



wombatsecurity.com

helping organizations combat phishing

An Empirical Evaluation of PhishGuru™ Embedded Training

Wombat Security Technologies

<http://wombatsecurity.com/>

April 2009

Wombat's scientific studies have demonstrated that employees are not motivated to pay attention to traditional cyber security training emails or read online training materials. However, by taking advantage of the teachable moment that occurs when someone believes they just fell for an attack (in reality an attack we generated), we are able to capture peoples' attention. Our PhishGuru™ cartoons provide training in a succinct and engaging format, developed using learning science principles. Our studies demonstrate people can learn important cyber security concepts from PhishGuru™ with only a couple of minutes of training, and can remember what they learned and apply it a month later.

Executive Summary

Wombat's PhishGuru™ embedded training system 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 that we delivered to users' inboxes. We validated the effectiveness of this approach in a series of laboratory and real-world empirical studies. We report here on two studies conducted at the CyLab Usable Privacy and Security Laboratory at Carnegie Mellon University,¹ as well as on a real-world study conducted with Carnegie Mellon students, faculty, and staff.

Key Findings

- **PhishGuru™ training succeeds where traditional email training fails.** A single round of PhishGuru™ embedded training reduced the number of lab study participants falling for phishing emails by 60%. The same training message delivered directly via email was completely ineffective.
- **PhishGuru™ training is fast.** Lab study participants spent an average of two minutes reading the training cartoon.
- **PhishGuru™ training is effective in the real world.** Participants who fell for a phishing email and received PhishGuru™ training in our real-world study were 50% less likely to fall for subsequent phishing emails than those who fell for the first phishing email but did not receive training.
- **Users remember PhishGuru™ training.** Users trained with PhishGuru™ in our real-world study were just as likely to be able to protect themselves from phishing 28 days after training as they were two days after training.
- **Multiple rounds of PhishGuru™ training are even more effective than a single round of training.** A second training message reinforces the original training and provides a second training opportunity. Twice-trained users in our real-world study were almost 50% less likely to provide their credentials to phishing websites than users trained only once. In addition, 18% of users were trained by the second PhishGuru™ message after not seeing the first training opportunity (because they did not fall for the first phishing message).
- **PhishGuru™ training does not decrease users' willingness to click on links in legitimate messages.** While many other forms of training increase users' suspicion of all email, PhishGuru™ teaches users techniques they can use successfully to distinguish legitimate emails from phishing messages.
- **People trained with PhishGuru™ enjoy this form of training and want to receive more of it.** 80% of study participants would recommended that their organization continue providing PhishGuru™ training.


Participants who fell for a phishing email and received PhishGuru™ training were 50% less likely to fall for subsequent phishing emails than those who did not receive training.

¹ Wombat Security Technologies was founded to commercialize products originally developed at Carnegie Mellon University as part of one of the largest anti-phishing research projects in the US.

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 an embedded training message when the user falls for the attack and clicks on the simulated phishing URL, thus taking advantage of a “teachable moment.” The training materials present the user with a succinct and engaging 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. Usable security experts developed these training materials using principles from learning science. Figure 1 shows an example of a PhishGuru™ training cartoon.


We conducted a series of laboratory and real-world studies to evaluate the effectiveness of PhishGuru™ training. This paper provides an overview of the methodology and results of these studies. Further details about these studies are available in the referenced publications.



PhishGuru™
Protect yourself from Phishing Scams

WARNING

Clicking on links like the one in the email you've just read puts you at risk for identity theft. A phishing scam uses fraudulent email and web pages to steal bank account information, passwords, and other confidential information.



How you were tricked:

This email is from my bank and it is asking me to update my information. I better click on the link and update it.

STOP!
Don't fall for this scam email.

Wombank
From: service@Wombank.com
Dear Jane,
Your account will be suspended if you do not update your information.
<http://www.Wombank.com/update>

How to help protect yourself:

- Don't trust links in an email.
<http://www.Wombank.com/update>
- Never give out personal information upon email request.
Name: Jane Smith
SSN: 123-456-7899
- Look carefully at the web address.
<http://www.Wombank.com>
- Type in the real website address into a web browser.
<http://www.Wombank.com>
- Don't call company phone numbers in emails or instant messages. Check a reliable source such as a phone book or credit card statement.
Credit Card Statement
For customer service call 1-800-xxx-xxxx
- Don't open unexpected email attachments or instant message download links.
My Inbox
Here is the updated document.
[attachment](#)

How phishers trick you:

Here is how con artists try to steal your personal information.

