12-10-2016

Phishing Training: A Preliminary Look at the Effects of Different Types of Training

Shamya Karumbaiah
University of Massachusetts - Amherst, shamya@cs.umass.edu

Ryan T. Wright
University of Virginia, rtw2n@comm.virginia.edu

Alexandra Durcikova
University of Oklahoma, alex@ou.edu

Matthew L. Jensen
University of Oklahoma, mjensen@ou.edu

Recommended Citation

Karumbaiah, Shamya; Wright, Ryan T.; Durcikova, Alexandra; and Jensen, Matthew L., "Phishing Training: A Preliminary Look at the Effects of Different Types of Training" (2016). *WISP 2016 Proceedings*. 11.
http://aisel.aisnet.org/wisp2016/11

Phishing Training: A Preliminary Look at the Effects of Different Types of Training

Shamya Karumbaiah
College of Information and Computer Sciences, University of Massachusetts Amherst, USA
shamya@cs.umass.edu

Ryan T Wright
McIntire School of Commerce, University of Virginia, USA
rtw2n@comm.virginia.edu

Alexandra Durcikova
Price College of Business, University of Oklahoma, USA
alex@ou.edu

Matthew Jensen
Price College of Business, University of Oklahoma, USA
mjensen@ou.edu

ABSTRACT

In this paper, we present the preliminary results of an experiment conducted to observe the impact of the different training techniques to increase the likelihood of participants identifying and reporting phishing messages. Three different training approaches were used – general video/quiz training, just-in-time training with simulated phishing emails, and a leaderboard, which awarded users points for forwarding correct phishing messages and penalized them for incorrect ones. The experiment emulated a normal working day of an executive assistant of a manager in an organization. Each participant was expected to accomplish work tasks and respond to work-related emails while watching for and reporting phishing messages. We observed that both general training and the presence of a leaderboard decreased the propensity to click on a phishing message, while we found no effect for different types of just-in-time training.

Keywords – phishing training, socially engineered attacks, security, phishing, leaderboard, just in time, video training

INTRODUCTION

It's a nightmare scenario; your customers' personal and credit card information is put on illicit websites for sale. Such was the case with Target when its customers found their credit and debit card information was now for sale on the Dark Web (Winter 2014). The cause of this breach was an old but pernicious attack: phishing. Phishing is a socially engineered attack aimed at fraudulently acquiring sensitive information from a victim. A single weak link in an organization opens the way for the attackers to cause damages like identity theft, loss of intellectual property, financial loss, and denial of access (Hong 2012). The problem of phishing is now worse than ever. According to a recent Federal Bureau of Investigation report (McCabe 2016), there has been a 270% increase in successful phishing attacks since 2015. This is the highest increase in activity since phishing attacks have been measured. The danger these attacks pose and the increase in prevalence, call for improved strategies to combat phishing attacks which are more personalized and sophisticated now. As a first step towards mitigating these attacks, individuals and organizations need to train users on how to identify phishing messages that get past the automatic detection systems. Although most sophisticated organizations offer phishing training of some sort, there is no clear research that points to the exact efficacy of different types of phishing training in organizations. In this research, we study the impact of different types training that have been designed to help users detect and report phishing attacks.

This research-in-progress paper outlines the preliminary findings of three different types of training in an experiment. These are: 1) general video/quiz training, 2) mock phishing training with just-in-time training (JIT), and 3) a phishing leaderboard, which evaluates the correctness of reported phishing emails by an individual. Preliminary results of the training indicate that both the general training and the leaderboard decreased the likelihood of clicking on a message, but just-in-time training did not. This paper unfolds first by briefly describing the theoretical development and hypotheses. This is followed by a description of our methods. Finally, we provide preliminary results and discuss these results.

THEORETICAL DEVELOPMENT AND HYPOTHESES

In the current research, we aim to measure the impact of different types of training on the accuracy of the phishing reports (general phishing training, JIT training, and leaderboard). All of these interventions are aimed at improving an individual’s awareness and intention to report the messages. Specifically, reporting an email as phishing/legitimate is a binary classification task measured by the accuracy of detection. The two measurements of interest are – true positives and false positives. Past research (Jensen et al. 2011) suggests two aspects of importance in training programs aimed at improving an individual’s detection abilities. First is to set proper thresholds of known characteristics used for identification like inclusion of suspicious link and cues to indicate urgency. Second is to carefully add new diagnostic characteristics to broaden the detection like adding a request for private information. Phishing training can improve the thresholds and diagnostic characteristics individuals use to identify phishing messages.

