Phishing Threats in the Age of Social Media: A User-Centric Approach

Ruxandra BADESCU
Bucharest University of Economic Studies, Romania
badescuruxandra@gmail.com

Even today, phishing attacks continue to have a high success rate, using impersonation to trick users into revealing sensitive information. Despite the fact that the number of technology users is growing rapidly, this does not seem to provide individuals with sufficient knowledge and experience to understand cybersecurity concepts. In addition to this, not only technological shortcomings are exploited, but also human characteristics, weaknesses, that make users vulnerable when faced with potential threats. This study aims to provide an overview of phishing attacks and the rapid development of detection techniques and phishing tools in a context where cybersecurity education or training is not compulsory. The article manifests the importance of education, regardless of the technological changes that influence the daily lives of countless individuals. The present research was conducted using a literature analysis and a survey.
Keywords: *Phishing, User susceptibility, Behavior, Human weaknesses*
DOI: 10.24818/issn14531305/28.4.2024.07

1 Introduction

In the age of social media, where anyone can become a content creator, people are much more likely to trust what they see online. Users tend to let their guard down when it comes to news, skincare advice, breakfast recommendations or fashion tips. Instead of using critical thinking, they get swayed by different trends, not knowing how vulnerable they have become. The term *privacy* has lost its meaning as numerous people share details about their daily lives online, with little to no regards for the consequences. Considering these observations, cyber criminals find it easy to engage in activities such as social engineering, fraud, scams, or phishing, as they can easily obtain personal information about individuals, and target their vulnerabilities. Navigating social media or conducting online searches does not mean an individual possesses digital skills. It is much more than just that. Digital literacy means being aware of the medium you navigate through, understanding how it works and identifying possible threats in conjunction with own vulnerabilities. Security is, therefore, the most crucial aspect of digital activities. Since users can be as young as a few years old, security measures should be taught step by step as individuals grow and understand more about the digital world.

The present thesis consists of a number of chapters, each addressing a different aspect of the research on phishing threats and the user-centric approach to mitigating them. The introduction delivers an overview of the presented topic, including the background, motivation, objectives, and scope of the study. The literature review chapter examines the existing literature on phishing threats. The following research methodology chapter describes how the research was conducted, and the methods and procedures used to collect and analyze data. The following sections describe the phishing domain, the human and technical vulnerabilities it exploits, as well as a proposed efficiency score for several types of phishing threats. The subsequent part analyses the current developments in phishing detection methods. The following chapter outlines the evolution of phishing techniques and the most recent examples that prove the ability to bypass security and pass as legitimate emails. The final chapter regarding the study's results presents the analysis and findings of the survey conducted as part of the research. The conclusions chapter presents a summary of the research's primary findings and makes recommendations for future study in the areas of phishing threat mitigation and phishing awareness.

2 Literature review

The increasing prevalence of phishing attacks in all forms and varieties in the last few years is a matter of concern that needs to be addressed. This is a clear sign that more and more people need to be made aware of the potential consequences of a phishing attack. Phishing awareness training sessions should be organized regularly in every organization to ensure a high level of understanding of potential online threats. On the other hand, ordinary users should be considered as important and receive training periodically. Education represents the most effective means of combating any form of cybersecurity attack. In the context of phishing attacks, it can be argued that the most effective means of protection is user awareness. This enables individuals to recognize and avoid scams, frauds and enticing offers, which are often employed by phishing attacks. Despite organizations' strong password policies, users generate weak passwords, often reuse them, and write them down on post its, as revealed by this systematic literature review [1]. This indicates that human factors are exposing the organization to cyber-attacks. In this case, companies need to invest in employee training. Simulated phishing attacks combined with informal training have proven to be an effective way to combat phishing attacks.

User awareness is a critical factor in cybersecurity and a topic that many are studying, as the goal is to clearly identify the reasons why people are being deceived. Several factors have been found to influence user awareness of phishing attacks. On the one hand, there are the human variables such as emotions, self-awareness, self-control, self-deception, motivation. On the other hand, users' IT skills, security awareness, and PC usage experience are also variables that influence people's awareness when exposed to phishing emails. The authors of [2] highlighted that the weaknesses of human behavior are an essential aspect to consider in developing relevant defense mechanisms. One of the most popular and effective ways to decrease the impact and damage produced by phishing attacks is to educate and train users. Training should be tailored to each age group and gender, as young people aged 18-25, older individuals and women are more likely to fall victim to phishing attacks. Another study mentioned in [2] highlights the characteristics of a phishing email that convince users to fall for these scams. The urgency of the message, the promise of a reward, the false credibility of the sender and their message, and the unfavorable outcomes are what make users fall into the trap. These features constituted the foundation for the survey questions regarding the potential indicators of a phishing email.

The study presented in [3] has identified a significant distinction between individuals with and without IT expertise in the context of phishing awareness. It was conducted in an academic community where phishing attacks were delivered to 1,350 students from different majors, such as social sciences, engineering, IT, Natural and Mathematical Sciences. The results show that students with no knowledge of phishing showed a lower susceptibility rate, while social studies majors had the highest click rate and engineering, and IT majors had the lowest. Interestingly, the study found that users more knowledgeable about phishing were often more susceptible, a result without a clear explanation. This highlights that even IT-trained students remain vulnerable to phishing.

In the paper [2], the authors mention a study that found no significant difference in susceptibility to phishing attacks between everyday technology users and occasional users. This is a concerning result because it seems that the everyday use of technology does not automatically provide users with adequate security knowledge. Therefore, an effective solution to this problem would be to educate all users on cybersecurity best practices, regardless of their technical expertise and experience. Another paper mentioned in the study [2] concludes that people pay the most attention to the subject and body of an email, even though the sender's email address can be an essential clue in the case of a phishing attack. Therefore, future studies should find ways to draw users' attention to this type of cue, as it is of the utmost importance. The questionnaire

used for the present study includes a section on the respondent's vigilance when dealing with incoming emails. This allows for the analysis of habits in conjunction with the level of awareness of phishing attacks. The following study [4] evaluates the effectiveness of security awareness and education programs in large organizations. A field investigation was conducted in a German company comprising 409 employees. The objective was to evaluate the effectiveness of a security awareness program in the context of phishing over a period of time. In addition, the effectiveness of four reminder measures was evaluated after the initial training program. The users' capability to accurately identify phishing emails greatly improved right after the training sessions and remained high enough until four months later. However, this effect diminished after six months since the initial training, indicating the need for a reminder measure. The study [4] demonstrated that a training session concerning phishing attacks containing multiple interactive examples and videos was the most effective, with a minimum of six months.

User awareness remains a significant factor in determining the cyber well-being of an institution. It is evident that innovative technologies cannot be the sole countermeasure against cybercrimes. Human factors play a significant role in users' susceptibility to falling victim to cybercrime. Furthermore, the continuous development of cybersecurity technologies takes into consideration human vulnerabilities in order to prevent and protect users' behavior. Consequently, further research is required in this field. This paper aims to provide an overview of phishing attacks, highlighting human and technical vulnerabilities exploited by cyber criminals. A brief overview of the detection schemes currently used to identify phishing attempts is provided. The case studies, along with the alarming evolution of phishing attacks, highlight the battle between cybersecurity professionals and malicious actors to stay one step ahead. A survey is used to assess the awareness and vigilance regarding phishing emails and email management using a target group of cybersecurity students from the master’s program. Regardless of the effectiveness of these mechanisms, education and training in basic cybersecurity concepts remain the pillars that can never be corrupted.

