Bolstering the Human Firewall: Evaluating the Impact of User Training in Mitigating Phishing Attacks

Andrew Simon
College of Computing
University of North Florida
Jacksonville, FL
n00695969@unf.edu

***Abstract*—The specific problem I look to address in this work is to discover the effectiveness of user training on phishing attacks and determine better solutions for the holes found in training. I plan to approach the training, by looking at multiple phishing strategies, such as authority, urgency, intimidation, scarcity, consensus, and familiarity. The training will aim to gather data on how well users can differentiate a malicious actor from a known entity and which of the strategies is the most effective on users. I also plan to include other common signs of phishing attempts in the training that involve URI, email address, or image/logo mismatch.**

The motivation for this work is founded on the reality that human error in the workplace can lead to major cyber security breaches, even when the best practices are in place at the organization to mitigate data loss. Phishing attacks are rapidly evolving and growing in sophistication and frequency. If organizations wish to keep individual and corporate data secure, user education on these cyber threats needs to be robust enough to counter the diverse array of strategies these phishing attacks utilize. The goal of this research is to gather data on how effective modern training solutions are for mitigating successful phishing attacks and discuss solutions to bolster these trainings based on the findings.

I. Introduction

Phishing attacks have emerged as a pervasive and highly detrimental cyber threat, posing significant risks to individuals and organizations. Characterized by deceitful tactics, these attacks manipulate human psychology and exploit vulnerabilities to trick users into revealing sensitive information, such as login credentials, financial data, and personal identification. The impact of successful phishing attacks extends beyond compromised data; it often includes financial losses, reputational damage, legal ramifications, and disruption of business operations. Consequently, organizations face a growing imperative to fortify their defenses against phishing attempts.

One crucial aspect of defense against phishing attacks is user training, which aims to educate individuals on recognizing and responding to phishing attempts effectively. Educating users about the common tactics and strategies employed by phishers is a proactive approach to mitigating the risks associated with such cyber threats. This study delves into the effectiveness of user training in preventing and mitigating the damage caused by phishing attacks. By analyzing the outcomes of targeted training initiatives, this research seeks to shed light on the extent to which informed and trained users can contribute to bolstering an organization's overall cyber security posture against phishing attacks. Understanding the efficacy of user training is essential for tailoring future educational programs, refining security protocols, and fostering a culture of cyber awareness within organizations. Ultimately, this knowledge will aid in the development of more robust strategies to combat phishing attacks and safeguard sensitive information.

This paper will review previous works on the question at hand, providing context and allowing us to build on existing knowledge in the field. These works will highlight key strategies used by others including personality traits, AI-assisted training, and real-time training alerts. This will be followed by an outline of the methodology, discussing the experimental approach of the work, and detailing the pre- and post-training assessments with K-5 teachers focused on email phishing strategies such as authority, urgency, scarcity, intimidation, consensus, and familiarity. The paper shall conclude with an analysis and proposed visualization of test results comparing the data from pre- and post-training assessments.

II. Previous Works

A 2022 study by A. Sumner, et al., suggests that user personality traits have a relationship with the impact on the effectiveness of phishing attacks. Their results show that when participants are measured between Dominant, Influence, Steadiness, and Conscientiousness personality types, that Influence types show the lowest susceptibility to the attacks, whereas Dominant and Steadiness have the highest susceptibility. This research determined that user accuracy increased by 8% after training and also measured their confidence in detection with a determined increase of 6% after training. [1]

S. Back and R. T. Guerette conducted a 2021 study analyzing the effectiveness of place management techniques in reducing phishing attacks and claim that threat actors heavily target organizations and geographic areas with known weak management [2]. Their results actually found a trend in the reverse of the expected direction, as participants who were exposed to the training became more likely to interact with links or provide personal information. The study suggests that these results might have occurred due to these individuals having an increased activity time online compared to the non-treated group.

In 2021, F. Shareveski and P. Jachim addressed the efficacy of using an AI voice assistant (Amazon Alexa in their case) to deliver phishing awareness training to users [3]. The voice assistant was chosen for ease of interaction with participants and utilized for an interaction-based awareness training. The training was composed of three categories: a facts/advice section providing contextual information, scenario-based stories to place participants in the minds of victims, and a gamification-style simulation to engage participants in the interaction with the attacks. The results demonstrated that participants who participated in the Alexa-led, interaction-based training significantly outperformed others when detecting phishing emails, supporting the claim that active engagement with the emails in the training can lead to positive results.