Wright and colleagues (Wright and Marett 2010; Wright et al. 2014; Durcikova et al. 2015; Wright et al. 2016) have noted the benefits of training in reducing the phishing vulnerability in individuals. But when they are left alone, such passive training has been observed to have limitations in helping them identify phishing attacks (e.g., Wright and Marett 2010). However, there has been a dearth of literature on the efficacy of this training techniques in comparison with other techniques. We expect general training to improve the identification ability of an individual. Thus:

> *H1a: General training on what is a phishing message will decrease the likelihood of clicking on a phishing messages.*

> *H1b: General training on what is a phishing message will increase likelihood of a user correctly reporting a phishing message.*

Just in time (JIT) uses mock phishing messages sent to employees to provide active training (Duffy and Jonassen 1992). Researchers have illustrated the power of learning-by-doing through improved productivity gains as the learners gain experience and accumulate needed knowledge (Epple 1991). JIT is typically seen in organizations that execute mock phishing attacks on their own subjects. Companies like Phishme.com, Phishing Labs, and PhishTank offer services that create campaigns to launch against employees in an organization (Phishme 2013). Initial tests of JIT have shown promise in reducing susceptibility to phishing attacks, but it is unclear what type of JIT training is most effective (e.g., Kumaraguru et al. 2007). We compare two versions of JIT training. The lean JIT just informs the participant that it was a phishing message whereas the rich JIT provides a detailed training on identifying such a phishing message. Thus:

> *H2a: Rich Just-in-Time training will decrease the likelihood of clicking on a phishing messages more than lean Just-in-Time training.*

> *H2b: Rich Just-in-Time training will increase the likelihood of a user reporting a phishing message over lean Just-in-Time training.*

Leaderboards used for phishing is drawn from the concept of gamification (Deterding 2011). It involves creating systems that rewards participants with badges or points for correct reporting of phishing messages. While it is known to be motivational and enjoyable in education and employee training (Landers et al. 2011; Glover 2013), this type of extrinsic motivation is also observed to encourage knowledge sharing in general (Hung et al. 2011) and is here applied to sharing knowledge about possible phishing messages. Thus:

> *H3a: The presence of a leaderboard will decrease the likelihood of clicking on a phishing messages.*

H3b: The presence of a leaderboard will increase likelihood of a user reporting a phishing message.

METHOD

A 2 (general training: presence/absence) X 2 (JIT: rich/lean) X 2 (leaderboard: presence/absence) factorial experiment was conducted with 422 undergraduate students at a large Northeastern university in the USA. For those in the general training condition, we introduced participants to a high-quality video describing how to identify phishing messages and follow-up quiz that reinforced the phishing training. This type of training is deployed by most organizations (Puhakainen and Siponen 2010). The quiz asked subjects about phishing and provided the correct answers to reinforce the video's message. This video and subsequent quiz was given 10 days before the experiment session.

JIT follows a learning-by-doing approach in a context representative of the real world. We had mock phishing messages sent to the participants while at task. For the two flavors of the second condition, we varied the amount of information the participants would get in the training. The rich JIT condition includes a link in the phishing message, which directs participants to a webpage and demonstrates a 3-step procedure that could have helped identify the message as phishing (similar to the video training). The lean JIT condition directed participants to a simple webpage that just informed them that they clicked on a phishing link.

For the third treatment, we developed a phishing leaderboard that tracks when participants identify phishing messages correctly. For the groups with leaderboard, a score is maintained based on if they report the message as phishing. An algorithm was written to score the forwarded messages. Participants, identified by their userID, received +100 points for correct reports and -25 points for wrong reports. A web portal with real time listing of participant names and scores in descending order by score is projected at the front of the lab.

A group of 30 participants was randomly assigned to a session and was supervised by 2-3 lab assistants. Each session was provided with the same experimental task. There were a total of 17 sessions in this experiment. Ten days before their scheduled session, the participants were asked to fill a pre survey and were given (or not given) the general training. After the training, a survey captured propensity to trust, perceived internet risk, and self-reported identification abilities (Wright and Marett 2010).

