Work 41 (2012) 3549-3552
DOI: 10.3233/WOR-2012-1054-3549
IOS Press

Training users to counteract phishing

Christopher B. Mayhorn[a,*] and Patrick G. Nyeste[a]
[a] *Department of Psychology, North Carolina State University, 640 Poe Hall, Campus Box 7650, Raleigh, NC, USA*

Abstract. Phishing is an increasingly more prevalent form of online, social engineered scams that escalate costs and risks to society year to year. This study demonstrates an association between anti-phishing training techniques used in previous research and individual differences which could affect phishing susceptibility. Results indicated that anti-phishing training in both a simple comic and more complex video game form is helpful in decreasing phishing susceptibility as measured by Miss rates for all individuals including college aged and computer savvy participants. Based on the results of the present study, implications for future efforts to combat phishing are discussed.

Keywords: Computer security, phishing, training

1. Introduction

A rampant and socially engineered tactic called "phishing" is a type of criminal behavior that many computer users are becoming intimately familiar with in terms of the costs associated with computer security. Phishing emails and scam sites prey upon the user's quick recognition of a trusted site with even large inconsistencies being ignored due to lack of attention. Results from a study by Dhamija, Tygar, and Hearst (2006), indicate that phishing websites fooled ninety percent of participants when the designs of the site closely mimicked the legitimate site.

Expanding on these previous efforts, the current study attempts to show an association between anti-phishing training techniques used in previous research and individual differences including: cognitive abilities (inhibition and working memory) and personality factors (Trust, Impulsivity, Computer Experience) which could affect phishing susceptibility.

2. Method

2.1. Design

The study utilized an experimental 2 (time tested: immediate vs. delayed) x 3 (training type: control vs. embedded vs. game plus embedded) mixed factorial design. The training variable was manipulated as a between-subjects variable while time tested was a within-subject variable.

[*] Corresponding author. E-mail: chris_mayhorn@ncsu.edu

2.2. Participants

Eighty-four participants (mean age = 19.5 years, SD = 2.3, range = 17-36) were recruited from the North Carolina State psychology participant pool for the main study. There were 28 participants assigned to each training group within constraints by randomizing the training group assignment. Sixty-five percent of the sample was female and self-report data regarding years of education completed indicated that participants had attended school for a mean of 12.7 years (SD = 1.9).

2.3. Procedure

Following completion of a consent form, participants were directed to a computer in the lab where they were asked to complete an online questionnaire that included a demographic survey as well as measures of trust (Joinson, 2007) and impulsivity (Kumaraguru et al, 2007a). Once the online questionnaire was completed, cognitive tests were administered in the lab. Participant's working memory was assessed using the Alphabet-span task (La Pointe & Engle, 1990). Then, a participant's ability to inhibit irrelevant information was ascertained using the Stroop test (Stroop, 1935).

Only the game training group had the Anti-Phishing Phil game training before the main experimental task. This game led the participant through two different training rounds with a teaching and then game section. Both the game and embedded training groups had training during the main experimental task with the use of comics that warned of the previous fake email stimuli.

For the main experimental task, the user role-played a friend named Bob Jones by viewing his email inbox and interacting with the email stimuli found there. These stimuli were presented on a desktop computer program, SuperLab, and participants made responses via the keyboard. Each stimulus was created using Adobe Photoshop with images and examples taken from The Anti-Phishing Work Group (2010), and PhishTank (2010). The email inbox included randomized email messages each with a link directing them to visit some corresponding website. Initially, the researcher read instructions to the participant. Participants used the Y and N keys to provide a "yes" or "no" response, respectively when they were tasked with answering whether they trusted each email.

During the first week, following training, the experimental task required participants to interact with 30 emails during the immediate assessment of training. Once the task was completed, the participant was scheduled to come back a week later for delayed assessment of training during which they encountered 40 emails (without any refresher training for any of the groups). Of the 40 emails encountered during the second week, 30 emails had been previously encountered during the first week to provide a measure of training retention. The remaining ten new emails were used as a measure of training transfer. In the first week, the experiment lasted a total of 45 minutes on average and the second week experiment lasted a total of 20 minutes on average. Once the second week role-play was completed, the participant was debriefed.