3 Research methodology

This section describes the research approach used in this dissertation. The study includes both a survey and a comprehensive literature review. The survey offers empirical data on user phishing awareness and email management habits. At the same time, the literature study attempts to synthesize current knowledge on phishing attacks, the vulnerabilities they exploit, and the latest phishing detection techniques. This approach ensures a rigorous examination of the topic. The review of existing literature on the evolution of phishing threats and current innovative detection and mitigation techniques involved an in-depth search of academic papers, books and respected online sources. Google Scholar, IEEE Xplore and ScienceDirect were the databases used to extract academic papers. Keywords included "phishing threats", "phishing awareness", "phishing attack detection", and "phishing attack mitigation". The search was conducted on literature published in various journals, academic publications, and conference proceedings between 2009 and 2024. This ensured both an exhaustive overview of the domain and the inclusion of recent developments and trends in phishing advancements, as well as detection and mitigation techniques. The selected sources that compose the references list were chosen based on their level of relevance, reliability, and ability to provide a complete understanding of phishing threats. The following themes were identified: phishing strategies, psychological aspects that affect phishing success, and technological countermeasures. The conclusions of each source were noted, and the data was synthesized to identify recurring themes, knowledge gaps, and newly concerning phishing threat areas.

The objective of the survey is to investigate whether young professionals who have received formal IT training are aware of the prevalence of phishing attacks in the digital

realm and possess a nuanced understanding of their objectives and methods. In addition, the email habits of the subjects are evaluated, as well as the importance they ascribe to various email cues, including the sender's email address, the subject line, hyperlinks, and attachments. The appropriate research method considered was conducting a survey at the Bucharest University of Economic Studies, targeting students from the IT&C security master's program in their first and second years of study. The survey was created using Google Forms. It was distributed to students via email and Slack. The data is collected and processed anonymously by the researcher. The questionnaire for master's students has two primary purposes. First, it is designed to investigate the habits of respondents when managing their email inboxes. Respondents are requested to evaluate the extent to which they perform thorough checks of the most relevant email cues on a Likert scale, ranging from one (representing "Never") to five (representing "Always"). Secondly, the master's survey includes an assessment of respondents' awareness of phishing emails. It contains a series of statements that represent the characteristics of a phishing email, including urgency, misspelt words, requests for personal information, tempting offers, a generic greeting (Dear Customer), malicious links and suspicious attachments. The respondents are required to evaluate each statement, which depicts a characteristic of a phishing email, on a Likert scale, where one represents "Not at all likely" and five represents "Very likely."

As cybersecurity master's students are presumed to have a basic understanding of security threats and vulnerabilities, the objective is to evaluate both their habits when managing emails and their awareness and degree of comprehension regarding potential phishing emails. The email habits question can address one possible issue: whether students apply the concepts they study in their daily lives. The target group can be considered homogeneous. It is comprised of cybersecurity master's students in their first and second years of study, aged twenty-two and above, with the majority having a background in computer science. It is assumed that they possess the same level of understanding of computer science and are proficient computer users.

4 A comprehensive overview of phishing attacks

Phishing is a computer crime that employs social engineering and psychological manipulation techniques to exploit human vulnerabilities, such as greed, curiosity, and fear. Cybercriminals' goal is to manipulate users to expose their confidential data, such as usernames, passwords, payment, or bank account information [5]. Not only that, but the attackers may also have the intention of installing malware programs to gain access to the victim's computer and to claim a certain amount of money, as in ransomware attacks. Phishing is derived from the word 'fishing' in the dictionary, as it closely resembles the activity of attempting to catch fish. The fishing rod or fishing net is represented by the tool developed by the attacker to expressively steal sensitive information. Phishing attacks are intended to get the victim's curiosity with enticing offers, gifts, or other tricks. When it comes to fishing, the fisherman knows that the prey is hungry and is looking for food, whereas in phishing attacks, the attackers are well aware of human nature and its weaknesses. To gain credibility, phishing attacks often use impersonation of legitimate companies or even people. Attackers use a variety of communication channels, such as email, short messaging services, and voice communication, to perform this type of attack. [6] claims that phishing attacks started in the 1990s, with the first recorded attempt involving hackers impersonating AOL employees to steal credit card information and passwords using instant messages and emails. This method allowed them to hijack the victims' accounts with ease.

Phishing attacks are performed by what they are called in the literature and professional language threat actors, also known as cybercriminals. They perform harmful actions in the digital realm, exploiting computer and system vulnerabilities and human weaknesses [7]. Their ultimate goal is financial gain by tricking victims into revealing sensitive and

personal information or performing an action that results in installing malicious software on the victim's computer. Before launching an attack, threat actors will engage in various activities, such as social engineering techniques and system reconnaissance, to gather more information about the victims.

5 Original solutions: efficiency in phishing and attack performance

Figure 1 illustrates the efficiency of phishing attacks. Multiple criteria were considered to assess the efficiency of the types of phishing attacks described. First, we evaluated the effort attackers invest in crafting convincing phishing emails or messages in the respective category of phishing. This includes the time required to gather information about the target through social engineering techniques to personalize the emails to sound plausible and to persuade the victim to perform the desired action. The complexity of the techniques and methods used to carry out the respective type of attack was considered. In the case of a successful attack, the potential damage, such as financial and reputational loss and business disruption, was estimated. In addition to this, the phishing attacks were evaluated from the perspective of number of individuals targeted. Nevertheless, it was taken into consideration that, for instance, spear-phishing or whaling attacks, which often target one person, can be very effective due to the personalized content, social engineering and psychological manipulation techniques used to persuade the victim. In this case, psychological manipulation was considered a factor for the success of the attacks.

