ACIG
APPLIED CYBERSECURITY & INTERNET GOVERNANCE

NASK

Exploiting Human Trust in Cybersecurity: Which Trust Development Process is Predominant in Phishing Attacks?

Morice Daudi | Computing Science Studies, Mzumbe University, Tanzania | ORCID: 0000-0001-7907-427X

Abstract

Humans live in an interconnected world that is increasingly featured with virtual interactions in cyberspace. That world has raised cybersecurity concerns, particularly on exploiting human trust through various means, such as phishing. Phishing remains one of the most prevalent forms of cybercrime. It exploits human trust to manipulate individuals into divulging sensitive information. This study investigates the trust development mechanisms most exploited by cybercriminals in phishing attacks. It focuses on two primary trust development processes: relationship history and future expectations. The study uses qualitative content analysis of 42 phishing messages collected from diverse secondary sources. The findings reveal that future expectations – such as promises of rewards, urgent requests, or threats of penalties – dominate phishing tactics. By contrast, relationship history mechanisms exploit the existing or fabricated relationships to evoke trust and compliance. These findings provide critical insights into the psychological manipulations leveraged in phishing schemes and highlight the need to integrate behavioural and cognitive principles into cybersecurity education. Practical implications include tailored training programs for distinct user groups, such as seniors, employees, and

Received: 12.09.2024

Accepted: 21.12.2024

Published: 30.12.2024

Cite this article as:
M. Daudi, "Exploiting human trust in cybersecurity: which trust development process is predominant in phishing attacks?," ACIG, vol. 4, no. 2, 2024. DOI: 10.60097/ACIG/199452

Corresponding author:
Morice Daudi, Computing Science Studies, Mzumbe University, Tanzania;
E-mail: dmorice@mzumbe.ac.tz
0000-0001-7907-427X

students. The training should emphasise on recognising urgency cues, emotional manipulation, and verification strategies.

Keywords

phishing attacks, human trust, trust development processes, future expectations, cybersecurity

1. Introduction

Cybersecurity discourse has traditionally framed humans as a problem – susceptible to social engineering, prone to error, and easily manipulated. This framing, however, presents a limited view [1]. It is limited because the exploitation of humans as the weakest link in cybersecurity stems from the interplay of human psychology, social engineering tactics, and system usability. The theoretical challenge behind this problem focuses on how to mitigate the inherent vulnerabilities of human factors in the cyber landscape. Despite substantial investments in technological defences, human errors remain the leading cause of security breaches, contributing to as much as 90% of cybersecurity incidents [2, 3]. These errors arise from various sources, such as insufficient awareness, inadequate training, and susceptibility to psychological manipulation through social engineering tactics [2, 4]. Those human factors in cybersecurity are multifaceted and include intentional or unintentional actions that compromise security. For example, social engineering tactics exploit cognitive biases and psychological triggers, deceiving individuals into revealing confidential information or performing actions that undermine security protocols [2, 5]. These attacks leverage psychological principles like authority, reciprocity, and scarcity to manipulate victims [6, 7]. The susceptibility of individuals to such manipulation highlights the critical need for comprehensive cybersecurity education and the fostering of a security-aware culture within organisations [3, 4]. Therefore, combining technological solutions with insights into human behaviour is crucial for strengthening organisational resilience against emerging cyber threats [5, 8].

The literature provides varied perspectives on examining daily cybersecurity incidents involving phishing. Mitnick and Simon [9] discuss the manipulative tactics employed by cybercriminals and highlight the calculated exploitation of human emotions and cognitive biases. Hadnagy [10] explores how attackers exploit cognitive biases, trust, and social norms to manipulate individuals. Investigating the relationship between trust and cybersecurity

risks, Alhasan [4] reveals that higher trust increases risky cybersecurity behaviours across cultures. Additionally, Khan et al. [3] and Triplett [11] explore how human factors, including decision-making processes, organisational culture, and leadership contribute to insider threat. Despite these contributions, a gap remains in understanding the specific trust development process that cybercriminals rely on in phishing attacks. While the psychological and organisational dimensions of trust exploitation have been studied, there is limited focus on attackers' exact mechanisms and stages of trust development processes. This gap is critical, as understanding these processes could lead to more effective countermeasures. To this end, the present paper investigates the trust development processes most commonly employed by cybercriminals in phishing attacks. The study addresses the following research question: 'Which trust development process do cybercriminals most often exploit in phishing?' The study contributes to cybersecurity education by identifying the prevalent trust-building processes used in these exploits. This contribution empowers users to protect themselves better.

