A Phishing Study of User Behavior with Incentive and Informed Intervention

Supriya Muthal, Sen Li, Yuan Huang, Xiangyang Li, Anton Dahbura
Johns Hopkins University Information Security Institute
3400 N. Charles St., Baltimore, MD 21218
{smuthal1, sli96, yhuan110, xyli, antondahbura}@jhu.edu

Nathan Bos, Kylie Molinaro
Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel, MD 20723
{Nathan.Bos, Kylie.Molinaro}@jhuapl.edu

ABSTRACT

A user study was designed to understand user security behavior when processing phishing emails. Previous research suggests that people are victimized by phishing emails due to a lack of awareness and the adverse effects of time pressure and distraction on information processing. We looked deeper to explore what phishing indicators users overlook more often than others, and whether applying interventions that emphasize such phishing tells and awarding incentives for good performance improve accuracy and influence task completion time. More specifically, 20 participants of mixed educational backgrounds were recruited to perform an email sorting task. Participants were instructed to move emails into a *suspicious* or *legitimate* folder. Phishing emails varied by three different phishing tells: sender's email address, link or attachment payload, and message composition. Each participant completed three rounds of the sorting task in one session. In the second round, one phishing tell, with which the participant struggled the most in the first round, was modified in a way to make it easier to recognize. Moreover, one group of participants was offered a financial reward if their classification accuracy reached 80% or better. Participants' performance data of classification accuracy and task completion time were analyzed and presented with a few interesting findings. This paper discusses the complexity of conducting such a user study and describes the research experience that the team had.

Categories and Subject Descriptors

Security and privacy: Human and societal aspects of security and privacy: Usability in security and privacy

Keywords

Security, User Study, Phishing, Intervention, Incentive

1. INTRODUCTION

Phishing attacks are becoming more and more sophisticated over time as adversaries are learning new techniques and strategies to attack Internet and email users to steal sensitive information. Hackers can conduct spear phishing attacks based on personalized communication to improve the effectiveness of such tactics. Defense mechanisms against phishing are not as effective and those protecting against phishing attacks have less knowledge about how users will behave in experiencing an attack, compared to the attackers who are more attentive of user behaviors and rely on the users' tendency to fall for these attacks.

This project examined the current trends in email phishing attacks and designed a simulated phishing study to better understand user perception, efficiency, and decision-making. To answer the questions and test our stated hypotheses, the study collected information on security related decisions by participants in an incentive group and a control group. Both groups completed three testing rounds: pre-intervention, intervention, and post-intervention. Participants were required to perform an email sorting task that included both phishing emails and non-phishing (with spam email allowed) emails. Participants were instructed to sort each email into one of two categories, suspicious or legitimate, based on their perception of the security threat risk of that email. Each phishing email had a specific phishing tell from one of the three categories that we focused on to study: 1) suspicious sender's email address; 2) malicious email payload; and 3) unprofessional/poor email composition.

We conducted the study in three rounds for each participant, where in the second round we introduced support for one of the three phishing tells that the participant had lowest accuracy score in round 1. The intervention introduced modifies phishing tells to make them appear more obvious, so that it should be easier for the participant to identify the phishing tells and classify emails accurately. Then, in the third round, the participant was tested again on a different set of emails similar to those in the first round. We aimed to understand if there was any training or learning effect produced by the intervention in the second round. We were also interested in user behaviors working under introduced "pressure" in the form of a financial reward for a high level of accuracy in all three rounds.

The performance data included the sorting accuracy and time taken to make a decision for each email. In our study, we also captured participant specific demographic and phishing related information through a post-experiment questionnaire that gave us insight into the participants' background, such as computer habits and general security awareness.