A 2019 study by Xiong, A. et al. suggests that the automatic Google Safe Browsing warning system that users are already receiving phishing alerts from can serve as an opportunity for awareness training in the same message [4]. Their results showed an increase in the ability to detect phishing attempts in future online activity. This evidence supports the conclusion that real-time training embedded into an alert system can be a complementary solution to the typical scheduled team-wide training most organizations incorporate into their security posture.

B. Weaver et al. conducted a 2021 study similar to this work concerning the effectiveness of awareness training on the mitigation of user exploitation of phishing attacks [5]. The primary takeaway from the results of the study is that even with a statistical increase in phishing detection, users still only successfully reported two-thirds of the illegitimate emails flagged as phishing attempts. B. Weaver et al. concluded that this result may be due to the breadth of phishing strategies implemented and the large surface area the attackers are utilizing for exploitation. These results suggest that a more robust phishing awareness training curriculum may be necessary to match the variety and depth of methods malicious threat actors are attacking with.

In 2022, T. Sutter et al. conducted one of the most expansive studies on the effectiveness of phishing awareness training, including over 31,000 participants and utilizing 144 different simulated phishing attacks in their research [6]. Their results show that 66% of users did not engage with phishing attempts after 12 weeks of exposure to anti-phishing simulations. The large sample size allowed T. Sutter et al. to develop a novel manifold learning-powered machine learning model. This model uses NLP features extracted from the emails to help predict the likelihood that participants would fall for a given phishing attempt. A predictive tool of this nature can help education professionals craft more efficient training simulations and better target strategies that users are more susceptible to.

D. Hillman et al's 2023 study focused on the impact of phishing awareness training at an enterprise level [7]. This test was run on an Israeli financial institution with approximately 5,000 participants. The metric measured was similar to the current study, being coined as Click-Through Rate (CTR). Their findings suggest that at an enterprise level participants tended to respond to phishing simulations that were more personalized. This result is theorized to be due to employees trusting individuals who know specific personal details in an enterprise work environment where such details are not commonly known. This study serves to support the idea that phishing awareness training should be nuanced and tailored to fit specific attack vectors in a given work environment.

III. Methodology/Evaluation

This research will be structured in an experimental approach, as I will be delivering a pre- and post-training assessment to my participants. The pre-training assessment will involve a variety of phishing emails incorporating different types and combinations of phishing techniques. My focus will be on authority, urgency, scarcity, intimidation, consensus, and familiarity for behavioral strategies. I will also include emails with slightly altered URIs, email addresses, and logos to test participant attention to detail, as this is often a sign of an illegitimate sender. I will also have a control group that will not receive the training but will still receive the post-training assessment emails. Participants in this research will include elementary grade (K - 5) school teachers in Duval County Public Schools. With my experience as a teacher of 5 years, I believe that this demographic would be an impactful group for this study. We constantly receive emails threatening to renew our certification at risk of being terminated, thanked for our service with a promise of a gift card, or offered a link to desirable resources from a ”fellow member” of a conference or the district. The training of the participants will address all of these techniques in a scenario format and allow participants to work through the sort of language used when confronting a phishing email. The structure of the training will be a PowerPoint presentation, diving into each technique, providing example emails, and having table discussions analyzing example emails. As we go through each technique, I will have participants speak out phrases that align with the strategy. The key focus of the training is to have participants understand the general language being leveraged in the attack. After delivering the content, I plan to have an application section, where participants will craft their own emails, send them to peers, and then analyze what techniques were used in the email. The presentation will close with aspects of a legitimate email and how best to determine the legitimacy of an email.

IV. Results

All results for this test will be mock results, as IRB approval would be needed to interact with the intended test group subjects. The pre- and post-assessments will include different emails, but each pre-assessment email will have a mirroring post-assessment partner for comparison. Each partner email grouping will be constructed with the aim of implementing the same phishing strategy to allow for a close to direct comparison of the results between the two emails. When analyzing the results, the data that is collected from each email will be directly compared to its mirrored partner as well as compared to the entire group of emails as a whole. For each email sent, the participant's response will be measured across three metrics: clicking the link in the email, not interacting with the link, but not reporting the message as a phishing attempt, and reporting the email as a phishing attempt to me. I will produce pie charts for each email, placed side by side with its mirrored partner for a visual comparison of the measurement of all three metrics. To visualize the overall efficacy of the training, I plan to compare the total numbers of each metric to demonstrate the difference of link interaction in the test group. When analyzing the results, conclusions will be made about the efficacy of each phishing technique, leading to discussions of which techniques seem to be most pervasive, and whether any findings might be due to details of the training itself or the overall effectiveness of the phishing technique. The primary aim of the research is to gather evidence of how proper awareness training impacts the user's interaction with phishing links. All results from this experiment will hinge on the participant's interaction with these links and how interaction with links differs before and after the training exercises.

