Mouse Behavior as an Index of Phishing Awareness

Kun Yu, Ronnie Taib, Marcus A. Butavicius, Kathryn Parsons, Fang Chen

Mouse Behavior as an Index of Phishing Awareness

Kun Yu(✉)*#, Ronnie Taib#, Marcus A. Butavicius^, Kathryn Parsons^ and Fang Chen*

* University of Technology Sydney, Ultimo, NSW 2007, Australia
# Data61, CSIRO, Eveleigh, NSW 2015, Australia
^ Defence Science and Technology Group, Edinburgh, SA 5111, Australia
{Kun.yu,Fang.chen}@uts.edu.au, Ronnie.taib@csiro.au,
{Marcus.Butavicius, Kathryn.Parsons}@dst.defence.gov.au

Abstract. Phishing attacks are one of the most common security challenges faced by individuals and organizations today. Although many techniques exist to filter out phishing emails, they are not always effective leaving humans as the most vulnerable links in the information security chain. This paper presents a study investigating how human behavior, especially mouse movements, may reflect cybersecurity awareness, in particular to phishing emails. Using an email sorting task, we examined three key mouse movement features: hover, slow movement, and response time. The results suggest that slow mouse movements indicate high awareness of phishing emails and could be used to determine the likelihood of users falling victim to phishing attacks. However, contrary to intuition, response time and mouse hovering behaviors do not correlate with phishing awareness.

Keywords: Cybersecurity; Phishing; Mouse Movements; Email Classification.

1 Introduction

Phishing refers to the fraudulent attempt to obtain sensitive information or access to the recipient's computer or system from people in electronic communications [1], [2] and is the most popular cyberattack method [3]. Although most organizations deploy technical defenses such as email filtering against cyber threats, employees still receive many phishing emails, due to more sophisticated and sometimes personalized content being crafted by attackers (e.g., spear phishing). In practice, humans remain the most vulnerable link in the information security chain, and in many cases it is the human behavior that makes a cyberattack possible and successful [4].

Extensive research has been conducted to investigate how and why people click on the suspicious URL or attachment in an email, and it is suggested that knowledge of phishing threats, usage of situational cues (such as the URLs in the email) and perceptions of severe negative consequences are possible indicators of phishing awareness [5]–[7]. However, most of these studies are focused on theory verification or validation with very few addressing empirical techniques to predict whether a suspicious URL is likely to be clicked by a given person via their behavioral data, or at what stage a person should be warned of a potential phishing threat.

Besides theoretical investigations, various intervention methods have been devised to educate people to avoid being phished, and typical anti-phishing skills include examination of the sender information, the addressing of the email, the hyperlinks involved in the email and typographic errors in the text [5], [8]. Focusing on these characteristics of emails, anti-phishing training can improve the general phishing awareness [9]. However, a common challenge faced by the anti-phishing education schemes is that they do not customize education to different users – an approach which has been shown to be effective when the training is tailored to individual differences such as learning style preferences [10].

Many emails are still processed on personal computers today, with the mouse as most widely used interface to browse the contents of the emails, check the sender information, click on URLs or open attachments. Existing research has revealed that tracking user's mouse movements is effective for website usability evaluation [11], [12], that a mouse position gives an indication of the user's gaze during their online tasks [13]–[15], and corresponds to user's attention [16]. Huang et al. suggested that mouse hovering gestures are related to user observation and thinking when browsing online search results [17]. This finding motivated us to consider whether typical features related to mouse movements, e.g. mouse hover or slow mouse movements can be used as indicators of a user's likelihood to click a phishing link. For example, slow mouse movements may suggest that the subject is reading slowly which may indicate a less impulsive decision making style that has been empirically demonstrated to be linked to phishing email resistance [18]. However, to our knowledge no similar examination has been conducted to date.

