Unlocking the Potential of Simulated Phishing Campaigns: Measuring the Impact of Interaction among Different Human Factors

Francesco Greco¹, Paolo Buono¹, Domenico Desiato¹, Giuseppe Desolda¹, Rosa Lanzilotti¹ and Grazia Ragone¹

¹University of Bari “Aldo Moro”, Via E. Orabona, 4, Bari, Italy, 70125

Abstract

Phishing poses a significant threat to companies and public administrations. Mostly, this attack is perpetrated by exploiting social engineering techniques, such as persuasion principles and emotional triggers. Moreover, technical defenses alone are insufficient to protect organizations from these socially engineered attacks. Therefore, countermeasures that address human vulnerabilities are essential. To this end, we present a framework dedicated to assess human vulnerabilities of employees within an organization by using simulated phishing campaigns. In detail, the proposed work consists of two activities. The first activity explores the interaction between persuasion principles, emotional triggers, and user profiles. Such aspect has not yet been investigated in the literature and it may provide more information on the human factors to which users are most exposed during a phishing attack. The second activity will focus on designing phishing campaigns in which we will measure the effectiveness of emails considering the emotional triggers and persuasion principles used to scam the users, as well as the interaction between these two dimensions and the user personality traits.

Keywords

phishing, human factors, persuasion principles, simulated phishing campaigns, big five personality traits

1. Introduction

Phishing is one of the major cyber threats in our society, being one of the top initial access vectors for cyber criminals [1]. It affects companies and public administrations (PAs) on a daily basis, with employees receiving malicious emails that appear to have been sent legitimately by colleagues, managers, or the IT department asking them to take immediate action such as clicking on a link or opening an attachment. In these attacks, criminals exploit users’ human factors, which increase their susceptibility to falling victim [2].

Given the important role that human factors play in the success of these attacks, phishing cannot be addressed solely on a technological level (e.g., by implementing automated phishing detection mechanisms). For this reason, organizations typically conduct “white hat” phishing
campaigns to assess the company’s susceptibility to phishing attacks. By sending fake emails, companies can estimate their exposure to attacks in terms of how many employees clicked on the phishing links in these emails [3, 4, 5].

While simulated phishing campaigns provide organizations with a tool to quantitatively assess their vulnerability to phishing attacks, they fall short in assessing the human factors at play when these attacks are successful [6]. For example, personality traits of an employee strongly impact their susceptibility to phishing [7]. Furthermore, the effectiveness of a phishing campaign can be significantly influenced by the nature of emails it comprises. Persuasion principles [8] are psychological techniques often used in phishing attacks, which can increase the user’s susceptibility [9, 10, 11]. Phishing emails also often leverage emotional drivers, such as creating a sense of urgency or fear, to increase the likelihood of users falling victim [12, 4, 13]. Although previous work has explored how individual user differences [14, 15, 3, 16] or the use of social engineering techniques [12, 4, 13, 9, 10, 11] may affect the susceptibility to phishing attacks, to date no approach comprehensively measures the interaction between (i) the users’ profile (in terms of personality traits), (ii) the use of persuasion principles and (iii) adoption of emotional triggers in phishing emails.

Our research proposes a new defensive solution in the context of the Italian national project DAMOCLES (Detection And Mitigation Of Cyber attacks that expLoit human vulnerabilitiES), which aims to develop a framework for the Italian Public Administration to assess human factors in cyber incidents and mitigate their impact through security awareness and customized user training. The ongoing work presented in this paper includes two main contributions. The first part provides insight into the relationships between persuasion principles, emotional triggers, and personality traits. To achieve this, a large-scale study will be conducted with over 1000 participants exposed to various emails that correspond to different combinations of persuasion principles and emotional triggers. The study results will reveal the most critical combinations of <persuasion principle, emotional trigger, personality trait> that make phishing emails most effective for certain users. The second part of our research will build on the knowledge gained in the first study to create more precise simulated phishing campaigns. These campaigns will enable companies and organizations to evaluate the susceptibility of their employees to emails that include (or exclude) the most effective phishing techniques for their profiles.

Understanding the individual vulnerabilities of the employees can lead to take more effective decisions from an organizational perspective, such as providing them with specific support in the form of personalized training material to address their vulnerabilities [17, 18]. Furthermore, with the right support and training, employees can become a valuable asset to the organization and an effective line of defense against phishing (i.e., also known as crowd-sourced phishing detection) [3].