Wombank
From: service@Wombank.com
Dear Jane,
Your account will be suspended if you do not update your information.
<http://www.Wombank.com/update>

I forged the address to look genuine.

I threatened the user with an urgent message.

I added a link that looks like it goes to Wombank - but it really sends people to my site so I can steal their information and money!

Thanks PhishGuru! Where can I learn more?

Visit PhishGuru.org

To learn more about protecting yourself from phishing scams visit <http://www.PhishGuru.org>

(c) 2008 Carnegie Mellon University

Figure 1. A PhishGuru™ training cartoon.

Laboratory study 1: Evaluation of Knowledge Acquisition

Our first laboratory study was designed to compare the effectiveness of PhishGuru™ embedded training cartoons with two other types of email-based training.²

Study design

We recruited 30 participants who were not computer security experts to participate in a study at the CyLab Usable Privacy and Security Laboratory at Carnegie Mellon University. Each participant was randomly placed in one of three conditions: notices, text/graphic, and comic. The user study consisted of a think-aloud session in which participants played the role of “Bobby Smith,” an employee of Cognix Inc. who works in the marketing department. Participants were told that the study investigated “how people effectively manage and use email.” They were told that they should interact with Bobby Smith’s email the way they would normally interact with their own email.

Each participant was shown 19 email messages. Nine of these messages were legitimate emails (from co-workers at Cognix, friends, and family) that requested that Bobby perform simple tasks and reply by email. Two messages were simulated legitimate emails containing links from organizations with which Bobby had an account. The mailbox also contained two spam emails, two fraudulent emails that appeared to come from organizations where Bobby had an account (which we refer to these as “phishing-account” emails), and two fraudulent emails that appeared to come from a bank with which Bobby did not have an account. The mailbox also had two training emails. Participants in the notices condition received training in the form of a typical email-based security notice sent by a major company to its customers. The training emails for participants in the other two conditions were phishing-account emails that included embedded training. When participants clicked on a link and fell for the phish they were directed to a website with a training intervention. In the text/graphics condition the training intervention was a single page of text and graphics that defined phishing and provided tips for detecting phishing messages. In the comic condition the intervention was an early version of a PhishGuru™ cartoon.

Our results suggest that the current practice of emailing out security notices is ineffective.

Results

In this study we measured knowledge acquisition based on whether or not users clicked on links in legitimate emails and phishing emails before and after training. We found statistically significant differences in knowledge acquisition between the three study conditions, with participants in the comic condition demonstrating the greatest knowledge acquisition. Participants in the comic condition also spent the

² P. Kumaraguru, Y. Rhee, A. Acquisti, L. Cranor, J. Hong, and E. Nunge. Protectin People from Phishing: The Design and Evaluation of an Embedded Training Email System. In *CHI 2007: Conference on Human Factors in Computing Systems*, San Jose, California, 28 April – May 3, 2007, p. 905-914. <http://doi.acm.org/10.1145/1240624.1240760>

most time reading training materials. Average time spent reading training materials was 9 seconds for the notice condition, 107 seconds for the text/graphics condition, and 120 seconds for the comic condition.

	Comic condition	Text/graphics condition	Security notices condition
Participants who fell for phishing-account email <i>before</i> training	100%	80%	90%
Participants who fell for phishing-account email <i>after</i> training	30%	70%	90%
Average time spent reading training materials	120 seconds	107 seconds	9 seconds

In the security notice condition ninety percent of participants fell for the phishing-account email before training, and we saw no improvement after training. The participants who had seen the security notices said that the information took too long to read and they were not sure what the messages were trying to convey.

Participants in the comic condition were significantly better at recognizing phishing emails than those in the notices condition.³ Participants in the text/graphics group also performed better than those in the notices condition, but this difference was not significant. In the text/graphic condition, 80% of the participants fell for the first phishing-account email and 70% fell for the final phishing-account email after training. In the comic condition, all participants fell for the first phishing-account email and only 30% fell for the final phishing-account email after training.