The user study captures users' actions and, according to their performance in handling different phishing tells, provides customized support, and furthermore, introduces a financial incentive based on task performance. Through this project, we aimed to answer the questions and test the hypotheses as stated below:

a. Of the three, which phishing tell is most likely to be overlooked by users?

b. Is the average time spent on legitimate emails more than that on phishing emails?

c. The average time taken by participants on emails with the design intervention will be less than that spent on other emails.

d. Participants in the incentive group will take more time than non-incentive group to sort emails.

e. As the time spent on each email increases, the participants' sorting accuracy will also increase.

f. There will be an increase in sorting accuracy from round 1 to round 2 due to the introduction of the design intervention in round 2.

g. There will be an increase in sorting accuracy from round 1 to round 3 due to a training effect produced by the design intervention in round 2.

h. Participants in the incentive group will have a higher sorting accuracy than those in the control group.

2. RELATED WORKS

2.1 Troublesome Phishing Emails

With spear phishing being the top attack vector and a common attack type on financial institutions and payment services, email has been the most common vehicle to conduct phishing attacks [1]. Companies must be prepared, as attacks are becoming more complicated and phishing emails are hard for users to distinguish from legitimate emails. According to one study, 20% of company staff ends up clicking on a phishing email during work [2]. Gmail is the most popular webmail service used by attackers to launch phishing emails to gain identity credentials and steal personal information. Another study estimated that spear phishing is responsible for 38% of cyber attacks on IT enterprises [3]. Banks suffered financial losses of $2.5 million to $10 million per bank, for a total of up to $1 billion. In a widespread attack on financial institutions, attackers used spear phishing emails containing weaponized .doc (Microsoft Word) and .cpl (Microsoft Control Panel) files as attachments to execute a backdoor software tool called Carbanak [4]. The growing number of incidents has led to an increase in focused research efforts to gain insight about what factors lead to phishing victimization and how design interventions and incentives are needed to prevent this exponentially increasing threat.

2.2 Phishing User Studies on PC

We have reviewed multiple reports on phishing studies and interactive phishing experiments that research why people fall for phishing and how to avoid it. As the number of emails a user must read increases, the more likely he or she is to be deceived and a user more likely attends to emails from senders that he or she feels familiar with [5]. Therefore, users are more likely to trust and read emails coming from popular financial institutions and commercial websites. In our project, we were interested in crafting emails based on this observation.

Phishing indicators are overlooked by a significant percentage of users, as they often do not understand what they should check in an email, and the inconsistent positioning on different web browsers makes the task of identifying a phishing email difficult [6]. The above study emphasized the importance of understanding user behavior in a phishing attack to better defend against it. Our design specifically focuses on several phishing indicators or tells, their significance, and placement in the email to test if participants can differentiate the various phishing tells. Habits form over time as people routinely use email and social media and as soon as a notification arrives, people with entrenched email/social media habits tend to click it even before realizing that they are clicking on it [7]. We collected such information in our study in a post experiment questionnaire.

2.3 Phishing User Studies on Mobile Devices

While the motives of cyber-attacks range from theft to cyber vandalism, activism, industrial and national espionage, almost all the attacks use spear phishing as the vector to initially gain access to an individual's computer or mobile device to infiltrate networks [8]. The use of mobile devices to access emails, bank accounts, and online shop has exponentially increased. It is equally important to study if the use of mobile devices influences individuals falling to phishing attacks.

One study simulated phishing attacks that varied in the cues available in the email [7]. It examined how the device used by subjects to access it influenced the outcome of the attack. The study results showed that there is not much significance of using heuristics in processing emails on mobile devices. Our study is currently limited to desktop computers, but we are looking into testing user behaviors on mobile platforms in future work.

2.4 Phishing User Study with Intervention

We would gain more insights into user security behaviors when interventions are introduced in phishing experiments. One study [9] studied the effectiveness of warning messages, with two user groups, one control group that received no warnings for phishing attack, and another group that received warnings. Out of nine participants, eight failed to act on warnings and fell to the phishing attack. During post-task interviews, most of the participants said they did not understand the meaning of the warning displayed and tended to ignore it in part due to the interface design. A second observation was that about half of the participants indicated that they did not know the definition of phishing.

There is a need to improve security awareness and training against phishing attacks. Instead of flooding users with constant warnings that could become intrusive and annoying, it is important to understand the user's perspective and decision-making process as an effective way of implementing security awareness programs, as proposed by another study on phishing [10]. We aim to build a user study that will test different phishing tells on user behavior as well as the impact of interventions customized to individual users.