The paper continues as following: Section 2 presents the related work on social engineering techniques commonly used in phishing email and user’s assessment; Section 3 discusses the 2-phase approach we propose to measure the effectiveness of phishing emails and to assess employees with a simulated phishing campaign; Section 4 draws conclusions and presents future work of the project.
2. Related Work

The causes of a phishing email’s effectiveness can be boiled down to two main factors: the characteristics of the email itself and the characteristics of the recipient.

Phishing emails often use Persuasion Principles to deceive users into clicking on phishing links or disclosing personal information [8, 11]. Cialdini [8] identifies 6 persuasion principles that are widely explored in the social engineering literature: authority, scarcity, liking, social proof, reciprocation, and consistency. The use of persuasion principles can ultimately affect the effectiveness of a phishing email, making it generally more deceptive to users [9, 10, 11, 19, 14]. Ferreira and Teles [11] identified a list of persuasion principles that are most prominent in phishing attacks, which include, in addition to authority and reciprocation, integrity and strong affect.

Phishing emails often exploit core emotions: curiosity (or anticipation), fear (or anxiety), greed (or desire), anger (or annoyance), joy (or excitement), confusion (surprise), and empathy (or compassion) [12, 4, 13]. This is usually accomplished by including emotional drivers (or triggers) that manipulate users and cause them to make irrational decisions [20, 21]. For instance, when experiencing sadness, individuals tend to gravitate toward high-risk/high-reward options, whereas those in anxious states prefer low-risk/low-reward choices [22]. In general, individuals who are under the influence of “visceral influences” do not consider the ramifications of their actions and seek immediate satisfaction of their visceral desires [23, 24].

Emails that employ these social engineering techniques (either alone or in combination) are typically more deceptive and can more easily lead users to become victims [9, 10, 11]. The quality of a phishing email can be measured using the Phish Scale developed by NIST [25]. This tool can help assess the difficulty of an email, in average, to be detected. This scale considers two main aspects: the email cues (i.e., the observable characteristics of an email such as language, presentation, correctness, etc.) and the alignment with the user premises (i.e., how closely an email matches the work roles or responsibilities of the recipient). The stronger an email’s premise alignment and the fewer cues it has, the more difficult it is to detect it as a phish. The difficulty of a phishing email can be classified in three categories, based on the number of cues: many cues (less difficult), some cues (medium), few cues (more difficult).

Regarding the characteristics of the recipient (i.e., the user), there are a number of human factors that play a critical role in influencing the susceptibility of users to phishing attacks [2, 26], including lack of knowledge, lack of resources, lack of awareness, norms, and complacency. Another important factor that affects an employee’s susceptibility to phishing is their personality [27, 28, 29]. Personality is undoubtedly a very complex factor to model; in the literature, the most widely adopted model in the literature is the Big Five Personality Traits [30], which describes an individual personality according to 5 traits: Openness, Agreeableness, Conscientiousness, Extraversion, and Neuroticism. These traits have been shown to be stable over time, and universally identifiable regardless of language, race, culture, or gender [31]. Other human factors, such as gender and age may play a role in influencing a user’s phishing susceptibility, but findings in literature are often contrasting [2]. Finally, emotions also play an important role in the susceptibility of users to fall for phishing attacks [32, 33]. The effectiveness of persuasion principles can be traced back to specific human factors. For example, extroverted individuals are particularly susceptible to the liking and scarcity persuasion principles, while agreeable
individuals are particularly susceptible to the authority principle [34, 29, 35].

Simulated phishing campaigns are typically used to deliver embedded training material [36, 37, 38, 3]: employees who fall victim to a fake phishing email are redirected to a training page that explains to them the risks of phishing attacks and why they should not trust the phishing email they received [36, 37]. This approach has proved to be much more effective than traditional frontal lessons, especially when the training material is embedded in warnings [38]. However, Lain et al. [3] conducted a large-scale long-term simulated phishing campaign in a company and gathered evidence that embedded training does not make employees more resilient to phishing, but rather may actually make them more susceptible.

3. Assessing users’ phishing vulnerabilities with simulated campaigns

The solution we propose in this paper will be carried out in two different and sequential activities:

1. Design of a user study to measure the three factors that may influence users’ susceptibility to phishing, i.e. personality traits, persuasion principles and emotional triggers;
2. Design of a simulated phishing campaign based on the results of Activity 1, i.e., the correlations between the three factors and users’ phishing susceptibility. A web platform will make it possible to apply the most effective combinations of these factors to test users with challenging fake phishing emails.

