School of Phish: A Real-Word Evaluation of Anti-Phishing Training

Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti,
Lorrie Cranor, Jason Hong, Mary Ann Blair, Theodore Pham
Carnegie Mellon University
{pkumarag, jcransh, acquisti,
lorrie, jasonhon, mc4t, telamon}@andrew.cmu.edu

ABSTRACT

PhishGuru is an embedded training system that teaches users to avoid falling for phishing attacks by delivering a training message when the user clicks on the URL in a simulated phishing email. In previous lab and real-world experiments, we validated the effectiveness of this approach. Here, we extend our previous work with a 515-participant, real-world study in which we focus on long-term retention and the effect of two training messages. We also investigate demographic factors that influence training and general phishing susceptibility. Results of this study show that (1) users trained with PhishGuru retain knowledge even after 28 days; (2) adding a second training message to reinforce the original training decreases the likelihood of people giving information to phishing websites; and (3) training does not decrease users' willingness to click on links in legitimate messages. We found no significant difference between males and females in the tendency to fall for phishing emails both before and after the training. We found that participants in the 18-25 age group were consistently more vulnerable to phishing attacks on all days of the study than older participants. Finally, our exit survey results indicate that most participants enjoyed receiving training during their normal use of email.

Categories and Subject Descriptors

H.1.2 [**Models and Applications**]: User / Machine systems—*human factors, human information processing*; K.6.5 [**Management of Computing and Information Systems**]: Security and protection education

General Terms

Design, experimentation, security, human factors

Keywords

Embedded training, phishing, email, usable privacy and security, real-world studies

Symposium on Usable Privacy and Security (SOUPS) 2009, July 15–17, 2009, Mountain View, CA USA
.

1. INTRODUCTION

PhishGuru is an embedded training system that teaches users to avoid falling for phishing attacks by sending them simulated phishing emails. These emails deliver a training message when the user falls for the attack and clicks on the simulated phishing URL, thus taking advantage of a "teachable moment." PhishGuru emails might be sent by a corporate system administrator, ISP, or training company. The training materials present the user with a comic strip that defines phishing, offers steps to follow to avoid falling for phishing attacks, and illustrates how easy it is for criminals to perpetrate such attacks.

Our prior studies tested users immediately and one week after the training, and demonstrated that PhishGuru improved users' ability to identify phishing emails and websites [9, 10, 12]. Training systems should be designed not only to convey knowledge, but also to help learners retain that knowledge for the long term [18]. In this study, we extend our previous work by presenting the results of a 515-participant, real-world experiment in which we measured long-term retention. In addition, while our previous studies focused on testing a single-training intervention, our embedded training approach allows for convenient, ongoing training. In this study we measure the effect of using a second training message to reinforce the original training. We also address some of the limitations of earlier laboratory [10] and real-world [12] PhishGuru studies.

Each simulated phishing email acts not only as a mechanism to deliver training, but also as a test of whether the recipient has learned how to distinguish legitimate from phishing messages. A real deployment of the system would not only train users, but also assess their performance at regular intervals. In this way, we can identify and present training interventions only to those users who continue to fall for simulated phishing attacks. In addition, this approach can be used to introduce recipients to new phishing threats over time and focus on those recipients who are most susceptible to the new threats. The issues of long term retention and repeated training interventions are essential to the validity and effectiveness of such long-term training and evaluation campaigns. If the training does not result in long term retention, such a deployment would require frequent training interventions, which could annoy users and even counter the effectiveness of the training. Similarly, if additional training interventions do not increase performance, the validity of a system that repeatedly trains users who continue to make mistakes is certainly called into question. The results from this study indicate that people trained with PhishGuru do

retain what they learned in the long term and that multiple training interventions increase performance.

Our study participants were Carnegie Mellon University (CMU) faculty, staff, and students. The simulated phishing emails we created were all spear-phishing emails targeted at the CMU community. Our results demonstrate that PhishGuru effectively trains users in the real world, and that people who were trained through PhishGuru retained this knowledge for at least 28 days. Results also show that people who were trained twice were significantly less likely to provide information to the simulated phishing web pages when tested 2 days, a week, and 2 weeks after training. We also found that training with PhishGuru does not increase the likelihood of false positive errors (participants identifying legitimate emails as phishing emails).

The large size and duration of this study allowed us to draw some conclusions about susceptibility to phishing based on certain demographic factors. As in the previous studies [4, 19], we did not observe a difference in susceptibility to phishing attacks with respect to gender. However we found that age is a factor in phishing susceptibility, as participants in the 18-25 age group were more likely to fall for phishing than those in older age groups.

The remainder of the paper is organized as follows: In the next section we relate phishing to relevant studies in deception theory, and we discuss related experimental studies on phishing. In Section 3, we present the study setup, participant demographics, and hypotheses. In Section 4, we present the results of our evaluation, demonstrating that PhishGuru effectively educates people in the real world. In Section 5, we present the challenges of conducting a field trial to study the effectiveness of phishing interventions and the ways in which we addressed them. Finally, in Section 6, we discuss the effect of training people in the real world.

2. BACKGROUND

In this section we present a brief background on phishing and highlight some lessons that can be learned from deception theory literature. We also describe some results from related empirical studies on phishing.

2.1 Deception theory

The Internet and other technological advancements have lowered the cost of perpetrating large-scale crimes. Recently, a dramatic increase has been observed in attacks known as "phishing," in which spoofed emails and fraudulent websites mislead victims, causing them to reveal private and potentially valuable information. Victims perceive that these emails are associated with a trusted brand, while in reality they are the work of con artists attempting to commit identity theft [13]. Phishers exploit the difference between the system model and the users' mental model to deceive and victimize users [14].

Psychologists and communication researchers have studied deception in detail. Deception is generally defined as "a message knowingly transmitted by a sender to foster a false belief or conclusion by the receiver" [2]. Communication literature suggests that many cues influence users when making trust decisions, including (1) verbal cues (*e.g.* language style, message content in the email); (2) non-verbal cues (*e.g.* time an email is received); and (3) contextual cues (*e.g.* feedback from toolbars) [3]. Studies have also shown that people fall for phishing attacks because many of the cues that people rely on can be easily spoofed by the phisher to deceive the victim [4].

Jonhson *et al.* have developed a generic model that can be used to detect deception by using the cues available in a given situation [8]. Grazioli adapted the model to detect deception over the Internet [21]. The model decomposes the action of detecting deception to (1) activation (allocating attention to cues, based on the presence of discrepancies between what is observed and what is expected, *e.g.* the information in the current email versus what is expected from the given sender); (2) hypothesis generation (generating hypothesis(es) to explain the next steps in the situation, *e.g.* "because there was some illegitimate access to my account, they want me to update my personal information"); (3) hypothesis testing (evaluating the hypotheses that were created *e.g.* "if I click on the link and the resulting website looks legitimate then it must be a legitimate email"); and (4) global assessment (making a decision on the given situation, *e.g.* a user decides that this is a legitimate website and provides personal information to the website). Researchers also propose computer awareness and training as a solution to prevent people from being deceived through computers [21]. One of the goals of anti-phishing work is to develop tools to educate users so that they are able to generate and test hypotheses properly and not be deceived.