In our study, we introduced a design intervention in both the incentive and non-incentive (control) groups (discussed below). Each participant was given help with the phishing tell that he or she struggled with the most in the first round by making that type of phishing tell easier to recognize in the second round.

2.5 Effect of Incentives

We are not aware of phishing user studies that involve incentives. As reported by a study of a statistical reasoning task, performance-based incentives produced significantly better performance than course credit and flat-fee rewards [11]. A strong incentive can promote more objective analysis in situations where there is an objectively correct answer. In our study, participants must select one of the two choices and there is a definite classification to every email. Introducing a performance-based incentive factor would help us understand how participant performance is affected. More specifically, in each round, participants in the incentive group were given a monetary award in addition to the base compensation they were guaranteed if they performed better than the required accuracy threshold (80%).

2.6 Our Approach

Our user study is different from the related works in several aspects of its design.

(1) Each participant has three rounds of the email sorting task, in which we can introduce a customized intervention to help participants in their weak areas and then test the retaining effect of such "training" in the last round. We specifically target one type of phishing tell for every participant to make it easier for participants to recognize that tell in the second round, and then remove the help in the third round to test if the intervention in the second round had lasting effect on performance improvement.
(2) Our study tests how monetary incentives impact participants' security decision making and the time to complete such tasks. There are two participant groups, the control group that receives a flat-fee compensation and a treatment group that receives performance-based compensation. We compare accuracy and email sorting time of both experimental groups.
(3) Real time data collection and analysis is critical to the individualized intervention scheme. We chose a web-mail system in order to automatically capture and analyze the performance data of each participant. Therefore, after the round 1, there was minimal time needed to set up the customized emails for round 2 to start.

3. USER STUDY DESIGN

Participants were tasked with accurately sorting emails by moving them into folders, one for legitimate emails and the other for suspicious.

3.1 Email Sorting Task

As previously mentioned, one important reason that victims fall for phishing emails is because of the sender's perceived familiarity. It is likely a user clicks a request to reset password seemingly from his or her bank. One key goal in a phishing study is to make the phishing emails personal to the participant. We chose a task design developed in a previous pilot study in the summer of 2016, where the participant was asked to screen the emails as the personal assistant for a professor. In this way, the participant, without the full knowledge of the professor's private life, has to deal with uncertainty in judging whether the emails are truly personal to the professor.

In each email sorting task, participants were presented 20 emails, with a mix of phishing and legitimate emails. There were 15 phishing emails, five of each type of phishing tell. The five legitimate emails could include spam emails. In the study, we defined spam emails as unwanted (e.g., advertising, promotions, etc.), but not malicious. The participant must move each email into one of the two email folders. Note that the participant was not allowed to click the link or check anything on Internet. He or she had to base the classification on the email itself.

This study had two user groups with 10 participants in each. One is the control group without any incentive (participants only received the base amount of $20) and the other is the incentive group, where each participant could be paid from $10 - $15. For participants in the incentive group, if their classification accuracy rate in every round is higher than 80%, a bonus of $5 is added to the $10 base compensation.

3.2 Phishing Tells

There are three different types of phishing tells that were studied in this project.

a. Phishing Tell 1: Suspicious sender's email address

Phishing emails in this category have a suspicious email address. For example, it can have a suspicious domain name, or misspelled addresses of popular social networking and domain names (e.g., number '0' in the place of letter 'o').

For the intervention for this type of phishing indicator, the email address has more suspicious domain names and the email address is always displayed.

b. Phishing Tell 2: Suspicious link/attachment

Phishing emails in this category have either a suspicious link or a suspicious attachment. For example, it could have exe/pdf file attachments or suspicious looking links, but with labels that do not match the URL addresses.

For the intervention, the URL address is displayed and the attachment is always an executable file type.

c. Phishing Tell 3: Suspicious email composition

Phishing emails in this category are suspicious in layout and writing. For example, the logo, images, spelling, or grammar in the email could be incorrect or improperly formatted.