3.1. Activity 1: User study to discover correlations between user profiles and persuasion techniques

To discover correlations between users’ profiles, persuasion principles, and emotional triggers, we need to construct a knowledge base with data about the phishing susceptibility of users (each with their own personality traits) to different phishing techniques. Therefore, a user study serves as a means for gathering the data. This will be done by firstly collecting data about the users to profile them according to the Big Five personality traits model by administering the NEO Five-Factor Inventory-3 [39], a 60-item questionnaire to measure their personality traits according to the Big 5 model. After a user profile of the employee is generated, the users will be exposed to a set of safe and phishing emails. The phishing emails included in the study will be crafted by applying different combinations of <persuasion principle, emotional trigger>. The persuasion principle will be one of the 6 persuasion principles (i.e., authority, scarcity, reciprocation, social proof, liking, consistency), while the emotional trigger will be one of the 7 emotional triggers (i.e., curiosity, fear, greed, anger, joy, confusion, empathy), leading to a total of $6 \times 7$ unique combinations.

In addition, to improve the external validity of the study, the topic of the phishing email is also varied, as done in [4]. The fake emails can be crafted by, e.g., following the modus operandi of Gallo et al. [14], starting from real phishing emails to include a unique combination of persuasion principle and emotional trigger.
For each of the 42 combinations, 3 variants are generated to have a more solid knowledge base. The variants are crafted to be of different levels of difficulty to include an additional dimension in the measurements. To objectively rate the overall level of difficulty for an average employee to detect an email, the Phish Scale [25] is used with the following scores: (1) low level of difficulty (cues category = "Many"), (2) medium level of difficulty (cues category = "Some"), and (3) high level of difficulty (cues category = "Few"). This results in \(42 \times 3 = 126\) fake emails that will be sent during the study; a fake phishing email contains a link that, when clicked, redirects an employee to a landing page where they are debriefed about the fake phishing email. At this point, the information about which employee clicked on the phishing link is saved. To avoid overloading users with too many emails, each of them will be exposed to a subset of the emails (e.g., 10 safe emails, 10 phishing emails). Eventually, each of the 42 combinations will be administered to an equal number of users.

3.2. Activity 2: Design of a simulated phishing campaign to measure more in-depth human factors

The findings from the previous study will highlight the most important interactions between persuasional principle, emotional trigger, personality trait that, for particular users, maximize the effectiveness of phishing emails. Building on the insights from the first study, the second activity of the research presented in this paper will develop more accurate simulated phishing campaigns. Through these campaigns, companies and organizations will be able to assess how vulnerable their staff members are to emails that contain (or don’t contain) the most successful phishing techniques specific to their profiles. To better illustrate this activity, we introduce a scenario that describes how this approach could be practically applied in a PA. The scenario is described below:

1. The National Institute for Social Security ("INPS", in Italian) is a PA with about 20,000
employees; faced with the ever-increasing risk of cyber-attacks, its IT director decided to
improve the organization’s defenses against phishing attacks by assessing the specific
human factors to which its employees are more susceptible to, with the subsequent goal of
addressing the specific deficiencies of employees through customized training programs.
2. 7 days in advance, employees are informed about the simulated phishing campaign that
will be conducted and its objective. They are also informed of the need to collect data
that can be used to create a profile, assuring them that their digital profile won’t be
directly traceable to them. These measures limit the extent of the ethical implications
that naturally come with a similar approach.
3. An initial model of the PA’s employees is created by administering the NEO Five-Factor
Inventory-3 [39], a 60-item questionnaire to measure their personality traits according
to the Big 5 model. To assess the employees’ initial ability to correctly recognize and
respond to phishing attacks, the survey-based Phishing Awareness Questionnaire [40] is also
administered. Finally, the employees’ risk-taking behavior is measured with the Balloon
Analogue Risk Task test [41], as higher risk-taking behaviours can negatively influence
phishing susceptibility [32]. The questionnaires are administered to the employees in the
workplace to ensure a more controlled environment.
4. A simulated phishing campaign has been designed to assess the long-term susceptibility of employees to phishing attacks, spanning a duration of 3 months. In this context, personalized phishing emails will be utilized, with a comprehensive approach tailored to each user. Specifically, a total of 30 emails will be meticulously crafted for every personality trait, drawing upon the top 10 combinations of persuasion principles and emotional triggers associated with that trait. Each of these combinations will generate 3 distinct emails varying in complexity. Consequently, throughout the campaign period, users will encounter the 30 emails tailored to the personality trait identified as most influential for them. This approach ensures a targeted exposure to a spectrum of psychological tactics employed in phishing attempts, facilitating a robust evaluation of susceptibility over time.