Compared with existing phishing awareness examination methods, there are several advantages of a mouse behavior-based method: as mouse movements occur naturally, the examination can be done in real time in a non-intrusive way without interfering with the user's online interactions. Furthermore, a user can be warned, or links can be deactivated, whenever relevant mouse behaviors are detected that reflect high susceptibility to clicking on suspicious content. Finally, the personalized nature of mouse movements makes it possible to design personalized phishing interventions. Hence, we propose three research hypotheses on the relationship between mouse features and user response to phishing emails:

1. Response time can indicate a user's decision when processing phishing emails;
2. Mouse hovers can be used to gauge the level of phishing threat awareness;
3. Slow mouse movements can be used to identify whether a phishing threat can be identified.

2 Method

Based on a user study involving 30 email sorting tasks, we collected behavioral data, as well as post hoc subjective feedback.

2.1 Participants

Thirty-three volunteers, including thirteen females, from a multi-national research organization participated in the experiment. Thirty participants finished all the tasks although two participants used an iPad and therefore had no mouse input. As a result, we collected mouse data from twenty-eight subjects (10 females) with an age range of 30-39 years old. Participants had reasonable self-reported proficiency in English (on average 7.6 out of a 10-point scale) and good computer skills (on average 7.7 out of a 10-point scale). Most participants (27 out of 28) finished the experiment within twenty minutes. Ethics approval was obtained for this study, and all the subjects confirmed their consent before the experiment and were aware that they could opt out at any time. They were unpaid but received a small snack as acknowledgement of their time.

2.2 Email Interface

We adapted the real-life examples of phishing emails from the UC Berkeley Information Security and Policy Phishing Examples Archive [19], and adapted the URLs and logos embedded in the email to suit the local population. We converted the emails to images of fixed size and calibrated the location of the URLs to enable the hover effect of the mouse (Fig. 1). Specifically, each email had a hyperlink-enabled element which could be either text, URL or a button. When the mouse hovered on this element for a short time (browser tooltips typically trigger after 500ms) without click, a popup message appeared, which revealed the real URL in a similar way that most modern email clients do. However, due to security reasons, none of the links were active during the experiment. In total thirty emails were crafted, including twenty phishing emails and ten legitimate ones, the latter being compiled from real emails received by the experiment designers in the past.

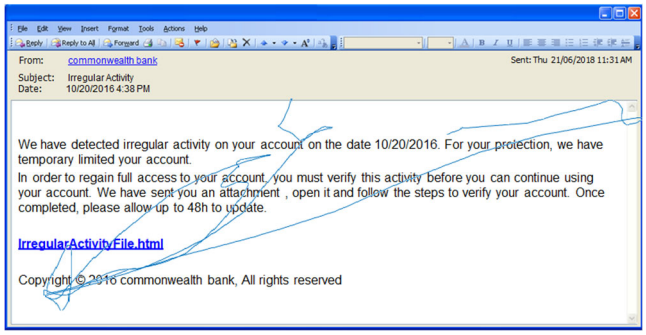

Fig. 1. A crafted phishing email with the retrieved traces of mouse movements from a user shown in blue.

2.3 Experiment Procedure

The emails were presented to the participants in a randomized order, and the participants were asked to finish the tasks as soon as possible, although there was no time limit for the tasks. For each email, the participant was asked to classify it into one of three categories: high priority, low priority, or suspicious. We assume that participants would only put emails with legitimate or innocuous content and links into the first two

categories. Upon completion of the email tasks, the participants answered three distinct open-ended questions (one per text in curly brackets): “Of the emails you categorized as {HIGH PRIORITY | LOW PRIORITY | SUSPICIOUS}, what aspects of the emails influenced your decision?”. They could answer ‘not applicable’ if they had never selected that category during the tasks.

2.4 Data Collection

We used a crowdsourcing tool to present the email interfaces, organize the questionnaires, and capture mouse movements. In particular, we collected timestamped mouse movements and click traces throughout the experiment. Using this information, we were able to replay the mouse behaviors, as shown in Fig. 1.