For the intervention, such traits are more obvious or spelling and grammar errors are shown in uppercase letters in the phishing email.

3.3 Email Sorting Rounds

There were three total rounds, each consisting of 20 total emails, five of which were legitimate and 15 phishing. The 15 phishing emails were comprised of five emails for each of the previously described phishing tells. Participants had 15 minutes for each round to sort the emails and a two-minute break in between each round. The accuracy for every phishing tell has a full score of five, so in total 15 for phishing emails and five for legitimate emails, thus making a total score of 20. After each round, a python script processed data collected for the participant automatically to calculate their scores. For the incentive group, the overall score determined the reward to the participant.

In round 2, we challenged participants with a different set of 20 emails, again mixed with emails of the three types of phishing tells and legitimate emails. The calculation of round 1 performance also determines the five emails of the type of phishing tell for which the participant scored the lowest. If there is a tie between scores of phishing tells, then we randomly choose a phishing tell.

Round 3 was the same as round 1. It had 20 emails (15 phishing, five legitimate) without any intervention, similar to those used in the first round. We captured the performance data to check whether the intervention resulted in a training effect.

3.4 Data Collection

We developed an experimentation infrastructure and data collection methods that automatically recorded detailed actions such as clicking, navigation, moving an email, etc. The data could then be imported and processed for further analysis.

The system structure has three components as shown in Figure 1, a RoundCube webmail server, a web based email client, and a BurpSuite proxy listener sitting in between. When processing the emails, the email client sends HTTP requests to the RoundCube email server and the server responds with HTTP traffic. Both requests and responses are relayed through the BurpSuite proxy server that was set up before the experiment began. The proxy listener intercepts communication between the email client and server and captures all the data used in the analysis in its logs.

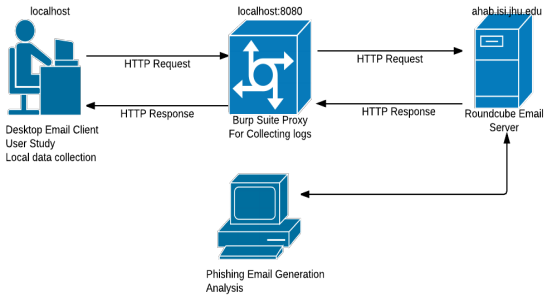

Figure 1. Data collection system architecture

After each round, BurpSuite logs were saved for each participant and initial analysis was performed. We first converted the logs to an XML file. A Python script then parsed this file to extract all email classification and timing data. It then calculated accuracy and time spent on each email.

4. USER STUDY METHODS

All research team members completed the required training certificate for conducting social research. The study gained appropriated Institutional Review Board (IRB) approval. The user study sessions and data management followed the approved protocols outlined in the IRB.

4.1 Recruitment

We posted recruitment announcements on the Johns Hopkins University announcement page, as well as, shared the advertisement with the Computer Science department and the Information Security Institute. The recruited participants had a mix of both computer science and non-computer science backgrounds. The average age of the participants was 23 years old with a maximum and minimum age of 18 and 38 years old, respectively. Seventeen of the 20 participants were students, six having a cybersecurity background, seven with a computer science background, and four with neither a cybersecurity nor computer science background. The remaining three participants were university faculty or staff members with a cybersecurity background.

4.2 Experimental Protocol

At the start of each session, we followed the IRB protocol to brief participants about the task and get an informed consent in writing. We then trained the participants on how to use the email client interface. This included a practice trial consisting of five emails to get familiar with the webmail environment and the task. Note that participants were only told to classify each email to one of the two email folders based on their perception of whether there is risk of personal identity information being stolen for a malicious purpose. They were not given specific instructions on where to look for pertaining information.

After the training, we started our data collection tool and logged into the webmail account that was pre-loaded with 20 emails for the first round. After the first round, we ran the Python code to process BurpSuite logs for performance analysis in order to decide the next set of emails to be used, and then logged into the corresponding webmail account with the right set of emails. The third round was very similar to the first round. We also video-recorded the user's screen as a backup to reduce the risk of total data loss in the event the data collection tool malfunctioned.