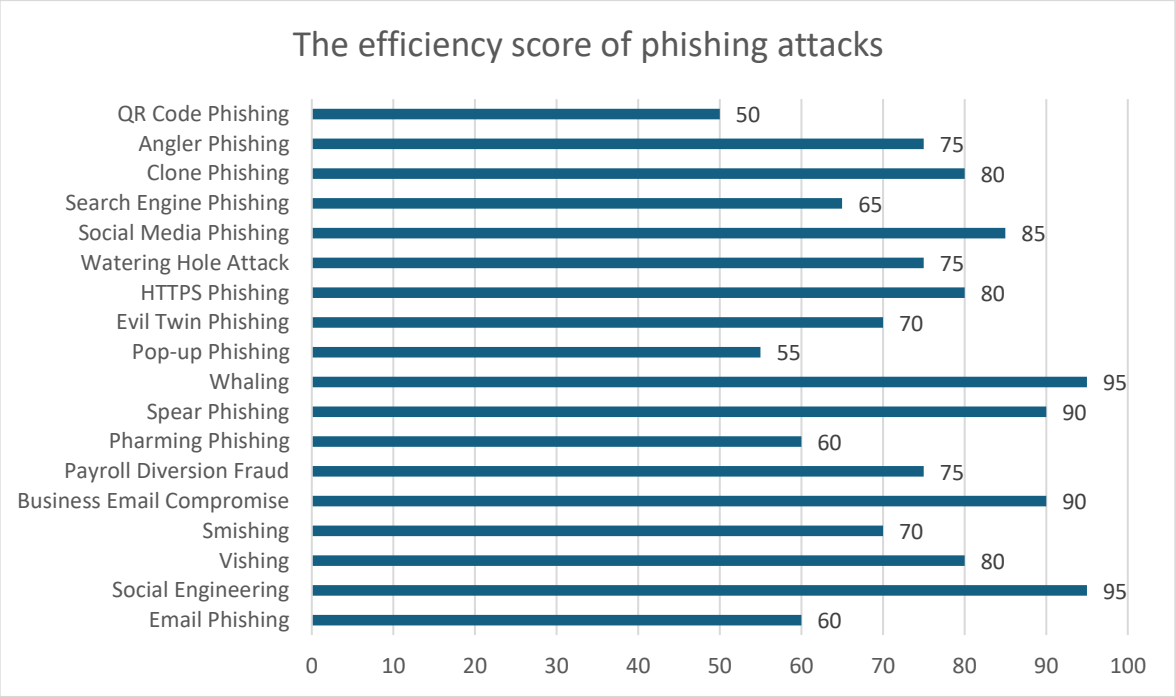

Fig. 1. Phishing types and their assigned efficiency score

The attacks that have a score between 90 and 100 are social engineering, business email compromise, whaling and spear phishing. Although the target group is not large, the effort the attackers put into gathering personal information about the target, their habits, and their weaknesses is greater than in any other type of phishing. The attacks include psychological manipulation techniques that are successful due to the amount of personal data the attackers have about the victim. The potential damage in this case is quite significant; in most real cases, large financial and reputational losses have been identified. The technical complexity of these attacks is relatively low compared to others, primarily using spoofed email addresses and phishing links.

The next category is phishing attacks with an

efficiency score between 80 and 90: vishing, HTTPS phishing, social media phishing, and clone phishing. The time is significantly reduced, as emails are no longer highly personalized. This increases the target group, especially in the case of social media phishing, where a single post or comment can reach a large number of users. Psychological manipulation techniques are still used, but in a simple form, exploiting people's trust in social media platforms, phone calls, HTTPS websites and previous email conversations. The potential damages from these attacks are either sensitive information provided by the victims or the installation of malicious software programs on their devices.

The subsequent category consists of phishing attacks with an efficiency score of 70 to 80: smishing, payroll diversion fraud, evil twin phishing, watering hole attack, and angler phishing. In the case of the last three types, the target group can be quite large, including passersby in a mall, employees from a certain department and unsatisfied customers who post their complaints online. Smishing has a limited target group, and payroll diversion fraud usually targets one victim at a time, as it requires significant time, and resources to be performed. Evil twin phishing is more complex than the others in terms of the technical setup required to execute the attack and relies on users' trust and habit of connecting to public Wi-Fi. A watering hole attack requires much preparation, from identifying the websites visited daily by a particular group of employees to identifying the vulnerabilities and methods to exploit them. It relies on the trust that people and organizations have in certain professional websites. Conversely, smishing, payroll diversion fraud and angler phishing, all rely on the manipulation of people's emotions to prompt the victims to act in accordance with the attacker's instructions.

The following category contains three types of phishing attacks: email phishing, pharming phishing, and search engine phishing, with an efficiency score between the values 60 and 70. Their target audience is quite diverse, potentially reaching hundreds and thousands of users, especially in the case of email phishing and search engine phishing. Pharming phishing is quite complex in terms of the techniques and methods used to carry out this attack and the poisoning or manipulation of the DNS table. Search engine phishing requires successful SEO manipulation to get the phishing sites to appear at the top of search results. Email phishing is much less complex than other forms of phishing, as the attack is not personalized and does not target a specific group of people or individual but rather a large pool of random individuals.

The final category includes pop-up phishing and QR code phishing, with scores between 50 and 60. These attacks are not at all personalized and targeted at a specific individual. However, they use general manipulation techniques to deceive victims, such as computer infection issues or account problems (pop-up phishing), or they pique the passersby's attention and curiosity with enticing offers (QR code phishing). Consequently, the time allocated for the planning stage is relatively short. The target group is quite large and aims to reach website visitors and individuals who pass by billboards and street advertisements. Pop-up phishing requires more technical knowledge, especially as browsers have updated their security mechanisms and users have become more aware of this type of scam. QR code phishing is quite simple, as it requires a phishing website and a QR code generator for the link. They are not to be overlooked, but highly personalized emails using social engineering manipulation techniques are much more successful nowadays.

6 Understanding the weaknesses phishers exploit

In today's digitalized landscape, phishing attacks remain a significant and prevalent threat. Instead of evading technological security measures, phishing focuses more on exploiting human vulnerabilities using complicated social engineering techniques. The majority of email filtering systems are not a hundred per cent effective when it comes to detecting and blocking phishing emails. The comprehension of social engineering and psychological tricks employed in phishing attacks

will help to identify these threats better, enabling users to protect themselves.

Greed is still the primary human weakness that is responsible for the success of a significant number of phishing attacks. In addition to greed, curiosity and fear are used to manipulate the victims and persuade them to disclose personal information. The goal is to make them act irrationally and on impulse without thinking of the potential consequences. In the same way, curiosity can be piqued by SMS messages regarding the distribution of free samples of a new product launched by a famous brand. Being driven by the fear of missing out on important opportunities, the victims are lured into filling in a form with their personal information in order to receive the promised goods. People can be more vulnerable to phishing attacks when it comes to love or family relationships, the strongest bonds in human nature. Attackers will tailor their messages or speeches to include information about a family member living abroad who has been in a car accident or is seriously injured and needs large sums of money [8]. The information in the message is accurate; the family member is indeed working abroad, and the family is faced with a tough decision. Large sums of money were stolen in this way by a group of convicts who were in prison and their accomplices who collected the money from the victims.

Trust is the subsequent psychological element exploited by phishers through impersonation of well acknowledged entities, such as banks, government agencies, popular brands, trusted people with authority. Attackers create a false sense of legitimacy by using the real logo or an almost identical one and maintaining the same tone when writing the email. When crafting the phishing email, attackers may use personal information about the victim collected from social media platforms to make the email and request sound more plausible and legitimate.

Authority is another psychological element that is exploited by phishing attacks, as most people, especially employees, are more likely to follow the request of a senior executive without questioning their solicitation. Therefore, impersonating such an important and trusted figure is a common tactic used in phishing attempts to ensure the success of the attack and the fulfilment of the request, whether it be the transfer of money or the disclosure of important information.

Social engineering techniques are often employed in the design of phishing emails. One of the most prevalent techniques is pretexting, creating a context in which the victim can be persuaded to divulge sensitive information. An easy way to convince people to reveal personal information is in the context of a questionnaire distributed by a so-called researcher. Cognitive biases make users more susceptible to phishing attacks. An example of a cognitive bias is often seen in IT professionals who are overconfident in believing they can easily identify phishing attempts and are unlikely to be deceived by such tricks. By letting their guard down, they become more susceptible to phishing attacks, as shown in the paper [2].

Furthermore, context and timing are two elements that significantly impact the success of a phishing attempt. If the attackers manage to gather information about a team's or department's schedule, they can launch their attack during the most stressful period of time, such as when a new application is going into production or the final days before the launch of a new product. People's vigilance is lower when they are in stressful situations or exhausted from too much work and concentration, which affects their judgement. The herd effect is also a factor that ensures the success of phishing attacks. Emails or messages are crafted to induce the idea that many other people either joined the challenge or the survey or completed the request. The fear of missing out will make the victims participate in the survey or challenge, not realizing that this is just a manipulation technique. The urgency of a phishing email is also a factor that compromises people's judgement and leads them to act on impulse, especially if the email contains negative consequences for not complying, such as losing money or having their account closed.