We grouped data for the High Priority and Low Priority categories as we were only interested in whether an email was considered suspicious or not. This resulted in a binary decision for each email (*Suspicious* / *Non-suspicious*), from each participant. We used a similar reclassification (i.e., *Suspicious* / *Non-suspicious*) for the qualitative answers to the final questionnaire.

In this paper, we focus on the participants’ responses to phishing emails only, hence the data collected from the non-phishing emails is out of scope and not discussed below. We derived the following variables for each email sorting task:

- Email category: the subjective decisions on whether a given email was phishing (*Suspicious*) or not (*Non-suspicious*);
- Response time T_r: the time elapsed before the participant selected a category for the email. We intentionally removed mouse movements within 200 milliseconds prior to the mouse click on the email category (*Low Priority*, *High Priority* or *Suspicious*) as we considered that the decision has already been made by that time, and hence the user behavior was no longer related to the decision process itself.;
- Mouse hover time T_h: a hover was registered if the mouse cursor did not change location between 100ms and 3s. This displayed a popup message when over a hyperlinked element. Longer mouse dwelling was considered as idle state and discarded;
- Mouse hover ratio R_h was calculated as

$$R_h = \sum_{task} T_h / T_r \tag{1}$$

  indicating the normalized mouse hove time per task;
- Mouse hover frequency F_h: the number of occurrences of hovers within an email sorting task;
- Slow mouse movements T_s: if the mouse movement speed fell in the bottom quartile (25%) of the overall speed within an email task, it was considered to be slow, and the corresponding time for slow movement was recorded;
- Slow mouse movement ratio R_s was calculated as

$$R_s = \sum_{task} T_s / T_r \quad (2)$$

- Slow mouse movement frequency F_s: the number of occurrences of slow mouse movements within an email sorting task.

3 Results

We analyzed the mouse motion features with respect to human decisions, and further refined these findings with an attempt to categorize subjects that are more vulnerable to phishing threats. We used Welch's *t*-test to examine the differences for all the reported analytics due to the unequal variances involved in the data.

3.1 Response Time

The response time T_r was compared between the correct and incorrect classification of a phishing email, as shown in Fig. 2. There was no significant difference ($t(199)=.73$, $p>.05$), suggesting that the response time does not account for the differences between phishing email identification, which suggests our first hypothesis to be invalid.

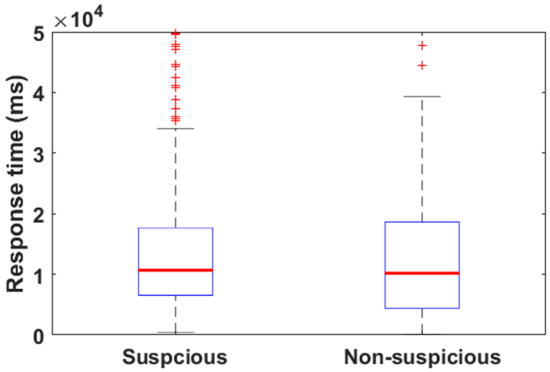

Fig. 2. Response time (T_r) of different decisions on a phishing email. The horizontal red line indicates the median value of each case.

3.2 Mouse Hover

As mouse hovering is considered to indicate when a user is observing and thinking about the email content [17], [19], a high hover frequency should suggest that the subject has examined several elements in the email, hence the email sorting decision should be more thorough. The other variable R_h reflects the total time intentionally spent on examining the elements in an email, and it was expected to exhibit a similar pattern to F_h. However, the Welch's *t*-test suggested no significant difference for F_h ($t(362)=.44$, $p>.05$) or R_h ($t(304)=.55$, $p>.05$) between the correct and incorrect decisions to classify a phishing email. Fig. 3 illustrates the hover frequency for correct and incorrect email classifications respectively. It can be seen that in either case, hover does not occur regularly – actually, for more than half of the emails the participants didn't hover their

mouse at all, hence insufficient hover data is collected and could not be used for subjective phishing awareness examination. These findings invalidate our second hypothesis for the feasibility of using mouse hovers as indicators of phishing awareness.