The remaining part of the paper is organised into seven sections. Section 2 discusses trust, phishing, and social engineering techniques for exploiting human trust. Section 3 outlines the methodology of this paper, followed by the presentation of results in Section 4. The findings presented in Section 5 are followed by their implications as discussed in Section 6. Section 7 provides practical recommendations. The paper ends with Section 8 by providing concluding remarks.

2. Literature Review

The present section comprises three subsections. It starts by discussing trust development processes (subsection 2.1), followed by the exploitation of human trust (subsection 2.2). Subsection 2.3 presents phishing techniques. The section ends by discussing social engineering techniques in subsection 2.4.

2.1. Trust Development Processes

Trust between parties evolves through specific processes. Before delving into these processes, it is essential to have a clear overview of the parties involved in trust transactions and the roles each party plays. For a trust exchange to be completed, two parties must be engaged: a trustor and a trustee. The trustor (e.g. a person) is an entity that develops a degree of reliance on another object

and accepts being vulnerable to the possible actions of that other object [12]. Similarly, the trustee (e.g. a person) is the party in whom the trust resides and can exploit the trustor's vulnerabilities [13]. To clarify further, the trustor is the party that puts its expectations in the other party, while the trustee is the party in which that expectation resides [14]. With this brief overview, the processes of trust development are discussed as follows.

Trust development processes can be understood through two primary mechanisms: relationship history and future expectations. Trust rooted in relationship history is built upon the experiences gained from past interactions between the trustor and the trustee [15]. Through relationship history, trust develops based on how parties have previously interacted and the experiences they have gained from one another. When parties have had no previous direct interactions, a reference from a third party is usually used to infer the development of trust [14]. Examples of bases of trust that employ relationship history include process-based, knowledge-based, and relational trust.

On the other hand, trust formed through future expectations is often driven by anticipated outcomes. Humans may trust the other party by relying on what they expect to gain after committing a trust transaction. This form of trust involves calculating the potential benefits and risks of engaging or not engaging in a particular trust transaction. Individuals assess whether entering a trusting relationship will yield favourable results or mitigate potential risks [14]. Examples of bases of trust that employ relationship history include calculus-based, deterrence-based, and competence-based trust [14, 15]. Both trust development processes (relationship history and future expectation) emphasise trust's dynamic nature.

2.2. Exploitations of Human Trust

Given the importance of trust in human interactions, cybercriminals exploit it as a key tactic in breaching cybersecurity. They leverage psychological principles, such as authority, reciprocity, and social proof to manipulate trust [1]. Those acts deceive individuals into compromising security systems. Trust exploitation is particularly effective because it taps into the inherent human tendency to trust familiar or authoritative sources [16].

One common method to exploit people's trust is phishing. Phishing relies heavily on manipulating human behaviour. In phishing attacks, cybercriminals craft messages that appear to originate

from trustworthy and legitimate sources to exploit the victim's inherent trust [4]. This deception is often amplified through urgent language or fabricated consequences. Through deception, individuals are compelled to respond quickly without fully verifying the communication's authenticity. The effectiveness of such attacks underscores the importance of raising awareness and educating individuals about the dangers of blindly trusting digital communications [7], particularly those that demand immediate action.

The cultural dimensions of trust also play a significant role in how individuals respond to phishing and other forms of deception. Research has shown that trust levels vary across cultures, with some cultures exhibiting higher baseline trust in digital communications [2]. Understanding these cultural differences is crucial for developing tailored cybersecurity strategies that address the specific trust-related vulnerabilities of different populations [6]. Generally, trust exploitation in cybersecurity highlights the relationship between psychology and technology. It also highlights the need for tactics combining technological protections with cultural and psychological knowledge.

2.3. Phishing Techniques in Cybersecurity

Phishing remains one of the most prevalent and effective techniques that cybercriminals employ to compromise cybersecurity. Phishing attacks typically involve sending fraudulent e-mails or messages that appear to come from legitimate sources [9]. The authors claim that those messages or e-mails lure individuals into providing sensitive information, such as passwords. The effectiveness of phishing lies in its ability to exploit basic human behaviour, such as trust and fear [10]. Trust and fear are sometimes triggered by falsified urgency and the authoritative nature of the messages. Despite widespread phishing awareness, the technique continues to evolve, becoming increasingly sophisticated and more challenging to detect [5].

Spear phishing, a more targeted form, has become a dangerous threat. Unlike typical phishing, which targets a large audience, spear phishing targets certain people or organisations [3]. It frequently relies on creating highly customised messages using data obtained from social media or other public sources. These messages are designed to appear credible and relevant to the recipient [11]. The precision and personalisation of spear phishing make it a formidable challenge for cybersecurity professionals, who must constantly adapt their defences to counter these evolving threats [4].