When study participants fell for our simulated phishing attacks they were motivated to spend time reading training materials.

Conclusions

Our results suggest that the current practice of emailing out security notices is ineffective because people are unwilling to spend time reading these notices. On the other hand, our comic strip embedded training is an effective way to teach people how to avoid falling for phishing attacks. When study participants fell for our simulated phishing attacks they were motivated to spend time reading training materials.

³ Chi-square was used to test for statistical significance in this study. P values <0.01 are reported as significant.

Laboratory study 2: Evaluation of Knowledge Retention

In our first laboratory study we tested users immediately after training. In our second laboratory study we examine (1) how well PhishGuru™ users can retain and transfer knowledge, and (2) how important it is to fall for simulated phishing attacks, as opposed to simply receiving PhishGuru™ interventions directly as email messages.⁴

Study design

Our second study used a similar design as the previous study, once again asking participants to play the role of Bobby Smith. We had four conditions: embedded, non-embedded, suspicion, and control. Participants in the embedded condition received a simulated phishing email and saw a revised version of the comic strip intervention when they clicked on a link in that email. Participants in the non-embedded condition received the same comic strip directly as part of an email message, without having to fall for a simulated phishing email. Participants in the suspicion condition received a brief email from a friend that mentioned phishing, without providing any information about how they could protect themselves. Participants in the control condition received an additional email from a friend, but received no training.

The second study was conducted in two parts, seven days apart. In the first part, participants saw 33 emails in Bobby's inbox: a set of 16 emails, a training intervention, and a set of 16 additional emails shown immediately after training. In the second part, the participants saw another 16 emails in Bobby's inbox.

Results

Our results provide evidence that embedded training enhances knowledge acquisition, knowledge retention, and knowledge transfer, allowing learners to effectively identify phishing messages without misidentifying legitimate messages.

We measured the percentage of correct decisions that participants in each condition made for phishing and legitimate emails before and after the training, as shown in Figure 2. Our results demonstrate that participants in the embedded training condition learned to detect phishing-account emails effectively while participants in the other conditions did not. While participants did not perform significantly differently in correctly identifying phishing-account emails before the training, those in the embedded condition performed significantly better than those in the other

Users in the embedded condition were able to retain the knowledge they acquired and use it to distinguish phishing and legitimate emails, even after a one-week delay.

⁴ P. Kumaraguru, Y. Rhee, A. Acquisti, L. Cranor, J. Hong, and E. Nunge. Protecting People from Phishing: The Design and Evaluation of an Embedded Training Email System In *CHI 2007: Conference on Human Factors in Computing Systems*, San Jose, California, 28 April – May 3, 2007, p. 905-914. <http://doi.acm.org/10.1145/1240624.1240760>

Participants in the embedded condition spent more than twice as much time reading the cartoon.

conditions immediately after training.⁵ Indeed, those in the non-embedded condition did not perform significantly differently than those in the control condition immediately after training. This may be due in part to the fact that participants in the embedded condition were motivated to spend more than twice as much time reading the cartoon than those in the non-embedded condition. Participants in the embedded condition spent an average of 97 seconds reading the cartoon, while participants in the non-embedded condition spent an average of 37 seconds.