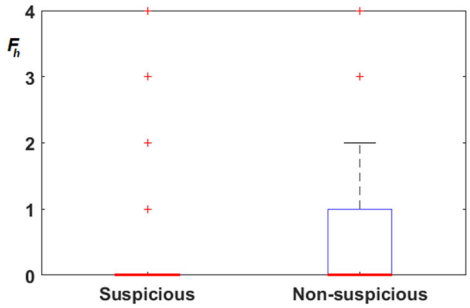

Fig. 3. Hover frequency (F_h) on a phishing email. The median values (red line) for both cases are zero, indicating that hover only occurred in less than half of the tasks.

3.3 Slow Mouse Movements

As illustrated in Fig. 4 (a), slow mouse movement frequency was, on average, lower for the *Suspicious* email decisions and this observation was confirmed by a significant difference in a Welch's *t*-test ($t(274)=-2.3$, $p<.05$).

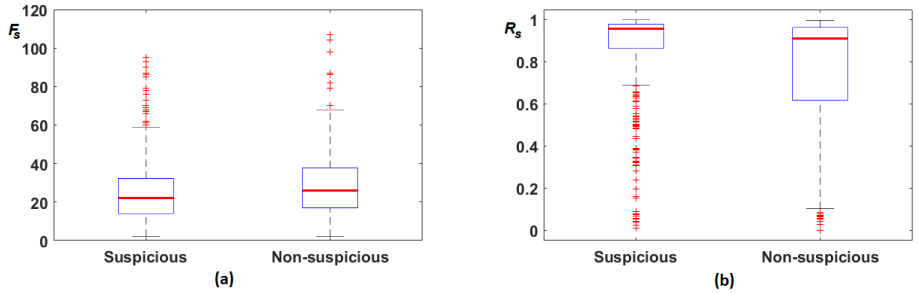

Fig. 4. (a) Slow mouse movement frequency and **(b)** slow mouse movement ratio of user decisions on a phishing email. The horizontal red line indicates the median value.

For the slow movement ratio R_s, a significant difference was also identified via a Welch's *t*-test ($t(236)=4.16$, $p<.01$), as shown in Fig. 4 (b). This suggests that to reach the correct decision to identify a phishing email, the participants overall spent more time performing slow mouse movements.

Combining the examination above on slow mouse movements, correct *Suspicious* email identification by the users relates to high ratio of slow mouse movement time but low occurrences of slow mouse movements. This suggests that, on average, if the mouse is moving slowly and takes a long time, there is a high chance that the phishing email is correctly classified, however frequent slow mouse movements do not account for a correct classification decision.

4 Discussion

In this paper, we explored whether mouse behavior data can be used to characterize susceptibility to phishing threats. The results confirmed slow mouse movements can reflect awareness. However, contrary to intuition, neither response time to sort an email, nor mouse hovers were able to indicate phishing awareness.

Response time has been shown to reflect several aspects of the mental process, including thinking, interpreting, processing and decision generating [20], [21]. It is notable that although different decisions are made when sorting phishing emails, the decision time didn't differ significantly, suggesting that any difference in cognitive processing time may be orders of magnitude smaller than the physical response of moving the mouse and selecting an outcome in this experiment.

As reported by prior research that user's mouse position gives an indication of the user's gaze during their online tasks and corresponds to the user's attention, the mouse hover can be interpreted as a visual fixation at the mouse location. In the email sorting task, only one hyperlink is embedded in each email, limiting the hover gestures the participants needed to examine the hyperlink. Examining all the realistic phishing emails from Berkeley Information Security and Policy Phishing Examples Archive [19], we found that similarly to our tasks, most of them only involve a single hyperlink-enabled element per email. Therefore, the hover gestures can be very limited when a phishing email is encountered, making them unsuitable for phishing awareness detection. However, in practice the hyperlink itself remains crucial to identify phishing emails, which is recognized by some participants: 11 out of the 28 of them reported that the URL in the email influenced their decision to categorize an email as *Suspicious*.