The techniques used to craft phishing emails are advancing along with the fast development

of technology and artificial intelligence. Emails are written using generative artificial intelligence technology to create convincing messages that no longer contain errors, using persuasive words and phrases [9]. Moreover, this new technology can process pieces of information from multiple sources to create believable spear-phishing and whaling emails with an urgent tone and an immediate call to action. Some experiments demonstrated the effectiveness of this kind of phishing email, which was more likely to trick individuals compared to those written by humans.

Furthermore, artificial intelligence is used in what is known as deep fake technology, which is used to create audio, video, and images of specific individuals [9]. These are often employed in phishing attempts for impersonation, deceiving victims into believing that they are hearing and seeing a real person. This is expected to become a significant challenge for cybersecurity, as these methods are very difficult to analyze and intercept. The effectiveness of this method is demonstrated by a real-life case where the CEO of a UK firm was deceived into wiring a large sum of money as a request from his superior, whose voice was used in the impersonation act.

7 Understanding phishing detection techniques

The following section presents general methods and techniques used to detect phishing attempts. The most efficient way of preventing catastrophic consequences is to have an approach based on both technical countermeasures and non-technical means of prevention when it comes to malicious emails or messages [10]. A non-technical way of preventing email phishing from affecting a device or an entire infrastructure is to educate users to recognize phishing threats. Training should be designed based on the target group of trainees so that the materials and concepts discussed are easy to understand based on their digital literacy. Multiple approaches would make training more effective. The training should start with a theoretical background regarding phishing threats, their particularities, and clues that may indicate a phishing attempt. The trainees should receive a set of questions they should ask when opening each email to assess the legitimacy of the email. On the other hand, individuals and organizations can use several methods and techniques to detect phishing attacks and protect users from this type of threat. The first method implies the use of blocklists and allow lists for phishing domains used in the most prevalent phishing attacks [10]. In an allow listing approach, all legitimate websites are saved in a database, letting users to access only those. However, this solution can be impractical in large corporations. It is impossible to know all the websites employees need to accomplish their work. New websites are often mistakenly considered suspicious and blocked because they do not appear in the database. Block listing saves known phishing websites, domains, or URLs in a database that is updated in real-time [10]. Users can check to see if the domain they are redirected to is known as phishing or not. This method reduces the traffic volume on a phishing website by over 90 per cent. Block listing uses automated detection of suspicious websites to update the database and to protect users from being deceived into entering sensitive information on phishing websites. Block listing is integrated with various security tools on the browser side. Another method is to compute a similarity coefficient between known legitimate websites and suspicious ones. If the value surpasses a set threshold, it means that the suspicious website is indeed used for phishing and impersonates the legitimate website [10]. However, this method is quite inaccurate and inefficient since small changes in a phishing website's appearance may radically change the coefficient and pass as legitimate while tricking the users into thinking they are browsing a legitimate website.

Phishing detection techniques that use machine learning have been researched over the last few years, as they are much more effective and flexible in detecting phishing attempts accurately. They have been used to detect phishing attacks distributed via email and to examine and compare the websites behind the URLs contained in the email and the

legitimate websites they impersonated. Moreover, machine learning techniques can prevent zero-day attacks only if the data models are trained on relevant high volumes of data and fine-tuned regularly to ensure accuracy. However, deep learning detection models for phishing emails and websites have shown higher precision than regular machine learning architecture [11].

An email security gateway is a piece of security software that monitors email traffic [12]. It is placed at the border between the infrastructure and the Internet to filter emails entering or exiting a company's infrastructure. This software uses signature analysis and techniques involving machine learning models to identify and block potentially malicious or sophisticated phishing emails. An email security gateway includes a sandbox that disarms the URLs and attachments contained in an email to observe their behavior. Moreover, it can examine the content of an email closely, determining if any malicious code is present and then removing it from the initial email. In the end, the user receives a clean version of the original email that does not represent a threat to their device or environment. Sandboxing is crucial in preventing malicious emails from entering a company's infrastructure or reaching a user. It is a safe environment in which URLs and attachments are opened, and their behavior is monitored for suspicious or malicious characteristics. Based on the verdict given by the sandbox, the email may be sent to the user or quarantined if it is found to be malicious. An extensive Secure Email Gateway has integrated a retrospective analysis component, part of a post-delivery protection feature. This feature is able to identify malicious emails after they have been delivered to the user, send alerts to the cybersecurity team to resolve the problem, and remove or quarantine the malicious emails that have been delivered. Furthermore, having integrated a data loss prevention system, this security software can help prevent the exposure of confidential information outside the company.

8 Artificial Intelligence: The future of phishing detection

Generative AI is a new approach in which machine learning models process and train on significant amounts of data. This way, generative AI can produce well-written software code, high-value text, and impressive artwork [13]. Both cybersecurity professionals and cybercriminals have embraced this emerging technology to conduct their respective daily tasks. There are several key areas where generative AI has proven to be beyond any other technology. For instance, it can help keep attackers up to date with the new zero-day vulnerabilities of various systems. Second, since it can be used for writing convincing, eye-catching posts or emails, it can be used to write persuasive phishing emails that are personalized to a target group or a specific individual. Spelling and grammar errors have been an indicator of phishing emails for years, but with the help of generative AI, it can write error-free emails that have a convincing tone. Moreover, with the introduction of chatbots on most e-commerce websites, conversations with artificial intelligence have become widespread in today's digital world. Cybercriminals can use this new feature to impersonate highly educated individuals and fool users into believing that they are having a conversation with a real person. It is alarming that a user will not be able to distinguish between a human and a robot interlocutor in an online interaction.

If generative AI helps attackers carry out successful attacks, then the same technology should be used as a method of detection for such new emerging threats. There are several ways to integrate artificial intelligence into email security systems to improve detection. First, it can help to create user profiles for all employees within an organization and all external parties with whom they communicate via email. It can learn the writing styles and tones of each individual, therefore enabling the detection of any anomalies. In the case of a phishing email impersonating a senior executive, the system can detect that it does not match the style of the real person.

Secondly, artificial intelligence techniques

can be employed to detect emails in which the attackers are asking for money or sensitive information. Likewise, it can be used to detect whether the attachments sent along with the email have pieces of code hidden within them. In addition, artificial intelligence can work against itself to detect whether the text of an email has been designed using generative AI techniques. On the other hand, the AI technology integrated within the email security appliance can use flagged phishing or malicious emails and their corresponding metadata to train the machine learning models to improve their accuracy.

9 Case study 1: Cisco's approach to email security

As was presented in the first part of this paper, users should be aware of the threats they may face when opening an email. Nevertheless, in this case, extensive training and awareness campaigns are needed to create the habit of constantly checking email indicators. Cisco's Secure Email Threat Defense is a solution to this problem [13]. It performs an assessment and then takes a decision for each element of an email. The sender's email address receives a score based on its reputation, as well as the email's source URL. The subject and body of the email are scanned for specific words and phrases that may indicate the presence of a phishing email. Each attachment is carefully scanned and opened in a sandbox; its content is scanned for suspicious words and phrases, and if it contains URLs, then these are analyzed as well. Spam email is accurately identified and quarantined or deleted if it is discovered to have malicious content. Furthermore, Cisco products are integrated with Talos, the leading supplier of security intelligence, and each new update from Talos is automatically integrated into this product.