In each session, participants were briefly introduced to the task and the leaderboard (if present in the condition). The task was designed to simulate a typical day in a knowledge workers job that included certain tasks and organizational priorities. The participants took the role of an executive assistant to the vice president (VP) of a tech company and were expected to manage the VPs email, forward personal messages, schedule appointments, and accomplish search tasks (e.g., a location for a large company meeting). They were detailed on email etiquette, priorities and constraints, along with the evaluation metrics for their tasks in the order of importance. Further, all students were provided a document that outlined the organization’s Information Technology (IT) security policy, which included asking subjects to forward all possible phishing emails to an address similar to phishing@company.com. Eight messages were prepopulated in their inboxes the previous night and the remaining 18 messages came during the experiment session.

Five of the 26 total messages were phishing messages and imitated real phishing messages like IT service alert, a cloud storage share request, a deal from a hotel chain, a payment receipt and a security alert. The link in these phishing messages redirected them to either the rich JIT webpage or the lean JIT webpage designed by us. After 30 minutes of working on the task, the participants were directed to a post-survey. After the experiment, the browser log of the participant's machine was collected to identify the phishing webpages visited by the participant.

RESULTS AND DISCUSSION

Our preliminary results show a number of interesting findings. First, it is important to note that it was necessary to separate testing of the hypotheses training into two models because all the participants for the general training and the leaderboard training experienced (or not) the intervention regardless of their actions in the experiment. In the JIT condition, participants only experienced training if they clicked on a phishing message. To test the efficacy of JIT training, we ran a binary logistic regression with subjects that clicked on at least one phishing message (i.e., received the JIT training). We found that those subjects in the rich JIT condition did not click on significantly fewer messages than those in the lean JIT condition (p = 0.516). In other words, the likelihood that a subject clicked on another phishing message after either the rich or lean JIT training is statistically similar. Further, we found no difference in the number of correct phishing messages that were reported by the subject for either the rich or lean JIT (e.g., forwarded these messages; p = .985).

In the next model, we executed a multivariate regression that included both the general training and the leaderboard training as fixed factors and propensity to trust, internet risk, experience with phishing, and sex as covariates. The dependent variables (DV) were 1) number of phishing messages clicked, and 2) number of messages reported as phishing (e.g., forwarded to the reporting account). All of the covariates in this model (trust, risk, sex) were not significantly related to the DVs at p > .10. Phishing experience was significantly related to the number of phishing messages the user clicked p = .08. Both the leaderboard and the general trainings were significantly related to how many phishing messages subjects clicked (p = 0.02, p

< 0.01, respectively). The interaction of the leaderboard and video/quiz was also significantly related to the number of phishing emails that subjects clicked on. There was no significant relationship between the fixed factors and the covariates to how many phishing messages were reported. Examining the means of the number of phishing messages clicked in each condition, we see that the general training reduced the number of messages click by 42% and the leaderboard reduced the amount of phishing messaged clicked by 41%.

Table 1. Mean Scores for # of Phishing Messages Clicked.

Training	Mean # of Phishing Messages Clicked (Presence / Absence)	Standard Error (Presence / Absence)
Pre-Training Video	0.57 / 0.99	0.10 / 0.09
Just-in-Time Training*	2.01/ 1.94	1.25 / 1.24
Leaderboard	0.58 / 0.98	0.09 / 0.10

Note: * The means of those who experienced JIT training (e.g., fell for at least one phishing message)

In our preliminary analysis of the data from this experiment we found that only the general training (H1a) and the leaderboard (H3a) induced significant differences in the number of phishing emails clicked on by the participants. Further, there is no evidence that JIT training (H2a) reduced the likelihood to click on a phishing message. Finally, there is no evidence that any of these trainings increase the propensity to report phishing messages.

Currently, we have several other factors that are being investigated in this data set. First, we are currently coding the response quality of the task that was given to the subjects. We will then use a regression model to analyze how the different interventions impacted work related observations such as coded quality of task, the number of messages sent, the number of non-phishing messages that were reported as phishing, etc. We have also collected data on stress related constructs which we will use as covariates in this model. They include computer anxiety, technology paralysis, perceived disruption of task, and perceived threat. Also, we are currently collecting data in a field experiment where we will examine the efficacy of training in the field using the participants' own organizational email accounts. Finally, we did find an interaction between the general training and the leaderboard. We need to explore these interactions further. The full results of the laboratory experiment and the field experiment will be reported in the presentation.