5. The simulated phishing campaign is launched. On Day 1, the first email is sent. The phishing link in the email redirects any employee who falls victim to a page where they are debriefed about the fake phishing email. Here they are reassured that no consequences will be taken against them, and that the data they will submit will be kept anonymous (in line with what is done in [4]). The causes that led them to click on the links are investigated by asking open-ended questions about (i) how did the email made them feel, to qualitatively collect their self-reported emotions (as in [4]), and (ii) what led them to click on the phishing link (as in [42]).

6. After Day 1, the remaining emails are sent at intervals of 3 days to avoid predictability (with an average of one email every 10 days). Furthermore, the minimum delay between one phishing email and another is necessary to avoid priming the employees to more secure behavior after exposure to a debriefing message (i.e., to reduce the expectancy effect [43]).

7. A dashboard can show the current situation for all employees by reporting, for each fake email sent, the percentage of employees who clicked on the phishing link. The employees’ personality traits are also displayed to highlight the correlation between them and the phishing susceptibility.

8. At the end of the simulated campaign, the company can address the individual vulnerabilities of each employee (whose identity remains undisclosed) by automatically delivering customized training/security awareness materials. For example, if an employee is found to be particularly vulnerable to the Authority principle used in IT communication emails, they are provided with examples of fake emails that include that specific persuasion technique; training material additionally suggests security measures to double check the sender’s identity (e.g., the address of legit communications). Moreover, they are provided with vital information such as some of the company norms (e.g., that the IT department will never ask employees to provide their passwords) and useful contacts to consult when they feel a communication is suspicious, so that they do not resort to alternative, less secure, sources.

4. Conclusions and Future Work

This work is part of the research conducted within the Italian national project DAMOCLES. The main project ultimately aims to develop a framework for the Italian PAs to assess and mitigate
human factors in cyber incidents. This would make it possible to uncover factors that may be overlooked in current cybersecurity training approaches and ultimately lead to better protection in these organizations. One line of action to enhance user protection is customized training that addresses the employees’ individual vulnerabilities.

This paper contributes to the first step of assessing the user vulnerability by proposing a methodology based on simulated phishing campaigns. This phase is only a part of a broader, iterative approach, that involves a continuous assessment-training process to progressively reduce an organization’s vulnerability to phishing (this methodology is also referred to as “Agile Phishing” by [4]).

Future work will include testing the proposed approach with user studies in a controlled setting. Moreover, much effort will be put in studying how to craft customized training material to specifically address one or more vulnerabilities. Another interest aspect to be investigated is the expectancy effect, i.e., the extent to which an employee is primed towards a safer behavior when they are aware that a phishing campaign is being conducted in the organization; analyses to assess this bias may involve comparing the click-rate in emails with similar difficulty sent with different delay from each other. While the proposed approach can certainly bring many benefits to organizations in their fight against phishing, there is a major ethical problem with collecting employees data in a safety critical context. Being able to identify each user and their actions with phishing emails could put their jobs at risk. Therefore, future works must include the development an anonymization mechanism to protect the user’s identity, while allowing targeted interventions to improve their susceptibility to phishing attacks.

Acknowledgments

This work has been supported by the Italian Ministry of University and Research (MUR) and by the European Union - NextGenerationEU, under grant PRIN 2022 PNRR “DAMOCLES: Detection And Mitigation Of Cyber attacks that exploit human vulnerability” (Grant P2022FXP5B) – CUP: H53D23008140001. The research of Francesco Greco is funded by a PhD fellowship within the framework of the Italian “D.M. n. 352, April 9, 2022” - under the National Recovery and Resilience Plan, Mission 4, Component 2, Investment 3.3 - PhD Project “Investigating XAI techniques to help user defend from phishing attacks”, co-supported by “Auriga S.p.A.” (CUP H91I22000410007).

References


[42] A. J. Ferguson, Fostering e-mail security awareness: The west point carronade, Educause