2.2 Related work

There are only a few published real-world studies that evaluate the effectiveness of anti-phishing training. The idea of sending fake phishing emails to test users' vulnerability [6, 7, 17] and evaluate the effectiveness of training delivered through other channels has been explored by researchers. Jagatic et al. studied the vulnerability of a university community towards a phishing email that pretends to come from somebody in their own social network, but did not study the effectiveness of training [7]. Researchers at West Point [6] and at the New York State Office of Cyber Security [17] conducted this type of study in two testing phases. Both studies showed an improvement in the participants' ability to identify phishing emails. Recently, the United States Department of Justice sent their employees fake phishing emails to test their vulnerability to phishing [20]. Sheng *et al.* have shown that people can be trained about phishing URLs through an online game called Anti-Phishing Phil [19]. They found the game to be effective in both a laboratory setting and in the real world [11, 19].

None of this previous research considers the question of how a user's behavior changes over time as a result of training. In our work, we send 7 simulated phishing emails to users over the course of 28 days. The long duration of our study allows us to focus on long term retention and the effect of providing more opportunities to learn.

Learning science literature shows that training is most effective when the training materials are presented in a real-world context [1]. Additionally, researchers have shown that providing immediate feedback during the learning phase results in more efficient learning [18]. One of our previous laboratory studies provides strong evidence that people make better decisions when they go through PhishGuru training than when they receive security notices emails [9]. We have also shown that people retain and transfer more knowledge when trained with embedded training than with non-embedded training [10]. Our previous work also suggests

that PhishGuru can effectively train employees in a real-world setting [12]. However, these studies don't address the primary foci of this paper: long term retention and reinforcement through additional training.

3. EVALUATION

In this section we present our participant demographics, methodology, and hypotheses.

3.1 Recruitment and demographics

We sent a recruitment email to all active CMU student, faculty, and staff Andrew email accounts[1] with the primary campus affiliation listed as "Pittsburgh." The email subject line read "Volunteers Needed: Help Us Protect the Carnegie Mellon Community from Identity Theft" and the email content described both what would be required of participants and what data would be collected from them. In addition, they were told that volunteers would be entered into a raffle to receive one of five $75 gift cards. Willing participants were instructed to reply to the recruitment email or go to a web link to opt in to the study. We also added "To verify the authenticity of this message, visit the ISO Security News & Events at https://www.cmu.edu/iso" in the email so that users could check the legitimacy of the message.[2] In total we sent 21,351 emails and recruited 515 volunteers. The CMU human resources department provided us with demographic information about each participant, summarized in Table 1.

Every person in the university is assigned a primary department, even if they are students with double majors or faculty with joint appointments. For the purpose of this study and our analysis, we looked only at primary departments. We grouped the 26 departments into 7 academic department clusters and 3 non-academic department clusters as shown in Table 1. For example, we grouped the Entertainment Technology Center and School of Computer Science together as Computer Science.

3.2 Study setup

Five hundred and fifteen participants were randomly assigned to *control*, *single-training*, and *multiple-training* conditions. There were 172 participants in control, 172 in single-training, and 171 in multiple-training. All participants, regardless of condition, were sent a series of 3 legitimate and 7 simulated spear-phishing emails over the course of 28 days, as shown in Table 2. In the body of each phishing email was a simulated phishing URL. Clicking on this link resulted in different scenarios depending on the study day and the participant's condition. Participants in the single-training condition who clicked the URL on day 0, and those in the multiple-training condition who clicked the URL on day 0 and/or day 14, saw one or both (one on each day) of the anti-phishing training interventions depicted in Figure 1. For all other study days in the single-training and multiple-training conditions, clicking on the URL led to a simulated phishing webpage where an HTML form asked users to provide private credentials. Participants in the control condition did not receive any anti-phishing training as part of the study. When they clicked on the URLs they were directed to simulated phishing webpages. We tested participants twice after each training email to test their immediate retention (2 days) and short-term retention (7 days).

Table 3 presents an overview of the 7 simulated phishing emails sent to participants. Except for the "Community Service" email—which proved to be a much less effective phishing lure than the other messages—we found no difference in the rate at which participants fell for each of the emails on day 0. However, to ensure that the aggregate response rates per day were not confounded by the potential differ-

Table 1: Percentage of people in the three conditions and percentage of people who fell on day 0, in each demographic (N = 515).

	% of control	% of single-training	% of multiple-training	% who fell for day 0 phish
Gender				
Female	44.8	48.8	39.8	48.5
Male	55.2	51.2	60.2	50.7
Affiliation				
Faculty	7.0	8.7	7.0	38.5
Staff	36.0	38.4	30.4	37.8
Students	56.4	52.9	62.6	58.6
Sponsored	0.6	0	0	0
Student year				
Doctoral	13.4	17.5	12.3	52.7
Masters	19.8	19.8	21.7	56.2
Undergraduate	20.9	18.6	28.0	62.9
Miscellaneous	2.3	1.1	0	66.7
None	43.6	43.0	38.0	37.9
Department type				
Academic	72.7	73.9	78.4	53.1
Administrative	24.4	24.4	19.3	39.3
Unknown	2.9	1.7	2.3	41.7
Academic departments				
IS and Public Policy	8.7	12.2	12.8	50
Humanities & Social Sciencs	7.6	8.7	8.1	59.5
Engineering	16.3	14.5	14.6	57.7
Fine Arts	4.6	6.4	3.5	48
Computer Science	16.3	14.5	18.7	48.2
Business	8.7	5.8	10.5	51.2
Sciences	10.5	11.6	11.1	52.6
Non-academic departments				
Computing Services and Research	5.8	5.8	5.2	34.5
Administration	18.6	18.0	13.6	41.2
Other	2.9	2.3	1.8	50

[1]The Andrew account is the main email account given to all CMU community members.

[2]ISO is the CMU Information Security Office

Table 2: Schedule of the emails, including day of study, calendar date (2008), and type of emails sent out that day. For example, on day 0 we sent test and legitimate emails to all participants.