3. Results

3.1. Descriptive Metrics of Phishing Performance: Signal Detection Theory (SDT)

The benefit of users' training was assessed by determining how susceptible they were to phishing at two different stages: during the first week of training, and one week after the training. These data were used as metrics of training acquisition, retention, and transfer (i.e., how well the users absorbed and learned the material). Following initial training, performance on the main experimental task (Bob's email list) created a baseline of Signal Detection hits, misses, false alarms, and correct rejections based on susceptibility to the phishing attacks according to Sheng et al (2007). Sheng et al.

(2007) proposed that Hits should be defined as correctly identifying phishing emails as untrustworthy. Misses are defined as incorrectly identifying the phishing emails as trustworthy. False Alarms (FA) would be demonstrated by incorrectly identifying the real emails as untrustworthy. Correct Rejections (CR) would be correctly identifying the real emails as trustworthy. Giving the participants more examples of anti-phishing data to create a baseline before the training could introduce training effects. Thus, the baseline comparison to a control condition in the present study was done between-subjects.

Descriptive data representing phishing performance is shown in Table 1 illustrating Hit, Miss, False Alarm, and Correct Rejection aspects of SDT with the full sample (n = 84) by the time of data collection (1st vs. 2nd week). The first week SDT sample consists of 30 emails showing the mean, standard error, and standard deviation.

Table 1

Signal Detection Theory Descriptive Statistics for Phishing Performance

	1st Week			2nd Week[a]		
	M	SE	SD	M	SE	SD
Hit	8.52	.278	2.54	10.4	.462	4.23
Miss	6.48	.278	2.54	9.55	.461	4.22
False Alarm	6.24	.325	2.98	8.27	.512	4.69
Correct Rejection	8.76	.325	2.98	11.8	.512	4.69

	2nd Week (Retention)[b]			2nd Week (Transfer)[c]		
	M	SE	SD	M	SE	SD
Hit	7.81	.348	3.19	2.65	.146	1.34
Miss	7.19	.348	3.19	2.33	.146	1.34
False Alarm	5.62	.376	3.44	2.64	.168	1.54
Correct Rejection	9.39	.376	3.44	2.37	.168	1.54

[a] 40 emails in total were viewed the second week; [b] and [c] are split from this and analyzed separately.

[b] 30 1st week emails tested retention of training for all emails shown in the first week.

[c] 10 new emails tested transfer of training not shown in first week.

3.2. ANOVA Analyses

It was predicted that the levels of phishing susceptibility would vary by training type (embedded, game, and control). In theory, the embedded and game training types should lead to a decrease in getting phished over the third level of training type (control). The game training type would show a greater decrease in getting phished over the embedded gaming type. A 3 (training type) x 2 (time tested) factor repeated measures ANOVA was used to test the hypotheses. With an alpha level of .05, a significant main effect of training was found when comparing Miss scores (trusting phish emails) across training conditions, $F\ (2, 81) = 6.258, p = 0.003$. A Tukey HSD posthoc test showed a significant difference between embedded and control as well as game and control training groups. However, the difference between embedded and gaming was not significant. A significant main effect for phishing susceptibility difference between times tested was found within-subjects such that people were more likely to be phished in the 2nd week than in the first, $F\ (1, 81) = 6.998, p = 0.01$. The dependent variables tested in this hypothesis were between first week Miss and second week Miss retention (30 emails). The interaction between times tested and levels of training was not statistically significant.