Cisco's Secure Email Threat Defense integrated AI to identify particularities of attacks by processing large volumes of data from the telemetry provided by most Cisco security systems. Much of the valuable data represents breadcrumbs left behind by attackers after performing various types of exploits. Threat data is another asset that helps enhance the security of a system by ensuring that the system is up-to-date and ready to detect and mitigate the majority of security threats. It contains information on the patterns of the latest threats, their signatures, what is behind successful breaches and how to respond best in each situation. Automated security systems enable real-time protection and mitigation of attacks, as well as valuable data and real-time alerts about the incidents that may occur.

The following passage will provide a short presentation of some efficient techniques used to detect business email compromise attacks, emphasizing the importance of artificial intelligence integration in email security systems. Business email compromise attacks do not require the distribution of malicious links or attachments. Instead, they use social engineering techniques to understand the relationships between several employees in a company and then manipulate them into revealing confidential information or performing a required action, such as transferring money. Therefore, only the text of the email, the tone of the writing, and the intent behind the message can be carefully examined. This type of analysis is called sentiment analysis [14]. It considers hierarchical relationships to understand if the request in the email is appropriate.

Most business email compromise attacks require the account takeover of a senior executive or other employee with authority and power in the company. After successfully compromising the account, the attacker observes the email threads for some time, waiting for the right moment to intervene and divert a payment. An exclusive way of detecting BEC attempts would be to detect these small, insignificant changes in an email thread where the attacker tries to infiltrate the conversation. The security team would be notified of this potential attempt in real time, and the email would be redirected to quarantine and blocked so that it would not reach the user.

The previous type of examination of emails is called conversation and thread analysis. It focuses on the intent of the message and the relationship between the two parties communicating. Any slight differences in the tone used for writing or in the length of the message are

taken into consideration in this kind of analysis. These represent crucial indicators that the person behind an email is not the real one but an attacker using impersonation. Advanced detection techniques for account impersonation and takeover involve creating social graphs that map the relationships between employees [14]. The graphs can point trust relationships and calculate a risk value that indicates the possible impersonation of a trusted party. These detection techniques also consider the history of all emails sent by critical employees to create a pattern for their communication habits and to highlight the people with whom they had the most conversations.

Cisco developed a novel technology to address these blended threats. It is known as CASE (Context Adaptive Scanning Engine) [15] and is designed to detect both traditional, simple phishing and spam emails, as well as advanced email threats. It also successfully detects viruses and malware in a single scan forty-two hours before their signature is available. This innovative technology considers the latest obfuscation mechanisms employed by cybercriminals. Therefore, emails must be analyzed in their full context, preferably with techniques similar to the human mind. CASE is built on mechanisms that involve highly developed machine-learning algorithms that can mimic human reasoning to assess an email's legitimacy. This reasoning is based on four essential components of an email: the identity of the sender, the destination of the hyperlinks, the way the message was designed, and the content of the message.

A sender's identity is assessed based on a **reputation score**, a concept introduced in 2003 by IronPort. The sender refers to the IP address, not just the email address. The reputation is evaluated by considering more than one hundred twenty characteristics, such as the volume of the traffic email and country of origin. The reputation score ranges from -10 to +10 [16]. For values between -10 and -3, the email is blocked or deleted. The email is accepted for values between -3 and -1, but other incoming emails from the same sender are halted. All emails with a reputation score of the sender between -1 and +10 are considered legitimate. This reputation-based filtering is able to stop above eighty per cent of the incoming spam traffic, thus increasing the availability and efficiency of the email infrastructure. The legitimacy of the hyperlinks in the email is assessed based on **web reputation filtering**, a novel perspective introduced by IronPort. It considers more than forty-five parameters, such as how old the destination domain is, the reputation of the IP address behind the URL, and the reputation and state of the host behind it. The content of the email is evaluated in its whole context, together with the sender's reputation, the legitimacy of the hyperlinks and the design of the email. The content analysis is built on the newest machine-learning techniques and can determine the categories of email content: financial, publicity, and spam-related content.

10 Case study 2: Exploring Trend Micro's Deep Discovery Email Inspector

A new solution meant to help companies detect and block spear-phishing attacks, ransomware, and other types of advanced threats is the Trend Micro email security appliance *Deep Discovery Email Inspector* [17], a security solution that carefully examines each email for malicious content, such as URLs or attachments, and performs a content analysis to assess its legitimacy. The investigation process combines various methods. First, the email is analyzed for known threats. After this, all the URLs embedded in the message, along with the attachments, are transmitted to the *Virtual Analyzer* sandbox, a custom sandbox used for investigating the potential behavior of such elements. *Virtual Analyzer* can scan URLs or files [18]. It effectively opens password-protected documents or archives using custom dictionaries created from all the keywords in the email message. It accesses the URLs, investigating redirects, downloads, suspicious connections, and other kinds of malicious behavior or particularities. Moreover, DDEI has embedded a real-time protection mechanism called **time-of-click protection**, meaning that the moment a user clicks on a link, behind the scenes, a real-time analysis of the respective URL is performed,

ensuring once more the safety of the user and its device. In addition, pattern-based and heuristic scanning mechanisms are employed by the *Advanced Threat Scan Engine* to detect zero-day vulnerabilities, embedded malicious code, document exploits and other types of known vulnerabilities. Furthermore, modern machine learning techniques are implemented to process threat data and conduct an in-depth file analysis to identify new types of advanced threats. Similar to Cisco's web reputation scoring, Trend claims to have one of the grandest reputation databases in the world [18]. The particularity of this web reputation service is that it sets a score for each page or link of a specific website and does not classify the entire website. Most of the time, only a group of pages or links are compromised from an entire website. Therefore, only those specific elements will be considered malicious and blocked, as the reputation will suffer changes over time.

Trend Micro implemented a novel technology that uses artificial intelligence to identify business email compromise attempts by creating a writing style DNA for senior executives, the people most likely to be impersonated by cyber criminals [19]. This new technology can examine an email for over seven thousand writing style particularities in less than a second. More than half a thousand sent emails are needed as input for AI technology to determine an executive's writing style accurately. Potential BEC emails go through this AI examination as the last layer of a comprehensive anti-spam and anti-virus analysis. If it is found that the writing style of an incoming email does not match the previously identified style, the email is marked accordingly. The recipient is warned not to take any action before confirming the request sent to the senior executive. This novel technology demonstrated little to no false positive alerts, being extremely efficient during the beta testing period.

11 The future of phishing: emerging techniques

The advances in the techniques employed in phishing attacks have determined the development of more intricate detection and prevention technologies. Complex and modern machine learning algorithms have been trained using tremendous amounts of data to identify phishing emails. Nevertheless, these new technologies that comprise the most advanced email security systems are still not a hundred per cent effective, and a few phishing emails still reach users. The scope of this section is to point out why some phishing emails are not detected, the methods implemented that help them bypass security systems, and the measures that security professionals can take to improve their detection systems.

The first approach to increasing the chances of a phishing email being considered legitimate is to improve the sender's reputation [20]. Emails from genuine and trusted email domains, such as *gmail.com* for Gmail or *mail.com*, have a good sender reputation and are free to use. The host from which the email is sent may have an IP address with a good reputation from Google, Cloudflare or Amazon. These IPs are considered legitimate most of the time, and cybersecurity teams cannot blacklist them because of the reputation of the company and because of the uncertainty of the actual reputation of that IP address. In addition, attackers can use email servers like those from Amazon Web Services to increase the chances of bypassing security appliances.