Study day	Day 0	Day 2	Day 7	Day 14	Day 16	Day 21	Day 28	Day 35
Date	Nov 10	Nov 12	Nov 17	Nov 24	Nov 26	Dec 1	Dec 8	Dec 15
Type of Emails Sent	Train and test, then legitimate	Test	Test, then legitimate	Train and test	Test	Test	Test, then legitimate	Post-study survey

ence in natural response rates for individual emails or by the interdependence of response rates among the emails, we developed a counterbalancing schedule. The counterbalancing schedule avoided these confounding issues by dividing the 515 participants randomly and equally per condition among 21 different viewing schedules for the 7 emails. The critical property of the 21 schedules was that, for any given day of the study, each of the 7 emails was sent out to an equal number of participants. This allowed us to compute the aggregate response rate for an entire day by summing the responses to each of the emails sent that day. Since the proportions were constant for all study days, different aggregate response rates across different days were comparable. To counterbalance the training materials, half of the participants in the single-training condition received intervention A and the other half received intervention B. Similarly, in the multiple-training condition, half of the participants received intervention A first and intervention B second and the other half received intervention B first and intervention A second. We found no significant difference in response rates among participants who received the training materials in different orders or among those who received different training material.

All emails that we constructed for the study were emails that the CMU community might normally receive, though they were not based on any information that a phisher would not be able to obtain from public webpages. From the email messages that participants sent us to sign up for the study, we determined that a large fraction of participants use an email client that might strip the HTML from the message body. Therefore, we did not replicate the common phishing tactic of using HTML to hide phishing URLs from users. All of our phishing messages displayed the phishing URLs in the body of the messages. Figure 2 (Top) shows an example of an email that was used in the study. This example asks the study participant to click on the link to change their Andrew password.

We registered all of the domain names we used in our simulated phishing emails using legitimate credentials—that is, a query to the associated "whois" database would show valid CMU affiliated contact information. In this way, if participants were skilled enough, they could easily infer that these domains were part of the study. Besides those shown in Table 3, we also registered another 10 similar-looking domains as backup.

Figure 2 (Middle) shows one of the simulated phishing websites. This example simulates the standard password change scenario at CMU. The site asks participants to provide their User ID, old password, and new password, and then to confirm their new password. All of the websites used in the study similarly collected some combination of user name and password. When participants submitted their information, they were taken to a "thank you" page, as shown in Figure 2 (Bottom). Participants saw a similar sequence of webpages ("login" followed by "thank you") in all email scenarios.

To estimate the false positive rate, we measured the response rate to three legitimate emails sent to study participants by the CMU Information Security Office (ISO). These messages were sent to all participants on day 0, day 7, and day 28 after the test/training emails were sent. The original recruitment email for this study was presented in the context of Cyber Security Awareness Month. The three legitimate emails were announcements for an ongoing security related scavenger hunt, begun during Cyber Security Awareness Month, which gave community members an opportunity to gain points in return for specified security related tasks. The subject line of the first email was "Earn Bonus Points #1: Win a Nintendo Wii, $250 Amazon Gift Card or other great prizes." The second and third emails had identical subjects, except that they were emails "#2" and "#3," respectively. The email itself indicated that the recipient needed to login with their Andrew password to claim their bonus points. Clicking the link took them to the real "webiso login page" (the standard login page for all CMU websites—the one that we spoofed in our phishing websites) where they were asked to provide their username and password.

So that we could track user responses, each participant was given a unique 4-character alpha-numeric hash that was appended as a parameter to the URL of all emails that participants received (*e.g.* in one email, participant 9009 received a URL that ended with `update.htm?ID=9009`). The hash also served as mechanism to allow us to protect the identity of participants during data analysis. To ensure that no sensitive data would be compromised, ISO did a complete penetration test on the machine that was used to host the phishing websites. In addition, the simulated phishing webpages were constructed so that no information was ever submitted to the webserver. Using JavaScript, all of the form data that the user submitted was discarded prior to form submission. To ensure that the emails were not blocked by CMU spam filters, the machine from which the emails were sent was put on a white list.

After all real and simulated phishing emails were sent, another email was sent to all participants asking them to complete a post-study survey. The survey consisted of questions regarding (1) the interest level of participants in receiving such training in the future, (2) participants' feedback on the training methodology, (3) participants' feedback on the interventions and instructions, (4) whether participants remembered registering for the study, and (5) demographic information such as age. Two hundred and seventy nine participants completed the post-study survey. These participants were distributed nearly equally across our three con-

