corbett, K. R. Koedinger and R. Pelletier. 1995. Cognitive Tutors: Lessons Learned. *The Journal of The Learning Science.* 4 (2), 167 – 207.

[5] Anti-Phishing Working Group (APWG). Retrieved on Sept 20, 2006. http://www.antiphishing.org/.

[6] Anti-Phishing Working Group. Phishing Activity Trends Report. 2006. http://www.antiphishing.org/reports/apwg_report_August_2006.pdf.

[7] Camtasia Studio. Retrieved Nov 9, 2006. http://www.techsmith.com/camtasia.asp.

[8] Clark, R. C. and R. E. Mayer. 2002. *E-Learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning*. Pfeiffer.

[9] Dhamija, R. and J. D. Tygar. 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.

[10] Dhamija, R., J. D. Tygar. and M. Hearst. 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.

[11] Domain keys. Retrieved Nov 5, 2006. http://en.wikipedia.org/wiki/Domain_keys.

[12] Downs, J., M. Holbrook and L. Cranor. 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.

[13] EarthLink. Retrieved Nov 3, 2006, http://www.earthlink.net/software/free/toolbar/.

[14] eBay. Spoof Email Tutorial. Retrieved March 7, 2006, http://pages.ebay.com/education/spooftutorial/.

[15] Emigh, A. Online Identity Theft: Phishing Technology, Chokepoints and Countermeasures. October, 2005. Retrieved Nov 3, 2006, http://www.antiphishing.org/Phishing-dhs-report.pdf.

[16] Evers, J. Security Expert: User education is pointless. Retrieved, Jan 13, 2007, http://news.com.com/2100-7350_3-6125213.html.

[17] Federal Trade Commission. An E-Card for You game. Retrieved Nov 7, 2006, http://www.ftc.gov/bcp/conline/ecards/phishing/index.html.

[18] Federal Trade Commission. How Not to Get Hooked by a Phishing Scam. Retrieved Nov 7, 2006, http://www.ftc.gov/bcp/edu/pubs/consumer/alerts/alt127.htm.

[19] Fette, I., N. Sadeh and A. Tomasic. Learning to Detect Phishing Emails. 2006. ISRI Technical report, CMU-ISRI-06-112. Retrieved Sep 2, 2006, http://reports-archive.adm.cs.cmu.edu/anon/isri2006/CMU-ISRI-06-112.pdf.

[20] 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.

[21] Google Toolbar. Google. Retrieved Nov 3, 2006, http://www.google.com/tools/firefox/safebrowsing/.

[22] Herzberg, A., and Gbara, A. 2004. TrustBar: Protecting (even Naïve) Web Users from Spoofing and Phishing Attacks. Cryptology ePrint Archive, Report 2004/155. http://eprint.iacr.org/2004/155.

[23] Jackson, C., Simon, D., Tan, D., and Barth, A. 2007. An evaluation of extended validation and picture-in-picture phishing attacks. In Usable Security (USEC'07). http://usablesecurity.org/papers/jackson.pdf.

[24] Jackson, J. W., A. J. Ferguson and M. J. Cobb. 2005. Building a University-wide Automated Information Assurance Awareness Exercise: The West Point Carronade. 35th ASEE/IEEE Frontiers in Education Conference. 2005. http://fie.engrng.pitt.edu/fie2005/papers/1694.pdf.

[25] 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.

[26] James, L. 2005. *Phishing Exposed*. Syngress, Canada.

[27] 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.

[28] Kumaraguru, P., Y. Rhee, A. Acquisti, L. Cranor, J. Hong and E. Nunge. 2007. Protecting People from Phishing: The Design and Evaluation of an Embedded Training Email System. To appear in the Conference on Computer Human Interaction.

[29] Lininger, R. and R. Dean. 2005. *Phishing: Cutting the Identity Theft Line*. Wiley, publishing Inc. Indianapolis, Indiana, USA.

[30] Mail frontier. Mailfrontier Phishing IQ test. Retrieved Sept 2, 2006, http://survey.mailfrontier.com/survey/quiztest.html.

[31] McMillan, R. Consumers to lose $2.8B to phishers in 2006: Gartner says fewer, but bigger, attacks will gain more for criminals. November, 2006, Retrieved Nov 10, 2006, http://www.infoworld.com/archives/emailPrint.jsp?R=printThis&A=/article/06/11/09/HNgartnerphishing_1.html.

[32] Microsoft. Recognizing phishing scams and fraudulent emails. Retrieved Oct 15, 2006. http://www.microsoft.com/athome/security/email/phishing.mspx.

[33] 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.

[34] MySecureCyberspace. Uniform Resource Locator (IRL). Retrieved Oct 15, 2006. http://www.mysecurecyberspace.com/encyclopedia/index/uniform-resource-locator-url-.html.

[35] Nielsen, J. 2004. User education is not the answer to security problems. http://www.useit.com/alertbox/20041025.html.

[36] Netcraft. Retrieved Nov 3, 2006, http://toolbar.netcraft.com/.

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

[38] 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. 2006. pp 237-241. New York, NY, USA.

[39] Sender Policy Framework (SPF). Retrieved Nov 5, 2006. http://en.wikipedia.org/wiki/Sender_Policy_Framework.

[40] SpamAssasin. Retrieved Nov 5, 2006, http://spamassassin.apache.org/.

[41] SpoofStick. Retrieved Sept 2, 2006, http://www.spoofstick.com/.

[42] SpoofGuard. Retrieved Sept 2, 2006, http://crypto.stanford.edu/SpoofGuard/.

[43] Whitten, A and J. D. Tygar. 1999. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. Proceedings of the 8th USENIX Security Symposium. http://www.cs.berkeley.edu/~tygar/papers/Why_Johnny_Cant_Encrypt/USENIX.pdf.

[44] Wu, M. Fighting Phishing at the User Interface. 2006. MIT PhD. thesis. http://groups.csail.mit.edu/uid/projects/phishing/minwu-thesis.pdf.

[45] Wu, M., R. C. Miller and Little, G. 2006. Web Wallet: Preventing Phishing Attacks By Revealing User Intentions. 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/10.1145/1143120.1143133.

[46] Wu, M., R. C. Miller and S. L. Garfinkel. 2006. Do Security Toolbars Actually Prevent Phishing Attacks? In the Conference on Human Factors in Computing Systems. http://www.simson.net/ref/2006/CHI-security-toolbar-final.pdf.

[47] Ye, Z. and Sean S. 2002. Trusted Paths for Browsers. Proceedings of the 11th USENIX Security Symposium. pp. 263 - 279. USENIX Association. Berkeley, CA, USA.

[48] Yee, K. P. and Sitaker K. PassPet: Convenient Password Management And Phishing Protection. 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/10.1145/1143120.1143126.

[49] 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.

[50] ZILLAbar. International Software Systems Solutions, Inc. Retrieved Nov 3, 2006, http://zillabar.com/home.do.