Most phishing emails contain a URL which directs users to a phishing domain. Several bypassing techniques have been identified [20]. Attackers use trusted domains such as Google and Amazon to hide malicious content. In some cases, the URL in the phishing email redirects to a file-sharing service, such as Google Drive, OneDrive, or SharePoint [21]. Second, the real domain a URL leads to can be hidden by shortening it. Alternatively, the URL can indicate a legitimate domain but redirect to a malicious domain. In both cases, the destination domain is concealed, and unless a sandboxing service is used, the website's real status cannot be evaluated. A captcha can be used to cover content that may indicate the true nature of the phishing website, such as a login form [20]. This will prevent anti-spam tools from analyzing and discovering malicious content. Phishing emails contain

malicious attachments, which are verified using two kinds of techniques. First, the file reputation system contains hashes of known documents used in phishing attacks. The email is considered legitimate and delivered to the user if the computed hash value is not found in the database. Second, the analysis can be done using a sandbox, which opens the attachments and inspects their content for potential links or code. Modern malicious programs are able to detect if they are executed in a sandbox-like environment [20]. Attackers developed sandbox circumvention mechanisms. The malware identifies the number and type of processes running, and libraries utilized, enabling the detection of the particularities of a sandbox environment. When detecting the sandbox environment and a debugger, the malware hides the malicious code, or it can postpone the execution.

Unfortunately, two- or multiple-factor authentication is not a security measure that can protect accounts from being compromised [22]. A unique phishing-as-a-service kit can effortlessly compromise Google and Microsoft 365 accounts; the victims only have to enter their username and password on a counterfeit login page. The tool compromises push notifications, phone call validation, and OTP codes sent by various apps or SMS for a Microsoft account. Unlike ordinary phishing attacks, when the victim can solve the issue by just changing the password of that account, the exposure persists in attacks where this new tool is involved. That is because the session cookies are saved on the attacker's server. This enables the hacker to replay the session and access the account even after changing the password.

Machine learning models proved to be much more efficient than the simple blacklisting methods because they can adapt and learn continuously based on the data that is used for training. Moreover, these models are automated and use resources efficiently. However, attacks are constantly developed to bypass machine learning detections, using methods such as obfuscation and polymorphism in order to conceal the malicious behavior. Therefore, these models require high-quality and relevant data for each customer to recognize targeted modern phishing attempts. In addition, the study [23] found that the performance of machine learning models is highly influenced by the features and classifiers used. Thus, the models have to be constantly trained and fine-tuned based on suitable data to keep pace with the rapid development. This is a challenging process. If the data used for training is publicly available, it can be manipulated to contain concealed vulnerabilities. Attackers cannot only try to bypass security analysis but sabotage machine learning models by introducing synthetic data to mislead the training process, exploiting the weaknesses of machine learning algorithms, influencing to their desire the feedback loops and interfering with the settings of hyperparameters.

As seen previously, emerging phishing technologies and toolkits help attackers succeed in their malicious attempts to steal credentials. To cope with such threats, companies must ensure that their machine-learning detection models can analyze both perceptible and imperceptible content [24]. The analysis has to be enhanced with multiple layers of artificial intelligence-based detection that can learn the natural way people communicate via email. Generative AI has a significant role in this scenario. The detection framework should consist of multiple layers based on machine learning models with enhanced learning capabilities. Most phishing detection systems rely on careful URL analysis, which is, in most cases, a clear indicator of a phishing attempt. The domain mainly consists of an enumeration of letters without a real meaning. In extenso, the URLs have domains with multiple extensions. Some of them are easily identified by users as phishing because of their peculiar appearance, while some require more work, such as a machine-learning model, to be rightfully classified. Unfortunately, recent discoveries [25] note that cybercriminals may have the necessary skills and knowledge to circumvent URL classification models by generating examples that trick these models. The author of the paper [25] tried to reproduce such innovation by using Generative Adversarial Networks to produce phishing URLs. The

constructed URLs were able to evade the Blackbox phishing detection models, although some of them were created to analyze intra-URL resemblance using complex procedures. Generative Adversarial Deep Neural Networks can help create realistic data automatically using a semi-supervised approach. Using such an approach, phishing URLs were successfully created, with a considerable number being able to deceive both simple and complex machine learning detection techniques.

12 Experimental results analysis: phishing insights from survey data

As previously stated in the research methodology, this paper includes a survey-based study to inquire about the students' awareness and caution in managing their incoming emails, which can sometimes be part of a phishing scheme. The structure and questions that compose the survey are detailed in the methodology chapter. The target group comprises cybersecurity master students in their first and second years of study, aged 22 and above. The study aims to confirm or infirm that cybersecurity students navigate the Internet in their daily lives according to their understanding after studying the potential threats that roam in the digital realm.

The survey was conducted among 70 students and collected 47 responses. Descriptive statistics were computed for each survey statement. Each of them could be assessed on a Likert scale from 1 to 5. The outcome is the following: most means are above 4, which indicates a high phishing awareness and email vigilance among students. This high level of awareness and vigilance is a proof of the effectiveness of their education and training in cybersecurity. Specifically, students were most cautious when checking the sender's email address, the links embedded in the email body and the attachments. Moreover, when identifying email phishing threats, the most recognized indicators were requests for sensitive information and gift offers, with mean scores of 4.81 and 4.68, respectively. In contrast to these results, some statements received a relatively low score, indicating medium awareness levels. This is the case with the statements regarding the presence of grammar and spelling mistakes in phishing emails and the generic greeting at the beginning of the message. The first statement had a mean score of 3.91, while the second showed the lowest mean awareness score of 3.17. These two areas present opportunities for improvement in the students' awareness of email phishing threats. The low scores are due to the respondents' habit of receiving and sending email messages that were not spell-checked beforehand. In addition, some companies, online shops, and brands may use a generic greeting when composing promotional emails. Therefore, since it may be used in legitimate emails, the generic greeting does not represent a vital clue in identifying phishing emails.

The survey data was subject to a correlation analysis, which revealed several noteworthy relationships between different behaviors of email phishing awareness. With a correlation coefficient of 0.8, a strong positive correlation was observed between the habit of carefully checking the sender's email address and the embedded links. This indicates that students who are cautious and check the sender's identity also check the links rigorously, meaning that they do not have the trust to access the links, even after ensuring the user is legitimate. A strong positive relation, indicated by a correlation coefficient of 0.7, was remarked between the verification of the email content and the email subject, which explains that these actions often occur together.

Moderate correlations between statements regarding the characteristics of phishing emails were found as well. A notable association was determined between the urgent tone of a phishing email and the account suspension tactic used by phishers, with a correlation coefficient of 0.75. This means the students know these two are clear indicators of a phishing threat, which employs such tactics to deceive users and manipulate their emotions. This is a clear indicator of how students perceive phishing threats. Emails that requested confidential data presented a correlation of 0.6 with emails that offered gifts or prizes. These correlations may also indicate what kind of

phishing attempts respondents encountered in their online experience.

The analysis of the survey's results highlights the necessity of cybersecurity education, not only for professionals but for all users. While the means for each statement were above 4, indicating a high level of awareness, they were not 5, suggesting that there is still room for improvement in users' vigilance when managing their emails. This restates the importance of continuous education and training in cybersecurity awareness. The analysis's results point out areas for improvement, such as the generic greeting included at the beginning of a phishing email and the request to call a phone number provided in the body of the email. Although they may be overlooked, these issues observed after the data analysis prove that cybersecurity education should be promoted and included for all users, regardless of their digital literacy.