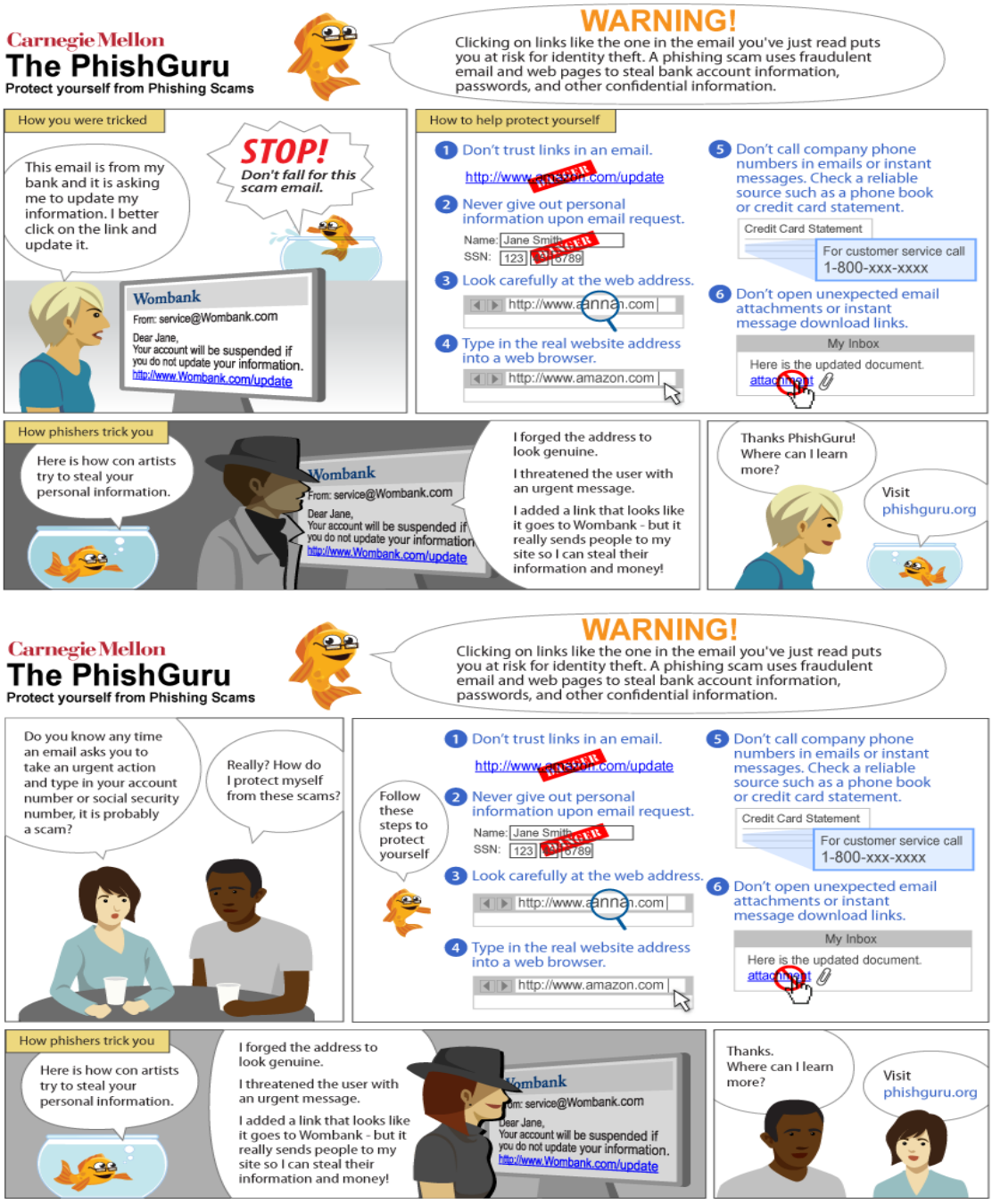

Figure 1: Above: Intervention A. One of the two training interventions used in the study. Half of the participants in the single-training and multiple-training conditions received this training intervention on day 0. The other half of the multiple-training condition received this on day 14. Below: Intervention B. The second training intervention used in the study. The instructions are the same as in Intervention A, but the characters and the story are slightly different. Half of the participants in single-training and multiple-training conditions received this training intervention on day 0. The other half of the multiple-training condition received this on day 14.

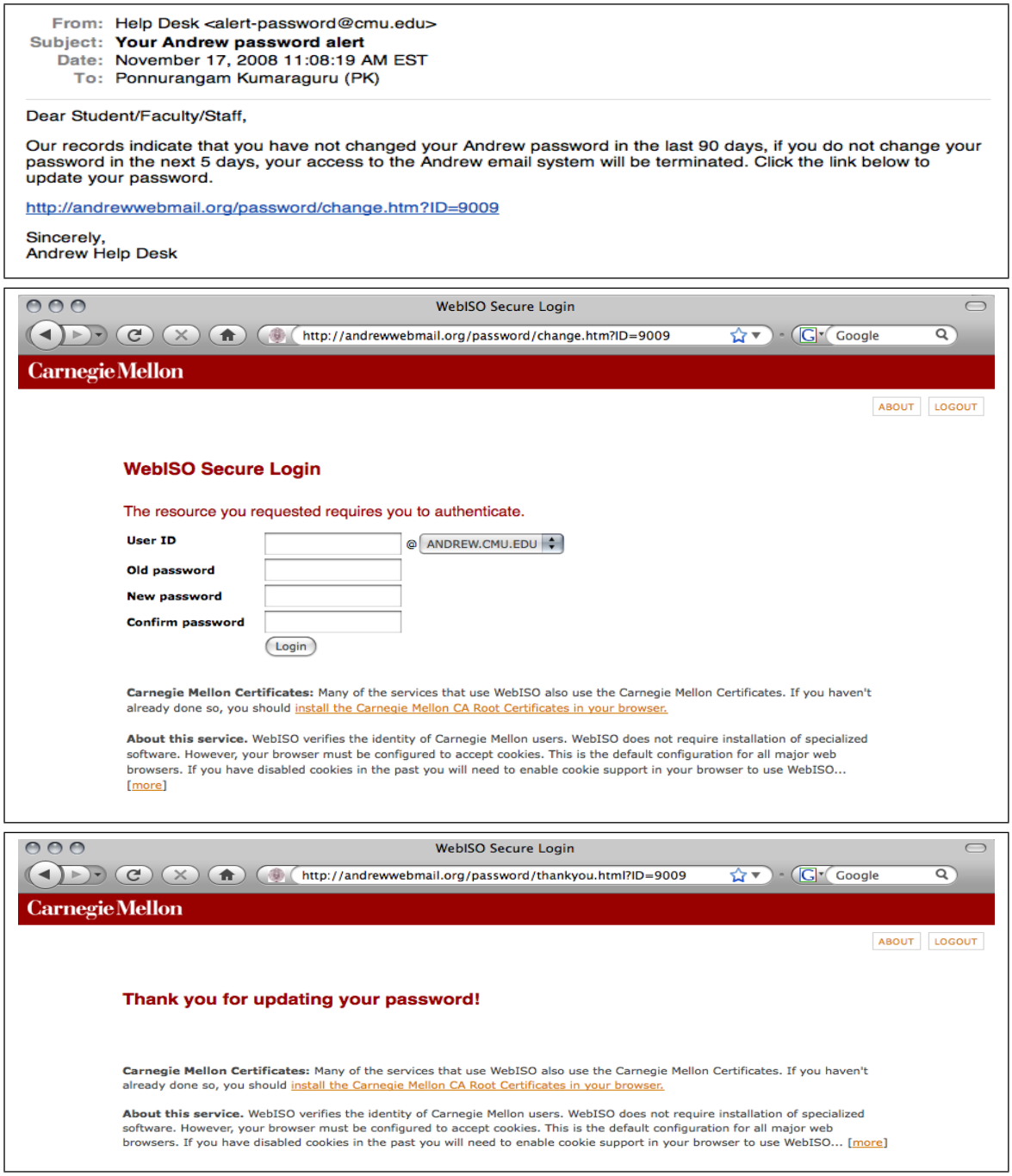

Figure 2: A sample of simulated phishing emails and websites. Top: A sample of the simulated phishing emails used in the study. The URL that appears in the email matches the target of the HREF statement. Middle: One of the seven simulated websites. Using JavaScript, all of the form data that the user submits was discarded prior to form submission. Bottom: "Thank you" webpage that was shown to the users when they gave credentials on the webpage presented in Middle. Similar pages were presented for other simulated websites.

Table 3: Summary of emails sent to study participants. In all emails, when the user clicked on the link in the email, she was taken to a page where her user name and password was requested. The "Bandwidth Quota Offer" email gave users an opportunity to increase their daily wireless bandwidth limit. The "Plaid Ca$h" email contained instructions to claim $100 in Plaid Ca$h (money to be used at CMU vendors). The remaining emails are sufficiently explained by the subject line. The legitimate email had "https" while all others had "http" in the URL.

Email type	From	Subject line	Domain name in URL
Test/Train	Info Sec <infosec@andrew.cmu.edu>	Bandwidth Quota Offer	cmubandwithamnesty.org
Test/Train	Networking Services <rec-networking@andrew.cmu.edu>	Register for Carnegie Mellon's annual networking event	carnegiemellonnetworking.org
Test/Train	Webmaster <webmaster@andrew.cmu.edu>	Change Andrew password	andrewpasswordexpiry.org
Test/Train	The Hub - Enrollment Services <thehub@andrew.cmu.edu>	Congratulation - Plaid Ca$h	idcardsforcmu.org
Test/Train	Sophie Jones <shjones@andrew.cmu.edu>	Please register for the conference	studenteventsatcmu.org
Test/Train	Community Service <community@andrew.cmu.edu>	Volunteer at Community Service Links	communityservicelinks.org
Test/Train	Help Desk <alert-password@cmu.edu>	Your Andrew password alert	andrewwebmail.org
Legitimate	Information Security Office <iso@andrew.cmu.edu>	Earn Bonus Points #1: Win a Nintendo Wii, $250 Amazon Gift Card or other great prizes	cmu.edu

ditions (control = 31.5%; single-training = 34.0%; multiple-training = 34.5%).