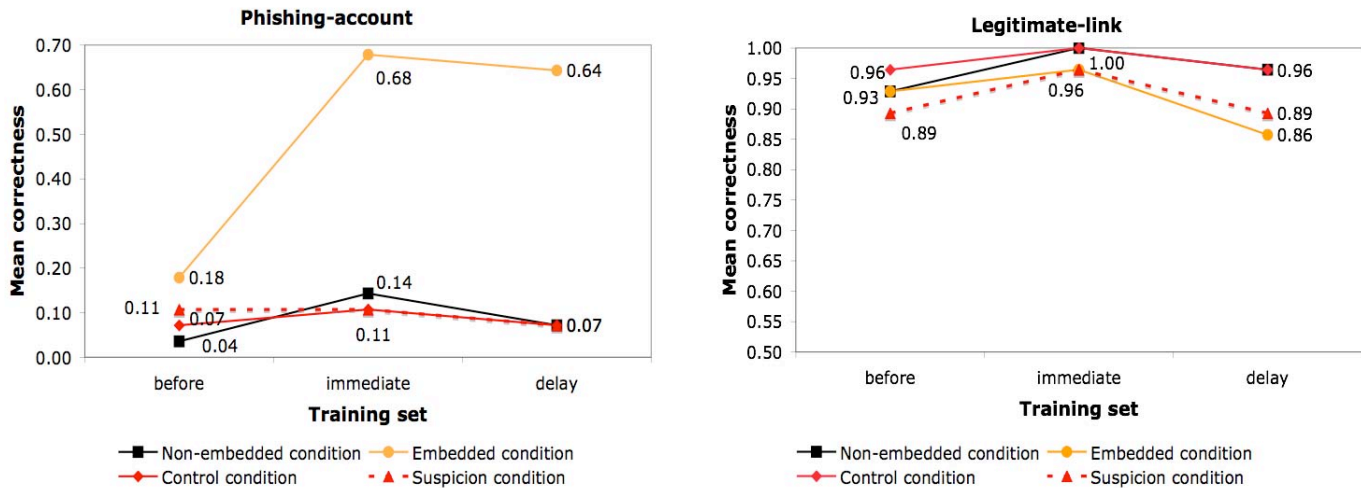


Figure 2. Mean correctness for identifying phishing-account and legitimate emails before training, immediately after training, and after a one-week delay. On the left, we can see that in the embedded condition, there is a sizeable improvement in participants’ abilities to detect phishing emails after having been trained with PhishGuru™. In contrast, there is no statistically significant improvement in the other conditions. On the right, we see that there are no statistically significant differences in the number of people clicking on legitimate links over time or across conditions. This indicates that PhishGuru™ training does not increase false positives.

Training is completely ineffective when sent directly via email, without leveraging the teachable moment after people fall for an attack.

To measure knowledge retention, we compared performance on phishing-account and legitimate emails before, immediately after training, and after a one-week delay. Our results suggest that users in the embedded condition were able to retain the knowledge they acquired and use it to distinguish phishing and legitimate emails, even after a one-week delay. In all conditions there was no significant difference between mean correctness on phishing-account emails or legitimate emails immediately after the training and after a one-week delay. Participants in the embedded condition improved their performance significantly on phishing-account emails after the delay compared to before the training, while participants in the other conditions did not improve. While 64% of the embedded-condition participants

⁵ Two-sample and paired t-tests were used to test for statistical significance in this study. P values <0.01 are reported as significant.

identified the phishing-account email correctly after a one-week delay, only 7% of the participants in the other conditions identified the email correctly.

Conclusions

Our results reinforce and extend the findings of our previous study, which suggested that embedded training can be an effective method to train users to distinguish between legitimate and phishing email messages. The fact that we saw no significant performance drop-off after one week suggests that users are likely to retain their training for longer time periods. Our observation that the suspicion condition was not significantly different from the control condition suggests that it is not helpful to tell users about phishing without providing them with information about how to identify phishing or actionable steps they can take to protect themselves.

Finally, our observation that the non-embedded condition was not significantly different from the control condition provides a strong indicator that the delivery method is an important factor in determining anti-phishing training effectiveness. Indeed, we found that our revised comic strip intervention, which provided effective training when displayed at that teachable moment after participants had fallen for a simulated phishing attack, was completely ineffective when sent directly via email.

It is not helpful to tell users about phishing without providing them with information about how to identify phishing or actionable steps they can take to protect themselves.

Real-world study

To evaluate PhishGuru™ in the real world, we conducted a 515-participant, real-world experiment in which we measured long-term retention over a 28-day period.⁶

Study design

Study participants were faculty, staff, and students from throughout the Carnegie Mellon University (CMU) community. The simulated phishing emails we created were all spear-phishing emails targeted at the CMU community.