13 Original solutions: remedial measures for raising users' phishing awareness

Despite organizations' efforts to have the most advanced security software to protect against phishing attacks, a few malevolent emails still reach users. Therefore, it is crucial to invest time and money in highly developed email security appliances and training sessions for employees. In order to be effective, training lessons should be organized regularly, taking into consideration the rapid development of new types and forms of phishing attacks. The training materials should contain theoretical concepts and practical, interactive examples. Research [4] has shown that video materials effectively increase people's understanding of phishing threats.

Employees in a company are very likely to receive such training, if not regularly, then from time to time. Nevertheless, the rest of the users, the young individuals who browse the Internet and social media platforms all day and those who work in different environments, remain susceptible to phishing without a chance to receive proper cybersecurity education. Therefore, some other ways must be found to educate ordinary users. Cybersecurity education should be integrated into the ICT (Information and Communication Technology) subject taught in middle and high school. In this case, the target group comprises children and teenagers aged 10 - 19. For children younger than ten years old, still in primary school, cybersecurity can be integrated into their science subject. In this case, the primary school teacher should receive proper training to add some basic concepts about online security into their lessons. Multiple projects have been developed to help children, teenagers, parents, and teachers better understand online security threats. Cybershield [26] offers various courses for teachers and young individuals and has partnerships with schools, offering teaching materials and providing opportunities for young people to learn more about cybersecurity and even help them pursue careers in this field. A national program, #SigurantaOnline [27], offers six free courses for children, parents and teachers, which help increase individuals' ability to recognize online threats and adequately understand and manage the Internet's risks.

To target the rest of the population, official institutions such as banks and telecommunications companies should advise and raise their clients' awareness of the potential threats they are exposed to. As mentioned in the previous sections, most phishers impersonate banks and use techniques such as vishing or smishing to deceive users and manipulate them to divulge personal information. Some financial institutions have started to develop cybersecurity awareness campaigns to inform users of the potential threats. They informed the users of the bank's legitimate activity, how it manages clients' data, what different kinds of attacks look like, and what techniques cybercriminals employ. The remaining category comprises older people, much more susceptible to vishing. To reach such a category, it is crucial to take into consideration their habits and lifestyle. Most individuals in this category watch the news and prefer some news channels over others. Therefore, the easiest way to reach them is to conduct an awareness campaign at the beginning of the news bulletin, when individuals pay the most attention.

With all the technological advances in

machine learning algorithms and the use of AI in detection and prevention security systems, education remains an asset that attackers cannot control or steal. If the security systems meant to protect users fail, education is the weapon that can help protect both users and organizations.

14 Conclusions

The present thesis analyzed the phishing domain from various perspectives to present an overview of the multiple facets of this cyber threat. Phishing attacks have developed significantly over time, using various approaches to exploit human vulnerabilities and technical issues. Social engineering and psychological manipulation techniques have proven to be the most effective in deceiving users to disclose personal information or download malicious programs. On the other hand, phishing attacks are not detected or prevented from reaching the user for various reasons. In the case of an ordinary user with little knowledge about cybersecurity, no significant phishing detection technologies may be installed on their devices, making them vulnerable to receiving phishing emails and messages. Specialized users with medium to high cybersecurity understanding may use different anti-phishing tools with different levels of accuracy. However, these detection models proved to be inaccurate, considering the hasty development of phishing tools and techniques. An efficiency score was proposed for each phishing type based on several criteria. The effort to craft the phishing email or message, the time required to gather information about the victim, the number of victims targeted, the complexity of the methods employed to reach the target and trick them, and the final damages in case the attack was successful were taken into consideration while assigning the efficiency score. Five phishing categories were identified. Simple and more complex phishing detection techniques were briefly described to provide an overview of the techniques employed by cybersecurity specialists to prevent phishing attacks from reaching systems. Since phishing threats have evolved and become increasingly sophisticated, the challenge is to identify and characterize these new forms of phishing and implement elaborate detection schemes that can recognize such attempts. Complex machine-learning models are the future of anti-phishing detection tools if they are trained on relevant and clean data, fine-tuned, and enhanced with multiple features regularly. Artificial intelligence is an excellent opportunity for cybersecurity professionals and cybercriminals to improve their tactics and upgrade their tools with automated features. AI enables attackers to craft well-written and plausible phishing emails and messages, bypassing the classic content analyzer and being considered legitimate from this point of view. In the same way, cyber professionals can enhance the features of their detection models with AI features, such as sentiment analysis, to detect business email compromise attacks. Furthermore, the case studies of Cisco email security appliance and Trend Micro Deep Discovery Email Inspector focus on the newest techniques great security vendors use to provide accuracy and efficiency in detecting advanced phishing threats. Each of their approaches is carefully described, focusing on the machine-learning or artificial intelligence modules adopted to enhance performance. The following chapter addresses the issue of the continuous development of phishing attacks having the scope of evading sophisticated anti-phishing tools. This particular study is intended to increase awareness of the multiple facets of phishing, understanding that along with the current evolution of technology, rapid changes happen on the dark side of the digital realm. Cybersecurity specialists are left in a challenging situation and must understand and keep up with the current complex techniques and tools to evade detection models. On the other hand, users must focus on basic cybersecurity education, which enables them to correctly perceive the digital environment, identify threats and protect themselves.

The study comprises a survey that was addressed to cybersecurity master students to assess their phishing awareness and vigilance when managing their emails. Email is the most used phishing vector, so it was considered essential to investigate how students

analyse their incoming emails. This approach can estimate their vulnerabilities and susceptibility. The analysis of the survey results indicated a high level of phishing awareness and cautiousness in email management. The clues of the generic greeting, as well as the request to call a phone number provided in the email body, do not represent relevant phishing indicators for most respondents. Remedial measures to increase general users' awareness of cybersecurity have been proposed. The subchapter provides different strategies for integrating primary cybersecurity education for children, young people and elders. Education remains the only weapon that can be effective regardless of the context, the complexity of the phishing threat or the ability of anti-phishing detection tools to prevent such threats from reaching the users. It represents the only asset that cybercriminals cannot corrupt.

Further recommendations in the context of the rapid development of technology and artificial intelligence are that the focus must be on including these new technological advancements in complex detection schemes. This approach may ensure satisfactory accuracy while considering the fast evolution of phishing tools, which enable attackers to carry out their attacks with minimum effort and in a very short time. Despite these challenges, the focus should also be on providing education and training for all users, regardless of their digital literacy. If it is well-tailored for each user group and enhanced with practical examples and relevant videos, training can fill the gaps where detection mechanisms are not efficient anymore.

References

[1] A. Smith, M. Papadaki, and S. M. Furnell, "Improving Awareness of Social Engineering At tacks," Bento Gonçalves, Brazil, 2013, pp. 249–256. doi: 10.1007/978-3-642-39377-8_29.

[2] G. Desolda, L. S. Ferro, A. Marrella, T. Catarci, and M. F. Costabile, "Human Factors in Phishing Attacks: A Systematic Literature Review," ACM Comput Surv, vol. 54, no. 8, pp. 1–35, Nov. 2022, doi: 10.1145/3469886.

[3] A. Diaz, A. T. Sherman, and A. Joshi, "Phishing in an academic community: A study of user susceptibility and behavior," Cryptologia, vol. 44, no. 1, pp. 53–67, Jan. 2020, doi: 10.1080/01611194.2019.1623343.