3.3 Hypotheses

In our previous work, we showed that people who were trained by PhishGuru, both in a laboratory setting [9, 10] and in real-world settings [12], effectively retained the knowledge they gained for a short period. Our goal in this study was to investigate whether PhishGuru helps people retain long term knowledge about phishing. In particular our aim was to study retention after 28 days.

Hypothesis 1: *Participants in the training conditions (single-training and multiple-training) identify phishing emails better than those in the control condition on every day except day 0.*

Our earlier studies only tested the effectiveness of the training methodology when participants were trained once, but learning science literature suggests that if people are provided with more opportunities to learn, they tend to remember instructions better [5]. In PhishGuru, the simulated email works for both training and testing purposes; people who continue to click on the simulated phishing URLs can be presented with further training materials. Our goal was to investigate whether participants who read the training materials twice had any advantage over participants who read the training materials only once.

Hypothesis 2: *Participants who see the training interventions twice perform better than participants who see the intervention once.*

Our earlier studies did not provide any conclusive evidence for whether training has any effect on false positive errors [12]. We believe that it is very important to consider this criterion when measuring training success. In this study we sent legitimate emails to participants on day 0, day 7, and day 28 to measure the false positive error rate.

Hypothesis 3: *When asked to identify legitimate emails participants who view the training materials in the training conditions will perform the same as participants in the control condition.*

4. RESULTS

Our results support Hypotheses 1, 2, and 3.

4.1 H1: Long-term retention

Our results show that people in the single-training and multiple-training training conditions who fell for our first phishing message performed significantly better when they received our second phishing message than those in the control condition. In addition, we observed no significant loss in retention after 28 days. Table 4 presents the percentage of participants who clicked and gave information on day 0 through day 28. Approximately 52.3% (90 participants) in control, 51.7% (89 participants) in single-training and 45.0% (77 participants) in multiple-training conditions clicked on the link in the email that they received on day 0. We found no significant difference among the click rates of participants across the three conditions on day 0 (ANOVA, $F(2,512) = 1.1$, p-value = 0.3). This implies that prior to any influence from the study, participants in all three conditions were similar. We also found no significant difference (ANOVA, $F(6,1203) = 1.7$, p-value = 0.3) in the click rate of participants in the control group across study days (day 0 until day 28). This implies that there was no change in the behavior of participants in the control group throughout the study.

On day 0, 48.4% of the participants in the training conditions viewed the PhishGuru intervention. To determine the effectiveness of the training, we conditioned the click rates of days 2 through 28 on those participants across all conditions

Table 4: Percentage of participants who clicked and gave information on days 0 through 28. N is the number of participants in each condition. Participants in the training conditions saw the interventions on day 0 and therefore had no opportunity to give information. We found no significant differences among the click rates of participants across the three conditions on day 0 and among participants in the control group on all days.

Conditions	N	Day 0		Day 2		Day 7		Day 14		Day 16		Day 21		Day 28	
		Click	Gave	Click	Gave	Click	Gave	Click	Gave	Click	Gave	Click	Gave	Click	Gave
Control	172	52.3	40.1	51.2	39.5	48.3	40.7	54.1	41.3	44.12	30.8	41.3	25.0	44.2	30.8
single-training	172	51.7	NA	35.5	29.1	34.9	26.7	35.5	25.0	23.8	19.2	29.7	22.1	23.8	17.4
multiple-training	171	45.0	NA	31.6	23.9	30.4	21.6	37.4	NA	29.2	21.6	26.9	18.1	25.6	17.5

who clicked the links in the email(s) on day 0. This way we could compare the participants who actually received the training in the single-training and multiple-training conditions to those in the control condition who took the analogous action on day 0. Figure 3 (Left) shows the percentage of these participants who clicked on links in emails and gave information to the fake phishing websites from day 2 until day 28. There is a significant difference (Chi-Sq = 14, p-value < 0.001) between the percentage of users who clicked in the control condition (54.4%) and the percentage who clicked in the single-training (27.0%) on day 28. Similarly, there is significant difference between the control and multiple-training (32.5%) conditions on day 28 (Chi-Sq = 8.9, p-value < 0.01). We also find that, in the single-training condition, participants who gave information to fake phishing websites on day 2 are not significantly different than on day 28 (Chi-Sq = 3.5, p-value < 0.1). Similarly, there is significant difference between the control and single-training and between the control and multiple-training conditions in the percentage of people who clicked on days 2 through 28. This shows that users trained with PhishGuru retain knowledge even after 28 days. This supports Hypothesis 1.

4.2 H2: Multiple training

Our results strongly suggest that users who saw the training intervention twice were less likely to give information to the fake phishing websites than those who only saw the training intervention once. Figure 3 (Right) shows the percentage of participants who clicked on links in emails from day 16 until day 28 conditioned on participants who clicked on the link on day 0 and those who clicked on day 14. There is a significant difference (Chi-Sq = 5.4, p-value =0.01) between the percentages of users who clicked in the single-training condition (42.9%) and those who clicked in the multiple-training (26.5%) on day 16 and a similar difference on day 21 (Chi-Sq = 7.8, p-value < 0.01). However, we did not find a significant difference between users who clicked in the single-training and multiple-training conditions on day 28 (Chi-Sq = 0.3, p-value =0.6). We also did not find any significant difference (Chi-Sq = 1.1, p-value = 0.3) in clicking between day 21 (26.5%) and day 28 (35.3%) in the multiple-training condition.

Figure 3 (Right) also shows that participants who were trained twice are doing significantly better than people who were trained once when it comes to giving their personal information to fake phishing websites. For example, on day 28, 31.4% of the participants in the single-training condition gave information to the website, while only 14.7% did in the multiple-training condition. This is significantly different (Chi-Sq = 7.3, p-value < 0.01). These results support Hypothesis 2.

We also found 30 participants (17.5%) in the multiple-training condition who did not see the intervention on day 0 but saw the intervention on day 14. These are the people who probably needed training, since they fell for the email on day 14. We saw no significant difference (t-test, t = 0.1, p-value = 0.8) between people in the single-training condition who clicked on day 14 but were trained on day 0 and people in the multiple-training condition who clicked on day 28 but were trained only on day 14. This suggests that multiple rounds of training is useful not only for re-inforcement but also for providing an additional opportunity for people who need training.