At the end of the session, we used a post-experiment questionnaire to collect participant demographics and other relevant information such as social media habits and email usage.

After each participant finished their task, we saved all the data collected for that participant using a common naming convention after removing any personally identifiable information, and stored it on a Google drive project folder with restricted access only to IRB certified team members.

5. DATA ANALYSIS AND RESULTS

This section presents results of this user study regarding the questions and hypotheses asked at the beginning of its design. We performed paired *t*-tests for each participant going through the three rounds.

5.1 Summary of Participant Performance

The performance metrics are the classification accuracy (number of emails correctly classified) and the email processing time (in seconds). Figures 2 and 3 summarize these two measures respectively for all 20 participants, numbered from A to T, in each of the three rounds. The participants are in the control and incentive groups in alternating order, i.e., participants A, C, E..., were in the control group, while participants B, D, F…, were in the incentive group.

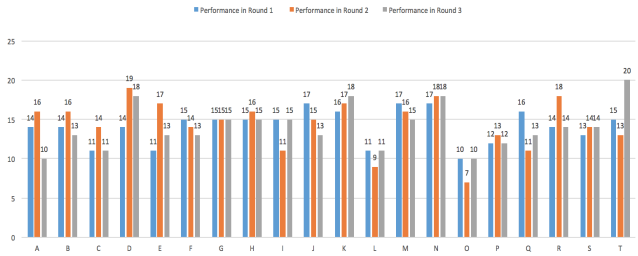

Figure 2. Classification accuracy of each participant by round

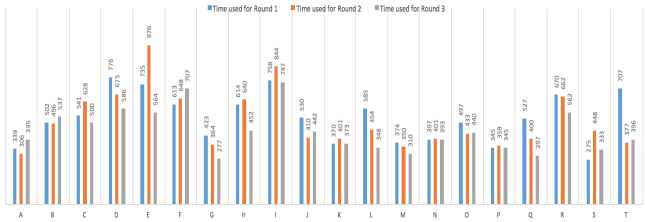

Figure 3. Time spent in each round by participant

5.2 Research Questions and Hypotheses

We used the IBM SPSS software to perform statistical analyses and significance tests. Our main research interest was to study the impact of the intervention mechanism for different phishing tell types and monetary incentives on the classification accuracy and time spent on the emails. Next, we will discuss the results for each of the questions and hypotheses from *a* to *h* presented in section 1.

a. Of the three, which phishing tell is most likely to be overlooked by users?

50% participants were victimized by phishing tell 1 (suspicious email address), 30% participants were victimized by phishing tell 3 (suspicious composition), and 20% participants were victimized by phishing tell 2 (suspicious links/attachments). This indicates that phishing tell 1 was missed the most by participants.

b. Is the average time spent on legitimate emails more than that on phishing emails?

The average time spent on individual legitimate and phishing emails was 23.97s and 27.61s respectively. We also noticed that participants spent more time on legitimate emails in round 1 and 2, but in round 3, they spent more time on phishing emails.

c. The average time taken by participants on emails with intervention is less than that spent on other emails.

Results show that in round 2, the average time that participants spent on intervention emails and non-intervention phishing emails is 22.14s and 23.20s respectively. However, there is no statistically significant difference.

	N	Minimum	Maximum	Mean	Std.Dev.
Avg_Time_Intervention	20	8.40	61.80	22.1400	13.39
Avg_Time_None_Intervention	20	12.78	39.95	23.2025	7.46

Figure 4. Average time spent on an email with intervention versus an email without intervention

We need to look more closely at the times spent in round 1 and round 2, for the type of phishing tell for which intervention was provided.

Paired Samples T-Test					
R1_P1_Time-R2_P1_Time	Mean	Std.Dev.	t	df	Sig.
	29.300	39.766	2.330	9	.045

Figure 5. Difference in average time spent on a phishing tell 1 email from round 1 to round 2 (incentive group)

For the incentive group, shown in Figure 5, the notation "*R1_P1_Time*" represents the average time to process one phishing tell 1 email in round 1. We found that the participants who were given the intervention for phishing tell 1 in round 2, on average, spent 29.3 seconds less time on one such email, with a *p*-value of 0.045. However, this does not indicate their classification accuracy changed in one way or the other.