ACKNOWLEDGEMENT

This research was supported by a grant from the National Science Foundation, Social and Economic Division, Project# XXXXX. The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies or endorsements, either expressed or implied, of NSF or the U.S. Government.

REFERENCES

Deterding, S., Khaled, R., Nacke, L., Dixon, D. (2011) "Gamification: Toward a Definition," in *CHI 2011 Gamification Workshop Proceedings*, Vancouver, BC, Canada

Duffy, T., and Jonassen, D. (1992) "Constructivism and the technology of instruction: A conversation," in *Hillsdale*, NJ: Erlbaum

Wright, R.T., Durcikova, A., Miranda, S., Jensen, M.L., Bernecker, S., & Kelly, K. (2016) The Building of an Organizational Framework for the Human Firewall in Briggs, Robert O. and Nunamaker, Jay F., Jr. (Eds.) *Report of the Hawaii International Conference on System Sciences*, Symposium on Credibility Assessment and Screening Technologies , Kauai, HI, Jan 4-6.

Durcikova, A., Jensen, M.L., Wright, R.T. (2015) Building the Human Firewall: Lessons from Organizational Anti-Phishing Initiatives, in Briggs, Robert O. and Nunamaker, Jay F., Jr. (Eds.) *Report of the Hawaii International Conference on System Sciences, Symposium on Credibility Assessment and Information Quality in Government and Business*, Kauai, HI, Jan 6-10.

Epple, D. (1991) "Organizational Learning Curves: A Method for Investigating Intra-Plant Transfer of Knowledge Acquired Through Learning by Doing" in *Organization Science*, Vol. 2, No. 1, pp. 58-70

Glover, I. (2013) "Play As You Learn: Gamification as a Technique for Motivating Learners" in *Ed Media*, pp. 1999–2008, 2013.

Hong, J. (2012) "The state of phishing attacks," in *Communications of the ACM*, 55(1): p. 74-81.

Hung, S.Y., Durcikova, A., Lai, H.M., and Lin, W.M. (2011) "The influence of intrinsic and extrinsic motivation on individuals' knowledge sharing behavior," in *International Journal of Human-Computer Studies*, 69(6): p. 415-427.

Jensen, M. L., Lowry, P. B., and Jenkins, J. L. (2011) "Effects of automated and participative decision support in computer-aided credibility assessment," in *Journal of Management Information Systems,* vol. 28(1), pp. 201- 234

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., Nunge, E. (2007) "Protecting People from Phishing: The Design and Evaluation of an Embedded Training Email System," in *Proceedings of the 2007 Computer Human Interaction, CHI.*

Landers, R. N., Callan R. C. (2011) "Casual Social Games as Serious Games: The Psychology of Gamification in Undergraduate Education and Employee Training," in *Serious Games and Edutainment Applications* pp 399-423

McCabe, J. (2016) "FBI Warns of Dramatic Increase in Business E-Mail Scams." https://www.fbi.gov/contact-us/field-offices/phoenix/news/press-releases/fbi-warns-of-dramatic-increase-in-business-e-mail-scams.

Phishme. (2013) "PhishMe surpasses 200 customers and 4 million users trained." http://phishme.com/phishme-surpasses-200-costumers-4million-users-trained/

Puhakainen, P., and Siponen, M. (2010) "Improving Employees' Compliance Through Information Systems Security Training: An Action Research Study" in *MIS Quarterly* Vol. 34, No. 4, pp. 757-778.

Vance, A., Siponen, M., and Pahnila, S. (2012) "Motivating IS Security Compliance: Insights from Habit and Protection Motivation Theory," in *Information & Management*, 49(3): p. 190-198

Winter, M. (2014) "Home Depot hackers used vendor log-on to steal data, e-mails." http://www.usatoday.com/story/money/business/2014/11/06/home-depot-hackers-stolen-data/18613167/

Wright, R.T., and Marett, K. (2010) "The influence of experiential and dispositional factors in phishing: An empirical investigation of the deceived," in *Journal of Management Information Systems,* 27(1): p. 273- 303.

Wright, R.T., Jensen, M. Thatcher, J.B., Dinger, M. & Marett, K. (2014) Influence Techniques in Phishing Attacks: An Examination of Vulnerability and Resistance. *Information Systems Research* 25(2) pp. 385-400