4.3 H3: Legitimate emails

Results from this study indicate that training users to recognize phishing emails using PhishGuru does not make them more likely to identify legitimate emails as phishing emails. Table 5 presents the percentage of participants who clicked and gave information in response to legitimate emails out of those participants who clicked on day 0. We found no significant difference between the three conditions on day 0 (ANOVA, F(2,512) = 2.7, p-value = 0.1) and on day 28 (ANOVA, F(2,512) = 1.2, p-value = 0.3). We also did not find any significant difference within the conditions among the three different emails (control - ANOVA, F(2,513) = 1.9, p-value = 0.2; single-training - ANOVA, F(2,513) = 1.7, p-value = 0.2; multiple-training - ANOVA, F(2,510) = 2.7, p-value = 0.1). This shows that user behavior did not change with respect to the legitimate emails that were tracked as part of the study, confirming that training people does not decrease their willingness to click on links in legitimate email messages. This result supports Hypothesis 3.

4.4 Analysis based on demographics

Multivariate regression analysis did not find any significant relationship between susceptibility to phishing on day 0 and gender (p-value = 0.9 for gender coefficient), student year (p-value = 0.5 for student year coefficient), or department (p-value = 0.8 for department coefficient). We did, however, find significant difference in the affiliation. In particular, we found significant difference (Std. error = 0.2, p-value < 0.05) between students and staff in falling for phishing on day 0. We found that students are more vulnerable to phishing emails before receiving any training from the study. We also found significant difference in the department type (different from primary department). In particular we found significant difference (Std. error = 0.2, p-value

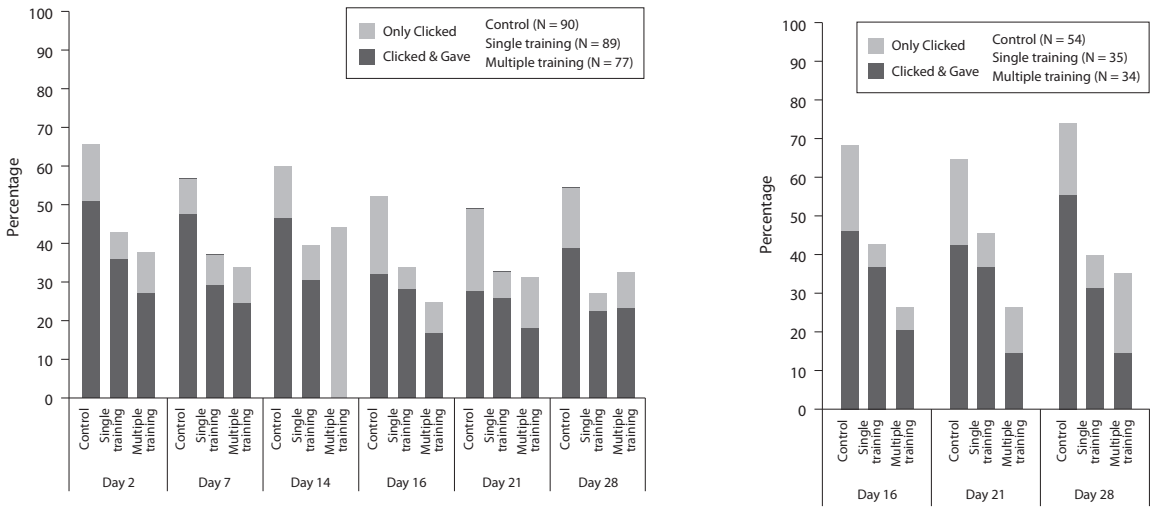

Figure 3: Percentage of participants who clicked on phishing links and gave information. Left: Days 2 through 28 conditioned on those participants who clicked the link on day 0. N is the number of people who clicked on day 0. Nobody gave information in the multiple-training condition on day 14 because it was a training email. There is significant difference between the control and single-training and between the control and multiple-training conditions in the percentage of people who clicked on days 2 through 28. Right: Days 16 through 28 conditioned on those participants who clicked on both day 0 and day 14. N is the number of people who clicked on day 0 and on day 14. There is significant difference between the single-training and multiple-training conditions in the percentage of people who gave information to phishing sites on days 16 through 28.

< 0.05) between the academic and administrative department types, with academics being more susceptible to falling for the phishing email. Investigating this further, we found that the difference could be attributed to the fact that all students are in the academic department type, making this group as a whole more vulnerable than others.

We investigated this difference between students and staff further to see if age was a factor in susceptibility to phishing. We used the age data collected through post-study surveys. Two hundred and sixty-seven participants provided their age in the survey. The minimum age in years was 18 and the maximum age was 77 (avg. = 32.3, SD = 12.8). We found a significant difference (Chi-Sq = 8, p-value < 0.01) in the likelihood of clicking on links on day 0 between age group 18 - 25 and those in all of the older age groups (Shown in Table 6). This shows that, prior to any training, those participants in the 18-25 age group are more likely to click on the links in the phishing emails than any other age group.

Among the participants who were trained on day 0, again, multivariate regression analysis did not find any significant relationship between susceptibility to phishing on day 28 and gender (p-value = 0.4 for gender coefficient), student year (p-value = 0.9 for student year coefficient), and department (p-value = 0.7 for department coefficient). We did find difference (Std. error = 0.3, = p-value < 0.001) between the academic and administrative department types, which was again attributable to students falling for phishing after training. Similar to day 0, on day 28 we found that the age group 18 - 25 was significantly (Chi-Sq = 10.5, p-value < 0.01) more likely to fall for phishing than other age groups (Table 6). We found that participants in the 18-25 age group were consistently more vulnerable to phishing attacks on all days of the study than older participants. These results are in line with risk averse literature, which says that younger people are more risk taking and impulsive, while older people are risk averse and less impulsive [16]. We were not able draw any concrete conclusions about faculty because the sample sizes were too small.

Table 5: Percentage of participants who clicked and gave information in response to the legitimate emails out of those participants who clicked on day 0. N is the number of participants in each condition. There is no significant difference between the three conditions on any given day.

Condition	N	Day 0		Day 7		Day 28	
		Click	Gave	Click	Gave	Click	Gave
Control	90	50.0	42.2	41.1	37.8	38.9	35.6
single-training	89	39.3	38.2	42.7	37.1	32.3	30.3
multiple-training	77	48.1	36.3	44.2	36.4	35.1	32.5

Computer savvy technical people (Software Engineering Institute, Computing Services) were less likely that others to fall for phish. In general, however, participants in our Computer Science and Computing Services and Research department clusters did not perform significantly different than participants in any other group on day 0.

Table 6: Percentage of participants who clicked on the link in the emails by age group. N = 267 people responded to the post-study survey with their age. This shows that age group 18 - 25 behaves in a significantly different way from all of the other age groups.