Paired Samples T-Test					
R1_P3_Time-R2_P3_Time	Mean	Std.Dev.	t	df	Sig.
	-34.60	28.563	-3.831	9	.004

Figure 6. Difference in average time spent on a phishing tell 3 email from round 1 to round 2 (control group)

An interesting result is shown in Figure 6. Participants with no incentive spent on average 34.6 seconds more time on a phishing tell 3 email from round 1 to round 2. Recall that the phishing tell 3 is that the logo, images, spelling, or grammar in the email could be incorrect or improperly formatted. This could suggest that these participants might become more attentive of such information emphasized in round 2.

d. Participants in the incentive group will take more time than the non-incentive group to sort emails.

As in Figure 7, the result shows that in round 3, the 10 participants from incentive group spent on average 24.60 seconds more to sort legitimate emails with a *p*-value of 0.024. In other word, the incentive group spends more time on legitimate emails in round 3. We did not find other significant results.

R3_Normal_Time	Incentive	Control	Mean Diff.	df	Sig. (2-tailed)
	105.20	80.60	24.60	18	.024

Figure 7. Difference in average time spent on a legitimate email in round 3 (incentive & control groups)

e. As the time spent on each email increases, the participants' sorting accuracy will also increase.

The results are not significant. We did not find a correlation between the time spent and the classification accuracy. We are further looking into the data.

f. There will be an increase in sorting accuracy from round 1 to round 3 due to a training effect produced by the design intervention in round 2.

We found that the participants receiving help for phishing tell 3 improved their accuracy. In Figure 8, "*R1_P3_Score*" represents the accuracy score for phishing tell 3 emails from round 1 to round 2, for those participants who received help with phishing tell 3. The *p*-value of 0.011 shows that a significant increase of 1.667 in accuracy for six participants. Recall that phishing tell 3 is email layout and composition errors. Likely the intervention of highlighting these issues in an email was noticeable.

Paired Samples T-Test					
R1_P3_Score-R2_P3_Score	Mean	Std.Dev.	t	df	Sig.
	-1.667	1.033	-3.953	5	.011

Figure 8. Difference in classification accuracy for phishing tell 3 intervention from round 1 to round 2 (incentive & control groups)

Of the above cases, we further show the performance improvement for those participants with an incentive in Figure 9. The improvement is significant with an even higher accuracy increase of 2.25 on average. This likely indicates that the monetary incentive made participants more attentive to the intervention received.

Paired Samples T-Test					
R1_P3_Score-R2_P3_Score	Mean	Std.Dev.	t	df	Sig.
	-2.250	.500	-9.000	3	.003

Figure 9. Difference in classification accuracy for phishing tell 3 intervention from round 1 to round 2 (incentive group)

However, the interventions for the other two phishing tells did not show a difference in classification accuracy. This highlights the complexity behind providing effective interventions to users.

g. There will be an increase in sorting accuracy from round 1 to round 3 due to a training effect produced by the design intervention in round 2.

Again, when combining all participant data, we did not find significant differences in the classification accuracy from round 1 to round 3. This includes the performance for the phishing tell 3 emails, for which, with intervention, significant improvement was seen from round 1 to round 2. This shows the challenge in designing interventions that have a lasting effect.

Figure 10 shows a less significant result for phishing tell 2. Four participants received the intervention for phishing tell 2. They had, on average, a 0.75 higher accuracy score for the phishing tell 2 emails from round 1 to round 3, but with a *p*-value of 0.058, this

is not a statistically significant difference. More samples are needed to better understand this potential difference.

Paired Samples T-Test					
R1_P2_Score-R3_P2_Score	Mean	Std.Dev.	t	df	Sig.
	-.750	.500	-3.000	3	.058

Figure 10. Difference in classification accuracy for phishing tell 3 intervention from round 1 to round 3 (incentive group)

h. Participants in the incentive group will have a higher sorting accuracy than those in the control group.