V. Conclusion and Future Work

This study aims to address the critical issue of phishing attacks in the workplace by investigating the efficacy of user training in mitigating the negative impact these threats can cause. Phishing attacks continue to pose significant risks to individuals and organizations, leading to financial losses, reputational damage, legal consequences, and disruptions in business operations. As this threat continues to evolve at a rapid rate, our defenses must also expand at a similar pace. While the results of this research are currently hypothetical, this research hopes to serve as a supportive framework for future studies on this problem.

This work does face several limitations that hinder the capture of the full real-world interactions of individuals and the validity of metrics. As noted, only mock results are to be gathered due to the need for IRB approval for the intended test group. This leads to the lack of real-world data on human interactions with phishing emails. Future work would either gain this approval to test these subjects or shift the focus to a scenario where this type of approval is not needed. While the scope of this study is properly motivated, the target test group of K-5 teachers limits the findings to a narrow demographic. Future work would look to expand this target group to other individuals in the education field at the administration level and teachers in secondary education. This expansion would aim to provide a more robust understanding of the phishing training landscape in the entire industry.

Other considerations for improvements in future work focus on an analysis of cost-benefit considerations and the exploration of varying methods over a longer portion of time. For cost-benefit analysis, it is important to compare the cost of implementing frequent and consistent security training with the current annual financial cost the district faces from phishing attacks. In doing this work, it's necessary to ensure that even if the training is successful it is leading to an overall financial benefit. Our study also only focuses on a singular method of scenario-based training, where a variety of methods could be implemented to provide a more well-rounded education on these phishing threats. Future work will strive to incorporate interactive simulations, gamification activities, and other real-time exercises to better capture some of the psychological stress factors that are involved when dealing with phishing attacks. The pinnacle goal of this work is to provide a training

References

[1] Sumner, A., Yuan, X., Anwar, M., McBride, M. (2022) Examining factors impacting the effectiveness of anti-phishing *JOURNAL OF COMPUTER INFORMATION SYSTEMS 11 trainings. J Comput Inf Syst.*, **6**2(5):975–97. doi:10. 1080/08874417.2021.1955638.

[2] Back, S.; Guerette, R.T. (2021) Cyber Place Management and Crime Prevention: The Effectiveness of Cybersecurity Awareness Training Against Phishing Attacks. 37, *J. Contemp. Crim. Justice*, **4**27–451.

[3] Sharevski, F., Jachim, P. (2022) "Alexa, What's a Phishing Email?": Training users to spot phishing emails using a voice assistant. *EURASIP J. on Info. Security*, **7**(1):7-7. doi: 10.1186/s13635-022-00133-w

[4] Xiong, A., Proctor, R. W., Yang, W., Li, N. (2019) Embedding Training Within Warnings Improves Skills of Identifying Phishing Webpages *SAGE Journals Premier: Human Factors*, **6**1 (4), p.577-595 doi: 10.1177/0018720818810942

[5] Weaver, B., Braly, A. M., Lane, D. M. (2021). Training users to identify phishing emails. *SAGE Journals: Journal of Educational Computing Research*, **5**9(6), p. 1169–1183 doi: 10.1177/0735633121992516

[6] Sutter, T., Bozkir, A. S., Gehring, B., and Berlich, P. (2022). "Avoiding the Hook: Influential Factors of Phishing Awareness Training on Click-Rates and a Data-Driven Approach to Predict Email Difficulty Perception,". *IEEE Access*, **1**0, pp. 100540-100565 doi: 10.1109/ACCESS.2022.3207272.

[7] Hillman, D., Harel, Y., Toch, E. (2023) Evaluating Organizational Phishing Awareness Training on an Enterprise Scale *Elsevier Ltd: Computers & Security*, **1**32 (103364) p. 103364 doi: 10.1016/j.cose.2023.103364