Age group	Day 0	Day 28
18 - 25	62.3	35.7
26 - 35	47.5	15.8
36 - 45	33.3	18.2
46 and more	42.5	10

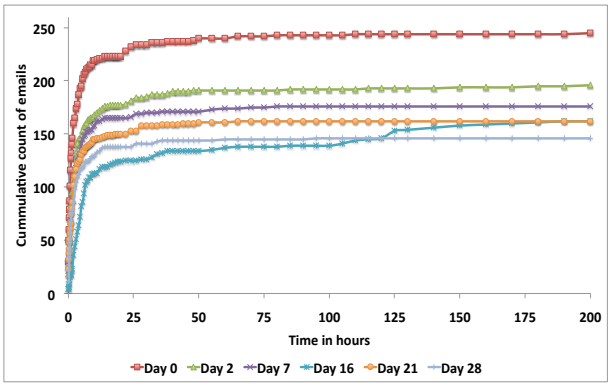

Figure 4: Cumulative number of emails that were clicked since the email was sent out. This shows that study participants who click on the links in emails will do so within 8 hours of the time the email was sent out. Because of a technical error, we were not able to capture the data for day 14. The day 16 time-window spans the Thanksgiving holiday, and the second peak coincides with the Monday after Thanksgiving.

4.5 Observations

In this section we describe the data that we collected in the study and through the post-study survey, as well as other observations from the data that we collected.

Our results indicate that most participants who will eventually click on the link in an email will do so within 8 hours from the time that the email is sent. To estimate the distribution of how long people took to read emails, we used the time at which a participant clicked on the phishing link as a proxy for the time the email was read. Figure 4 presents the cumulative number of emails that were clicked on for each study day since the study email was sent out. This shows that, 2 hours after the emails were sent, at least half of the people who will eventually click on the link have already done so; after 8 hours, nearly all people (90%) who will click have already done so. This suggests that anti-phishing methods that rely on black-lists should aim to update their lists before this window has passed; otherwise, users will click on the link and become a victim for phishing. This further supports the effectiveness of methodologies such as PhishGuru that work from the start of a phishing attack.

Some of the post-study survey questions were designed to gauge the receptiveness of the participants to PhishGuru training. Eighty percent of participants liked the idea of conducting such campus studies at regular intervals and ninety percent would recommend this type of training to a friend. One participant wrote, "I really like this study, and we should have this kind of program every year to increase the awareness." Another wrote, "This should be one of the first things that incoming CMU students learn." Some participants liked the idea of being reminded of the instructions periodically. One participant wrote, "It is always good to be reminded. Sometimes you forget, so I think getting reminders once a month is a good way of helping us to remember." We were also interested in finding out how often the emails should be sent to the participants. We asked, "How often would you like to receive educational materials like this picture(s) in your email inbox?" Eighty five participants responded to the question. Forty percent answered "Once a month," while 22.3% said they never want to see such training emails.

When asked to give an open-ended comment about the study, one of the participants said "One thing I did not like about the study is that I was tricked by one email that was part of the study, but I had to call to be reassured that I did not have to change my Andrew password." Since we were working with the ISO team, they presented a canned response to inquiries from participants. We believe this mitigated potential backlash to the study. We also believe that, when it comes to training emails, participants who click on the link should be quickly and courteously alerted to the fact that they have been tricked. We incorporated such friendly alerts into the training messages. In the case of testing emails, it is important to debrief people about the study and provide them with opportunities to give their feedback. In our study, we debriefed participants through an email and plan to conduct a university-wide presentation about the results.

Unlike in our previous PhishGuru field study [12], we found little interaction between participants discussing the study. Only 13% of participants indicated that they had talked about the tips presented in the PhishGuru training with other members of the CMU community in the prior 30 days. Six of the participants who said they had discussed the training provided information about their discussions. A typical response was: "Just talked about the fact that I fell for one scam that offered $100 prize" or "I did talk about how I was tricked VERY easily into giving away my username/password to my andrew account." To further understand potential contamination across study conditions, we asked "How did you get to see the picture(s)?" in the post-study. Of those who responded, 87% reported seeing the training cartoons through a link in an email from the study. Only 5% reported seeing the training through a link in an email that was forwarded by a friend or a colleague at CMU, and 5% reported that a friend or a colleague at CMU showed them the training. The remaining participants said they couldn't remember how they they got to the training. These results show that most of the participants received the training material through the emails sent through the study; therefore, there was little chance for interaction among participants regarding the study, and so little chance of the conditions being contaminated.

5. CHALLENGES IN ADMINISTERING REAL-WORLD PHISHING STUDIES

We have taken measures in this study to address many lessons that we learned from earlier work. Real-world studies can provide more ecological validity and richer data than laboratory studies, but are often difficult to conduct. The challenges we faced included making sure our study emails reached participants' inboxes, maintaining participants' privacy, avoiding contamination between study conditions, and coordinating with relevant third parties.

Simulated emails may get deleted before they reach the user's inbox if, for instance, filters determine that the message is Spam. Additionally, since many web-browsers often come equipped with anti-phishing tools, researchers must be careful that the study material isn't blocked. In particular, researchers should be aware of the possibility that study websites might end up on a black-list. To be prepared for problems of this nature, we registered multiple dummy domains and prepared multiple sets of emails as backup. Furthermore, since email reading behavior may be different over university holidays than it is during the regular semester, we carefully timed our study schedule so that our study emails were not sent during university holidays.

In order to maintain the privacy of the participants, study administrators should not/cannot collect any personal information. Furthermore, to understand the users' behavior over time, users' responses must be tracked in a way that respects their privacy. We accomplished this in the study by assigning an anonymous hash to each participant, tracking each participant only through the hash.

To avoid subject contamination, study designers should try to minimize the chance that participants in different conditions will interact with each other; such interactions may invalidate the study data. Working to prevent these interactions, study designers must ensure that the study sample is embedded within a large, geographically separate population. In our previous field study, significant contamination occurred because study participants all worked on one floor of an office building [12]. In our current study, even though all participants were from the same university campus, they represented a small fraction of the campus population and were spread across 26 departments and many buildings, which limited contamination.

It is important to coordinate with any relevant third parties that might be affected by the study. We worked very closely with ISO in both the design and implementation stages of this study. In addition, ISO aided us in getting permission from the Institutional Review Board (IRB), in coordinating with campus help desks, and in getting permission from all the campus offices spoofed in the study. As a courtesy and to minimize accidental external interference in the study, researchers should work with system administrators and help desk officials of the organization to inform them about the study. If possible, researchers should also provide system administrators with a "canned" response which they can use to respond to any inquires from participants. This helps minimize the chance that system administrators will send an email to the entire population warning them to avoid opening an email that was actually part of the study (we have seen this happen in a prior study). Finally, it is essential that any university phishing study go through the university's IRB. Having a well defined plan to address the challenges we mentioned here could help prevent potential difficulties in the review process.