Figure 11 shows that in round 3, those 10 participants from the incentive group achieved, on average, a 1.3 higher accuracy score for phishing tell type 1 emails with a *p*-value of 0.027. The incentive group also performed better in sorting phishing tell type 1 emails in round 3.

R3_P1_ Score	Incentive	Control	Mean Diff.	df	Sig. (2-tailed)
	3.50	2.20	1.30	18	.027

Figure 11. Difference in classification accuracy for phishing tell 3 emails in round 3 (incentive & control groups)

When combining all participant data, we did not find significant differences. However, the findings for hypothesis *f* suggest that an appropriate incentive may result in better improvement in coordination with a helpful intervention.

6. CONCLUSIONS AND FUTURE WORK

This paper presented a user study of phishing email recognition. We evaluated how participants performed an email processing task while varying the help according to their capability of correctly classifying emails in multiple rounds. Moreover, we offered monetary rewards based on the accuracy of their performance to incentivize participants in a treatment group.

Preliminary data analysis has shown several interesting insights, but, more importantly, demonstrated the complexity of user security behaviors and the challenges when developing lasting and meaningful design interventions. Specifically, we saw that although there was performance improvement when certain types of intervention were provided to the participant, that effect did not carry over after the intervention was removed. On the other hand, the use of monetary incentives may make participants more attentive to benefit from an intervention, compared to the control group. However, that did not always translate to a higher classification accuracy.

We are continuing our research to consider more realistic scenarios where users handle multiple tasks in a phishing email recognition setting. This calls for further effort to carefully amend the user study protocol and to fundamentally understand how participants react to this environment.

7. ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation under Award 1544493. We also would like to thank Brynne Harrison from the University at Buffalo for her help with the user study experimental design and data analysis.

8. REFERENCES

[1] APWG (2016), *Phishing Activity Trends Report*. Retrieved from http://docs.apwg.org/reports/apwg_trends_report_q2_2016.pdf

[2] Finextra (2015), *JPMorgan dupes 20% of staff into opening fake phishing email*. Retrieved from https://www.finextra.com/news/fullstory.aspx?newsitemid=28278

[3] Cloudmark (2016), *Survey Reveals Spear Phishing as a Top Security Concern to Enterprises*. Retrieved from https://blog.cloudmark.com/2016/01/13/survey-spear-phishing-a-top-security-concern-to-enterprises

[4] Cloudmark (2016), *Spear Phishing: The Secret Weapon Behind the Worst Cyber Attacks*. Kaspersky Report. Retrieved from https://blog.cloudmark.com/2016/01/13/spear-phishing-secret-weapon-in-worst-cyber-attacks/

[5] A. Vishwanath, T. Herath , R. Chen, J. Wang, H. Raghav Rao (2011), Why do people get phished? Testing individual differences in phishing vulnerability within an integrated information processing model, *Decision Support Systems*, Vol. 51, No. 3, Pp. 576-586.

[6] R. Dhamija, J. D. Tygar, & M. Hearst (2006), Why phishing works, *Proceedings of the SIGCHI Conference on User Factors in Computing Systems*, pages 581-590

[7] A. Vishwanath, B. Harrison, & Y. J. Ng (2016), Suspicion, Cognition, and Automaticity Model of phishing Susceptibility, *Communication Research*, February 2016.

[8] FireEye (2013), *Exposing One of China's Cyber espionage units*, Threat Intelligence Report.

[9] W. Yang, J. Chen, A. Xiong, R. W. Proctor, & N. Li, (2015). Effectiveness of a phishing warning in field settings, *Proceedings of the 2015 Symposium and Bootcamp on the Science of Security*, Urbana, Illinois, April 21-22, 2015.

[10] M. A. Sasse., & I. Kirlappos (2012), Security education against phishing: A modest proposal for a major rethink, *IEEE Security & Privacy*, Vol. 10, No. 2, Pp. 24-32.

[11] G. L. Brase (2009), How different types of participant payments alter task performance, *Judgment and Decision Making*, Vol. 4, No. 5, Pp. 419-428.