[4] B. Reinheimer et al., "An investigation of phishing awareness and education over time: when and how to best remind users.," in Sixteenth USENIX Conference on Usable Privacy and Security (SOUPS'20), Virtual Conference: USENIX Association, Aug. 2020, pp. 259–284.

[5] K. Jansson and R. von Solms, "Phishing for phishing awareness," Behaviour & Information Technology, vol. 32, no. 6, pp. 584–593, Jun. 2013, doi: 10.1080/0144929X.2011.632650.

[6] P. Gillin, "The History of Phishing Attacks," Verizon Business. Accessed: Feb. 05, 2024. [Online]. Available: https://www.verizon.com/business/resources/articles/s/the-history-of-phishing/

[7] B. Lenaerts-Bergmans, "What is a Cyber Threat Actor?," CrowdStrike. Accessed: May 16, 2024. [Online]. Available: https://www.crowdstrike.com/cybersecurity-101/threat-actor/

[8] "Patru escroci au reusit sa stranga peste 100.000 de euro, din puscarie. Victimele, pacalite prin cea mai cunoscuta metoda," Stirileprotv. Accessed: May 20, 2024. [Online]. Available: https://stirileprotv.ro/stiri/actualitate/patru-escroci-au-reusit-sa-stranga-peste-100-000-de-euro-din-puscarie-victimele-au-fost-pacalite-prin-mai-cunoscuta-metoda.html

[9] T. Shloman, "The Psychology of Phishing: Unraveling the Success Behind Phishing Attacks and Effective Countermeasures," Trellix. Accessed: May 20, 2024. [Online]. Available: https://www.trellix.com/blogs/research/understanding-phishing-psychology-effective-strategies-and-tips/

[10] R. Alabdan, "Phishing Attacks Survey: Types, Vectors, and Technical Approaches," Future Internet, vol. 12, no.

10, p. 168, Sep. 2020, doi: 10.3390/fi12100168.

[11] V. Sapkal, N. More, and R. Agme, "A Briefed Review on Phishing Attacks and Detection Approaches," in 7th International Conference on Innovations and Research in Technology and Engineering (ICIRTE-2022), Mumbai-22, India, 2022. doi: 10.2139/ssrn.4108334.

[12] "What is a secure email gateway?," DarkTrace. Accessed: May 25, 2024. [Online]. Available: https://darktrace.com/cyber-ai-glossary/secure-email-gateway-seg

[13] F. Smith and G. Bridgers, Advanced Email Threats For Dummies, Special Edition. Hoboken, New Jersey: John Wiley & Sons, Inc., 2024.

[14] CloudFlare, "How to Stop Business Email Compromise Threats. Advanced techniques for fighting," 2022. Accessed: May 15, 2024. [Online]. Available: https://cf-assets.www.cloudflare.com/slt3lc6tev37/4sivQSVRKbmDGL6VqSLLmM/3f62dd0c9dd7534ca9b43d012ddd8635/Cloudflare_Area_1_Whitepaper_-_How_to_Stop_Business_Email_Compromise_May_2022.pdf

[15] Cisco, "Cisco Email Security: Understanding Context Adaptive Scanning Engine (CASE)," Cisco. Accessed: May 26, 2024. [Online]. Available: https://www.cisco.com/c/en/us/support/docs/security/email-security-appliance/215533-cisco-email-security-understanding-cont.html

[16] Cisco, "Email Security Using Cisco ESA. Technology design guide," Cisco. Accessed: May 26, 2024. [Online]. Available: https://www.cisco.com/c/dam/en/us/td/docs/solutions/CVD/Aug2013/CVD-EmailSecurityUsingCiscoESADesignGuide-AUG13.pdf

[17] Trend Micro, "Trend MicroTM DEEP DISCOVERYTM EMAIL INSPECTOR. Stop targeted email attacks leading to data breaches or ransomware," Trend Micro. Accessed: May 26, 2024. [Online]. Available: https://www.trendmicro.com/en_us/business/products/user-protection/sps/email-and-collaboration/email-inspector.html?modal=s8c-btn-learn-more-4ae27b

[18] Trend MicroTM, "A new solution | Trend Micro Service Central," Trend Micro. Accessed: May 26, 2024. [Online]. Available: https://docs.trendmicro.com/en-us/documentation/article/deep-discovery-email-inspector-51-online-help-a-new-solution

[19] C. Taylor, "Using writing style checks to stop CEO impersonations," Trend Micro. Accessed: May 27, 2024. [Online]. Available: https://www.trendmicro.com/en_us/research/18/i/stop-impersonations-of-your-ceo-by-checking-the-writing-style.html

[20] Mantra Team, "How hackers bypass anti-spam to deliver phishing emails," Mantra. Accessed: May 30, 2024. [Online]. Available: https://www.mantra.ms/blog/bypassing-antispam

[21] Perception Point, "Techniques Used by Hackers to Bypass Email Security Solutions," Perception Point. Accessed: May 30, 2024. [Online]. Available: https://perception-point.io/news/technique-used-by-pass-email-security-solutions/

[22] S. Ikeda, "MFA Bypass Kit simplifies phishing attacks on Gmail and Microsoft 365 accounts," CPO Magazine. Accessed: May 30, 2024. [Online]. Available: https://www.cpomagazine.com/cyber-security/mfa-bypass-kit-simplifies-phishing-attacks-on-gmail-and-microsoft-365-accounts/

[23] S. Kapan and E. Sora Gunal, "Improved Phishing Attack Detection with Machine Learning: A Comprehensive Evaluation of Classifiers and Features," Applied Sciences, vol. 13, no. 24, p. 13269, Dec. 2023, doi: 10.3390/app132413269.

[24] R. Fernandez, "New phishing attacks bypass machine learning Security: Experts talk solutions," Techopedia. Accessed: May 30, 2024. [Online]. Available: https://www.techopedia.com/new-

phishing-attacks-bypass-ml-security-expert-talk-solutions
[25] A. AlEroud and G. Karabatis, "Bypassing Detection of URL-based Phishing Attacks Using Generative Adversarial Deep Neural Networks," in Proceedings of the Sixth International Workshop on Security and Privacy Analytics, New York, NY, USA: ACM, Mar. 2020, pp. 53–60. doi: 10.1145/3375708.3380315.
[26] "CyberShield," CyberShield. Accessed: May 29, 2024. [Online]. Available: https://www.cybershield.org/
[27] A. Suciu, "Protecție prin educație: Securitate digitală. Cursuri de securitate cibernetică pentru elevi și adulți," LIVRESQ. Accessed: May 29, 2024. [Online]. Available: https://livresq.com/ro/news/protectie-prin-educatie-securitate-digitala-cursuri-de-securitate-cibernetica-pentru-elevi-si-adulti/

Ruxandra BADESCU holds a bachelor's degree in Economic Informatics and a master's degree in IT&C Security from the Faculty of Economic Cybernetics, Statistics, and Informatics. Currently serving as an analyst in the Security Operations Centre of a prominent financial institution, she specializes in safeguarding digital environments and enhancing user safety. Her research interests encompass the development of advanced threat detection methodologies, implementing proactive defense mechanisms, and optimizing incident response processes. She is also engaged in malware analysis, phishing prevention, vulnerability management, and securing endpoint devices, with an additional focus on digital forensics. Dedicated to promoting cybersecurity best practices, she actively contributes to a safer digital landscape.