6. DISCUSSION

In this paper, we investigated the effectiveness of an embedded training methodology called PhishGuru that teaches people about phishing during their normal use of email. We showed that, even 28 days after training, users trained by PhishGuru were less likely to click on the link in a simulated phishing email than those who were not trained. Furthermore, users who saw the training intervention twice were less likely to give information to fake phishing websites than those who only saw the training intervention once. Additionally, results from this study indicate that training users to recognize phishing emails using PhishGuru does not increase their concern towards email in general or cause them to make more false positive mistakes. Another surprising result was that around 90% of the participants who eventually clicked on the link in an email did so within 8 hours of the time the email was sent. We believe this behavior generalizes to other university populations, though non-university populations may behave quite differently when reading emails. In analyzing the demographics, our results showed that younger people (in the 18-25 age group) were more prone to falling for phishing emails consistently on all days of the study than older participants. This suggests a need for: (1) training before college; and (2) training that specifically targets high school and college students.

The study presented in this paper addresses some of the limitations of earlier laboratory [10] and real-world [12] studies of PhishGuru. To address these limitations, we employed a larger sample size, extended the study duration, counterbalanced the email and training interventions, minimized the chance of contamination from participants talking about the study amongst themselves, and provided good incentives for participants to complete the post-study survey. In the process of addressing these limitations, we successfully showed that PhishGuru can be deployed both on a large scale and in the real world as an embedded training system where users can be educated about phishing during their regular use of email. This study included only a small fraction of our campus population due to IRB requirements that participants opt in to the study before receiving any study emails. However, if this deployment had been done as a real training exercise—that is, without an academic IRB requirement—we believe it would have been easy to train the entire campus with only minimal changes to the study setup.

This study affirms prior research [10] suggesting that the PhishGuru methodology is an unobtrusive way to train users about phishing. Some comments from the post-study survey include: (1) "I really liked the idea of sending CMU students fake phishing emails and then saying to them, essentially, HEY! You could've just gotten scammed! You should be more careful – here's how...." (2) "I think the idea of using something fun, like a cartoon, to teach people about a serious subject is awesome!."

Furthermore, the fact that knowledge gained from the training materials is retained for at least 28 days suggests that very frequent interventions, which could annoy users, are not necessary. In practice, this should be balanced with the fact that repeated training does improve user performance; a proper trade-off between usability and accuracy can and should be optimized.

In addition to increasing user awareness about phishing emails, there was evidence that the study had the unintended consequence of assessing both the users' awareness of proper response channels for phishing attacks and the ability of ISO to react to phishing attacks. Many users properly contacted the ISO help desk to alert them of the emails, either by phone or through the official email address. However, some were apparently unaware of ISO's role in protecting the campus, and instead contacted some other "trusted source" like a professor or departmental system administrator to seek advice. This suggests that ISO may want to explore ways to increase awareness of the proper channels for reporting phishing attacks and other cyber security related issues. In a real deployment of PhishGuru, training interventions could be one way to distribute this information to the public.

This study is proof that it is possible to effectively educate users about security in the real world and on a large scale. Our findings suggest that security researchers and practitioners should implement user training as a complementary strategy to other technological solutions for security problems.

7. ACKNOWLEDGMENTS

This work was supported by the National Science Foundation grant number CCF-0524189, Army Research Office grant number DAAD19-02-1-0389, Fundaçã paraa Ciĕnciae Tecnologia (FCT) Portugal under a grant from the Information and Communications Technology Institute (ICTI) at CMU. The authors also thank all CMU faculty, staff, and students who took part in the study. The authors would like to thank all members of the Supporting Trust Decisions project, Dr. Vincent Aleven, Julian Cantella, and the ISO team at CMU.

8. REFERENCES

[1] J. R. Anderson and H. A. Simon. Situated learning and education. *Educational Researcher*, 25:5–11, 1996.

[2] D. B. Buller and J. K. Burgoon. Interpersonal deception theory. *Communication Theory*, 6(3):203 – 242, 1996.

[3] J. R. Carlson, J. F. George, J. K. Burgoon, M. Adkins, and C. H. White. Deception in computer-mediated communication. *Group Decision and Negotiation*, 13(1):5 – 28, 2004.

[4] R. Dhamija, J. D. Tygar, and M. Hearst. Why Phishing Works. *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, 2006.

[5] E. Ellis, L. Worthington, and M. Larkin. Research synthesis on effective teaching principles and the design of quality tools for educators. Technical report, National center to improve the tools of educators, 1994.

[6] A. J. Ferguson. Fostering E-Mail Security Awareness: The West Point Carronade. *EDUCASE Quarterly*, (1), 2005.

[7] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, 50(10):94–100, October 2007.

[8] P. Johnson, S. Grazioli, K. Jamal, and G. Berryman. Detecting deception: adversarial problem solving in a low base-rate world. *Cognitive Science: A Multidisciplinary Journal*, 25(3):355 – 392, 2001.

[9] P. Kumaraguru, Y. Rhee, A. Acquisti, L. F. Cranor, J. Hong, and E. Nunge. Protecting people from phishing: the design and evaluation of an embedded training email system. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 905–914, 2007.

[10] P. Kumaraguru, Y. Rhee, S. Sheng, S. Hasan, A. Acquisti, L. F. Cranor, and J. Hong. Getting users to pay attention to anti-phishing education: Evaluation of retention and transfer. *e-Crime Researchers Summit, Anti-Phishing Working Group*, 2007.

[11] P. Kumaraguru, S. Sheng, A. Acquisti, L. Cranor, and J. Hong. Under review.

[12] P. Kumaraguru, S. Sheng, A. Acquisti, L. F. Cranor, and J. Hong. Lessons from a real world evaluation of anti-phishing training. *e-Crime Researchers Summit, Anti-Phishing Working Group*, October 2008.

[13] R. Lininger and R. D. Vines. *Phishing: Cutting the Identity Theft Line*. Indianapolis, Indiana, USA, 2005.

[14] R. C. Miller and M. Wu. Fighting Phishing at the User Interface. *O'Reilly*, August 2005. In Lorrie Cranor and Simson Garfinkel (Eds.) Security and Usability: Designing Secure Systems that People Can Use.

[15] T. Moore and R. Anderson. How brain type influences online safety. Working paper, July 2008.

[16] R. A. Morin and A. Fernandez Suarez. Risk aversion revisited. *Journal of Finance*, 38(4):1201–16, September 1983.

[17] 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., 2005.

[18] R. A. Schmidt and R. A. Bjork. New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3(4):207–217, July 1992.

[19] S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L. F. Cranor, J. Hong, and E. Nunge. Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish. In *SOUPS '07: Proceedings of the 3rd symposium on Usable privacy and security*, pages 88–99, 2007.

[20] E. Spagat. Justice department hoaxes employees. News article, January 2009. http://news.yahoo.com/s/ap/20090129/ap_on_go_ca_st_pe/justice_hoax.

[21] G. Stefano. Where did they go wrong? an analysis of the failure of knowledgeable internet consumers to detect deception over the internet. *Group Decision and Negotiation*, 13:149 – 172, March 2004.