Five hundred and fifteen participants were randomly assigned to three conditions: control, single training, and multiple training. All participants, regardless of condition, were sent a series of three legitimate and seven simulated spear-phishing emails over the course of 28 days (on days 0, 2, 7, 14, 16, 21, and 28). In the body of each spear-phishing email was a simulated phishing URL that directed participants to a simulated phishing web site that requested the private credentials necessary to login to CMU websites. Participants in the single- and multiple-training conditions who clicked the URL on Day 0 saw a PhishGuru™ training cartoon instead of a simulated phishing website. Participants in the multiple-training condition who clicked the URL on Day 14 also saw a PhishGuru™ training cartoon (the second cartoon contained the same training content as the first, but included different characters and a slightly

⁶ P. Kumaraguru, J. Cranshaw, A. Acquisti, L. Cranor, J. Hong, M.A. Blair, T. Pham. School of Phish: A Real-World Evaluation of Anti-Phishing Training. CyLab Technical Report CMU-CyLab-09-002, 2009. http://www.cylab.cmu.edu/research/techreports/tr_cylab09002.html

different story line). Participants in the control condition did not receive any anti-phishing training as part of the study.

We developed seven plain-text, spear-phishing emails with subject lines relating to password changes, bandwidth quota, event registration, prizes, and volunteering for community service. These were all 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 web pages. All of our phishing messages displayed the phishing URLs in the body of the messages. We did not replicate the common phishing tactic of using HTML to hide phishing URLs from users.

To ensure that the aggregate response rates per day were not confounded by the potential difference 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.

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) on Day 0, Day 7, and Day 28 after the test/training emails were sent. The three legitimate emails were announcements for an ongoing security-related scavenger hunt begun during Cyber Security Awareness Month that gave community members an opportunity to gain points in return for specified security-related tasks. The emails indicated that the recipient needed to login with their password to claim their bonus points. Clicking the link took them to a real CMU website 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 sent to participants. 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. 279 of our participants completed the post-study survey. These participants were distributed nearly equally across our three conditions.

Results

Our results show that people in the single- and multiple-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. We found no significant differences among the click rate of participants across the three conditions on day 0 or in the click rate of participants in the control group across study days.⁷

Users trained with PhishGuru™ retain knowledge even after 28 days.

⁷ ANOVA was used to test for statistical significance. P values <0.01 are reported as significant.

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 who clicked the links on day 0. This way we could compare the participants who actually received the training in the single- 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 difference⁸ between the percentage of users who clicked in the control condition (54.4%) and the percentage who clicked in the single-training condition (27.0%) on day 28. Similarly, there is a significant difference between the control and multiple-training (32.5%) conditions on day 28. 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. This shows that users trained with PhishGuru™ retain knowledge even after 28 days.