4. Discussion

The current research contributes to the existing psychological literature on computer security and phishing in a number of ways. For instance, inferential results included various effects of training and individual differences. A positive effect of training was found in reducing phishing susceptibility (using Miss as the dependant variable) and increasing awareness of fake emails. The positive effect of both the embedded and game training groups continued into the second week compared to the control group when looking at the means even though the results were not statistically significant. In terms of how individual differences impact phishing, Stroop scores as a measure of ability to inhibit irrelevant information had a significant inverse relationship with phishing susceptibility in the embedded training group. As the ability to inhibit increased, phishing susceptibility decreased. Working memory capacity as measured by the Alphabet span task had a significant inverse relationship with phishing susceptibility (Miss rate) in the game training group. As working memory ability increased, phishing susceptibility decreased. Finally, anti-phishing training appears to be an excellent way of reducing phishing susceptibility (Miss rate) in terms of increasing skepticism towards fake emails amongst a wide variety of individual differences. However, the training might gain greater strength in those individuals with increased inhibitory and working memory ability.

References

[1] Adams, A., & Sasse, M. A. (1999). Users are not the enemy: Why users compromise computer security mechanisms and how to take remedial measures. Communications of the Association of Computing Machinery,42, 40-46.

[2] America Online and the National Cyber Security Alliance (2004). AOL/NCSA Online Safety Study. http://www.staysafeonline.info/news/safety_study_v04.pdf

[3] Gaudin, S. (2007). Human error more dangerous than hackers. TechWeb. http://www.techweb.com/showArticle.jhtml?articleID=197801676

[4] Gordon, L. A., Loeb, M.P., Lucyshyn, W., & Richardson, R. (2006). 2006 CSI/FBI computer crime and security survey. Baltimore, MD: Computer Security Institute.

[5] Hardee, J. B., West, R., & Mayhorn, C. B. (2006). To download or not to download: An examination of computer security decision-making. Association of Computing Machinery: Interactions, 13(3), 32-37.

[6] Sanders, M. S., & McCormick, E. J. (1993). Human Factors in Engineering and Design (7th Ed.). New York, NY: McGraw-Hill Inc.

[7] Schneier, B. (2000). Secrets and Lies: Digital Security in a Networked World. New York: Wiley & Sons.

[8] Smetters, D. K., & Grinter, R. E. (2002). Moving from the design of usable security technologies to the design of useful secure applications. Proceedings of the New Security Paradigms Workshop 02, pp. 82-89, ACM Press.

[9] West, R. T., Mayhorn, C.B., Hardee, J. B., & Mendel, J. (In press). The weakest link: A human factors perspective on user-security interaction. To appear in M. Gupta & R. Sharman (Eds.), Social and Human Elements of Information Security: Emerging Trends and Countermeasures. Idea Group, Inc.

[10] Whitten, A., & Tygar, J. D. (1999). Why Johnny can't encrypt: A usability evaluation of PGP 5.0. Proceedings of the 8th Usenix Security Symposium, pp. 169-184, Usenix Association.

[11] Wogalter, M. S. (2006). Handbook of Warnings. Mahwah, NJ: Lawrence Erlbaum Associates.

[12] Wogalter, M. S., & Mayhorn, C. B. (2008). Trusting the internet: Cues affecting perceived credibility. International Journal of Technology and Human Interaction, 4 (1), 76-94.

[13] Wogalter, M. S., & Mayhorn, C. B. (2005). Providing cognitive support with technology-based warning systems. Ergonomics, 48(5), 522-533.

[14] Yee, K.P. (2002). User interaction design for secure systems. Proceedings of the 4th International Conference on Information and Communications Security, Singapore, 9-12 December 2002, pp. 278–290. London: Springer-Verlag.

[15] Zurko, M. E., & Simon, R. T., (1997). User-Centered Security. New Security Paradigms Workshop, Lake Arrowhead, 17-20 September 1997, pp. 27-33. New York: ACM Press.

[16] Zurko, M. E., Simon, R., & Sanfilippo, A. (1999). A user-centered, modular authorization service built on an RBAC foundation. IEEE Symposium on Security and Privacy, Oakland, 9-12 May 1999, pp. 57-71. Oakland, CA: IEEE.