Compared with mouse hovers, slow mouse movements exhibited a much higher rate of occurrences. Two features we derived – mouse movement frequency and ratio – are both good indicators of phishing awareness. We attribute this result to the amount of data available and to the mental processes involved. The slow mouse movements refer to how the participant slows down from time to time as the eyes may slow down as well to follow the mouse cursor, suggesting that the participant is focusing their attention on a specific part of the email message. This research established that such cognitive processing mechanism is related to the ability to discriminate suspicious emails. Specifically, we found that if many slow movements, or in other words, frequent fluctuation in mouse movement speed exist in one task, chances are that the phishing email will be misclassified. In comparison, long slow mouse movements with few fluctuations in mouse movement speed often result in correct identification of a phishing threat.

Our findings regarding two qualitatively different processing styles for emails aligns with the Dual Process Theory of human reasoning [22]. Accordingly, we have two qualitatively different systems for making decisions, i.e., System 1 focusing on implicit, automatic reasoning and System 2 which is deliberate and analytic. The former is characterized by intuitive, heuristic decision making that is highly efficient and automatic but also relatively error prone. The latter is more controlled, systematic and effortful.

Given that previous research has suggested that System 2 thinking is linked to better phishing email detection, the long slow mouse movements with few fluctuations in mouse movement speed evident in correct phishing classification in our study may in

fact indicate functioning of System 2. In other words, the users' mouse activity (i.e., long, slow mouse movements) may be reflective of their decision-making quality (i.e., focus on details and deliberation with a high level of attentional control) when determining the legitimacy of an email. Future research into phishing email resistance should examine, empirically, the link between mouse movements, performance on the Cognitive Reflection Test [23] and accuracy in email categorization.

The features we derived, including the slow mouse movement frequency and ratio can both be calculated in real time before the user decision, and thus it would be possible to provide decision support to users based on their own mouse behavior pattern before they are about to click a link in a (phishing) email. The difference observed in slow mouse movement frequency and ratio before a decision can serve many other purposes, including the evaluation of anti-phishing education effectiveness, behavior recommendation for people who may fall victim to phishing attacks, or real-time inspection of phishing emails via crowdsourcing, or even general web browsing safety. For example, such decision support could dynamically block links or prompt the user to exert caution, when the user is deemed vulnerable to clicking on suspicious content, leading to better protection from drive-by download attacks or other malign web links.

There are several limitations to the current study. First, the sample was drawn from participants with a research background, which may not be representative of the wider population. Secondly, we posed the experiment as an email sorting task but didn't specify it was a phishing email detection study, to avoid the priming effect that previous research has shown will artificially improve performance [18]. In the post-experiment questionnaires we discovered one participant focused specifically on the discrimination between high priority and low priority emails (although their data indicated that appropriate phishing email selections were made).

As mouse remains a natural choice for email processing, we only analyzed mouse movement data in this paper. However, other user interaction data, e.g., mobile device use, or physiological data (eye tracking, galvanic skin response, or blood volume pulse) could be used to examine phishing awareness. However, these may lead to a more complicated experimental setup, hence may not be suitable for data collection via crowdsourcing platforms. It is also possible to combine mouse movement analysis with email text/content processing, as compound features to quantify phishing awareness, which will form part of our future work.

5 Conclusion

In this paper, we present a study using mouse movement features to identify subjective phishing awareness. Our experiment results show that slow mouse movements, rather than mouse hover and subjective response time, are good indicators of phishing threat detection. We also discuss the feasibility of using different mouse movement features to identify subjective phishing awareness, which is essential for the development of techniques that can protect vulnerable email users in real time. Our study adds to the general understanding of how humans use the mouse before processing potentially harmful email content, which could provide crucial novel, non-intrusive methods to assist people in phishing prevention and new ways of anti-phishing education through customized phishing detection.