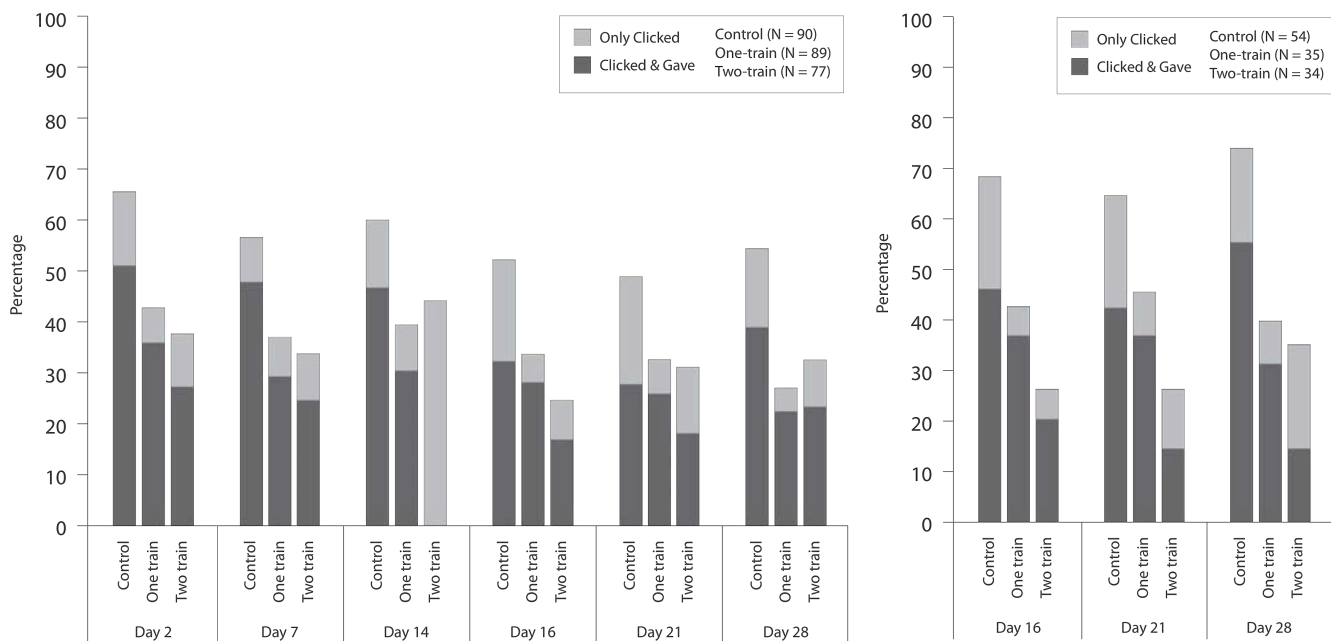


Figure 3. Percentage of participants who clicked on phishing links and gave information. Left: Days 2 through 28 conditioned on those participants who clicked on day 0. N is the number of people who clicked on day 0. There is significant difference between the control and single-training condition (one-train)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

⁸ Chi-square was used to test for statistical significance. P values <0.01 are reported as significant.

Multi-round training is useful not only for re-enforcement, but also for providing an additional opportunity for people who need training.

We found 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 between the percentages of users who clicked in the single-training condition (42.9%) and those who clicked in the multiple-training condition (26.5%) on day 16 and a similar difference on day 21. However, we did not find a significant difference between users who clicked in the single-training and multiple-training conditions on day 28. Figure 3 (Right) also shows that participants who were trained twice are doing significantly better than participants 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% of participants in the multiple-training condition gave information.

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 people probably needed training, since they fell for the email on day 14. This suggests that multi-round training is useful not only for re-enforcement, but also for providing an additional opportunity for people who need training.

Our results indicate that training users to recognize phishing emails using PhishGuru does not make them more likely to identify legitimate emails as phishing emails. We found no significant difference between the three conditions in response rate to the legitimate emails on day 0 and on day 28.

Training with PhishGuru™ does not increase the likelihood of false positive errors.

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 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 our real-world study also allowed us to draw some conclusions about susceptibility to phishing based on certain demographic factors. Our results suggest that there is little or no 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.

In our post-study questionnaire 80% of participants said they would recommend that CMU continue providing Phishguru™ training. One participant wrote, “We should have this kind of program every year to increase the awareness.” Another wrote, “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....” Most participants liked the idea of multi-round training. One participant wrote, “I think getting reminders once a month is a good way of helping us to remember.” Participants commented that they enjoyed receiving training in

80% of participants said they would recommend that CMU continue providing PhishGuru™ training.

cartoon form: “I think the idea of using something fun, like a cartoon, to teach people about a serious subject is awesome!”

Conclusions

The CMU study demonstrated the effectiveness of the PhishGuru™ approach of using teachable moments combined with engaging and actionable training materials.

Our embedded training approach allows for convenient and fast ongoing training in which each simulated phishing email acts both as a mechanism to deliver training and as a test of whether the recipient has learned how to distinguish legitimate from phishing messages. A PhishGuru™ deployment not only trains users in about two minutes, but also assesses 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.

- **PhishGuru™ training is effective in the real world.** Participants who fell for a phishing email and received PhishGuru™ training were 50% less likely to fall for subsequent phishing emails than those who fell for the first phishing email but did not receive training.
- **Users remember PhishGuru™ training.** Users trained with PhishGuru™ were just as likely to be able to protect themselves from phishing 28 days after training as they were two days after training.
- **Multiple rounds of PhishGuru™ training are even more effective than a single round of training.** A second training message reinforces the original training and provides a second training opportunity. Twice-trained users were almost 50% less likely to provide their credentials to phishing websites than users trained only once. In addition, 18% of users were trained by the second PhishGuru™ message after missing the first training opportunity (because they did not fall for the first phishing message).
- **PhishGuru™ training does not decrease users’ willingness to click on links in legitimate messages.** While many other forms of training increase users’ suspicion of all email, PhishGuru™ teaches users techniques they can use successfully to distinguish legitimate emails from phishing messages.
- **People trained with PhishGuru™ enjoy this form of training and want to receive more of it.** 80% of study participants would recommended that their organization continue providing PhishGuru™ training.