References

1. Drake, C. E., Oliver, J. J., & Koontz, E. J. Anatomy of a Phishing Email. In *CEAS* (2004).
2. Butavicius, M., Parsons, K., Pattinson, M., & McCormac, A. Breaching the human firewall: Social engineering in phishing and spear-phishing emails. *arXiv preprint arXiv:1606.00887*. (2016).
3. Coopers, P. Turnaround and transformation in cybersecurity: Key findings from the global state of information security survey (2016).
4. Dhamija, R., Tygar, J. D., & Hearst, M. Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 581-590. ACM (2006).
5. Downs, J. S., Holbrook, M., & Cranor, L. F. Behavioral response to phishing risk. In *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit* (pp. 37-44). ACM (2007).
6. Parsons, K., McCormac, A., Butavicius, M., Pattinson, M., & Jerram, C. Determining employee awareness using the human aspects of information security questionnaire (HAIS-Q). *Computers & security*, *42*, pp. 165-176 (2014).
7. Conway, D., Taib, R., Harris, M., Yu, K., Berkovsky, S., & Chen, F. (2017). A qualitative investigation of bank employee experiences of information security and phishing. In *Thirteenth Symposium on Usable Privacy and Security*, pp. 115-129. (2017).
8. Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., & Pham, T. School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security* p. 3. ACM (2009).
9. Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 373-382. ACM (2010).
10. Pattinson, M., Butavicius, M., Ciccarello, B., Lillie, M., Parsons, K., Calic, D. & McCormac, A. Adapting Cyber Security Training to Your Employees. Proceedings of the 12th International Symposium on Human Aspects of Information Security & Assurance, Dundee, Scotland (2018).
11. Arroyo, E., Selker, T., & Wei, W. Usability tool for analysis of web designs using mouse tracks. In *CHI'06 extended abstracts on Human factors in computing systems*, pp. 484-489. ACM (2006).
12. Atterer, R., Wnuk, M., & Schmidt, A. Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the 15th international conference on World Wide Web*, pp. 203-212. ACM (2006).
13. Navalpakkam, V., Jentzsch, L., Sayres, R., Ravi, S., Ahmed, A., & Smola, A. Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 953-964). ACM (2013).
14. Hauger, D., Paramythis, A., & Weibelzahl, S. Using browser interaction data to determine page reading behavior. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 147-158). Springer, Berlin, Heidelberg (2011).
15. Huang, J., White, R., & Buscher, G. User see, user point: gaze and cursor alignment in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1341-1350). ACM (2012).
16. Smucker, M. D., Guo, X. S., & Toulis, A. Mouse movement during relevance judging: implications for determining user attention. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pp. 979-982. ACM (2014).

17. Huang, J., White, R. W., & Dumais, S. No clicks, no problem: using cursor movements to understand and improve search. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1225-1234. ACM (2011).
18. Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. Phishing for the truth: A scenario-based experiment of users' behavioural response to emails. In *IFIP International Information Security Conference* (pp. 366-378). Springer, Berlin, Heidelberg (2013).
19. "The Phish Tank | Information Security and Policy." [Online]. Available: https://security.berkeley.edu/resources/phishing/phish-tank. [Accessed: 20-Mar-2019].
20. Anderson, E. W., Potter, K. C., Matzen, L. E., Shepherd, J. F., Preston, G. A., & Silva, C. T. A user study of visualization effectiveness using EEG and cognitive load. In *Computer graphics forum* (Vol. 30, No. 3, pp. 791-800). Oxford, UK: Blackwell Publishing Ltd. (2011).
21. Sweller, J. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295-312 (1994).
22. Evans, J. S. B. Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning*, *18*(1), 5-31 (2012).
23. Toplak, M. E., West, R. F., & Stanovich, K. E. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & cognition*, *39*(7), 1275. (2011).