Phishing attacks have expanded beyond e-mail to include other communication platforms, such as SMS (smishing) and voice calls (vishing). These multi-vector attacks allow cybercriminals to simultaneously exploit different aspects of human behaviour and technological vulnerabilities [14]. For instance, smishing messages may appear to come from a trusted source, like a bank, and include a link that directs the victim to a fake website where their credentials are stolen [2]. Diversifying phishing techniques across multiple channels shows cybercriminals' adaptability and the need for comprehensive cybersecurity strategies covering many phishing vectors [7].

2.4. Social Engineering Techniques

Social engineering encompasses various tactics to manipulate individuals into divulging confidential information or performing actions compromising security. Social engineering is highly effective at breaching security systems because it exploits human psychological and cognitive biases [10]. Unlike traditional hacking, which targets technical vulnerabilities, social engineering exploits the human element, often seen as cybersecurity's weakest link [9]. By exploiting psychological principles like trust, authority, and reciprocity, social engineers can bypass technological defences and gain unauthorised access to systems or data [16].

Pretexting is a widely used social engineering technique where attackers create a fictitious scenario to deceive victims into revealing sensitive information. This often involves impersonating a trusted individual or authority figure, such as an IT support technician or a government official, to make the request appear legitimate [6]. The technique is particularly effective in organisational settings, where employees may feel obligated to comply with requests from perceived authorities [5]. The success of pretexting hinges on the attacker's ability to craft a convincing narrative that resonates with the victim's expectations and prior experiences [7].

Baiting is another common social engineering tactic that involves tempting victims with an enticing offer. The offer may comprise a gift to manipulate victims into actions compromising their security. This method exploits the human inclination for free or valuable items, often resulting in the spread of malware or the theft of sensitive information [2]. Baiting capitalises on individuals' curiosity and their tendency to take risks for potential rewards. Like other social engineering techniques, the success of baiting highlights the critical need for robust cybersecurity education that fosters skepticism and critical thinking in digital interactions [3].

3. Methodology

This study utilises a qualitative content analysis approach to investigate the trust development processes exploited by cybercriminals in phishing attacks. The research explores the two primary trust-building mechanisms – *relationship history* and *forthcoming expectations* – and their prevalence in phishing messages. The study categorises and analyses phishing messages to identify patterns and trends using these trust development processes. The data for this study was collected from various secondary sources, such as academic publications, cybersecurity reports, and online repositories of phishing messages. Specifically, phishing messages were extracted through search engines. These sources were chosen due to their comprehensive coverage of phishing tactics and their relevance to the research topic. Most spam messages from those sources are generic, which is considered a reference for many spam messages. A total of 42 phishing messages were selected for analysis to comprehensively represent various phishing tactics. These messages were intentionally chosen to capture the trust development processes related to relationship history and future expectations.

The phishing messages included in this study were purposely selected based on the relevance, variety, and recency criteria. Firstly, messages were included if they explicitly or implicitly involved trust development tactics to deceive the recipient. Secondly, a diverse set of messages was selected to cover different types of phishing attempts, such as those related to financial incentives, urgent requests, or personal relationships. Finally, preference was given to messages representative of contemporary phishing tactics to ensure that the findings are relevant to current cybersecurity challenges. The selected phishing messages were analysed using a thematic content analysis method. Each message was reviewed to identify the trust development process utilised – either relationship history or forthcoming expectations. Each phishing message was coded according to the identified trust development process. The frequency of each type was recorded and analysed to determine which process is more commonly exploited by cybercriminals.

To ensure the reliability of the analysis, two assistant researchers independently coded phishing messages. Any discrepancies in coding were discussed and resolved through consensus to mitigate potential biases in message classification. Validity was addressed by triangulating the findings with existing literature (such as that in Daud [14]) on phishing tactics and trust development processes in cybersecurity. The results were compared with previous studies

to ensure that the identified patterns align with established knowledge in the field. Furthermore, as the study utilised publicly available data from secondary sources, no personal information was collected or analysed. All sources of phishing messages were adequately cited, and care was taken to ensure that the analysis did not involve any unethical data manipulation.

4. Results

Table 1 presents the results of spam messages used mainly by cybercriminals. These messages were extracted from literature sources [14, 17–21]. Of the 42 spam messages, 33 were based on the future expectation trust development process. The remaining nine messages were based on the relationship history trust development process.

This study categorises spam messages presented in Table 1 into various groups: account verification, billing statements, credit card offers, customer service inquiries, family matters, job offers, and package delivery notifications (Figure 1). The most commonly identified categories were prizes or gift cards and account verification requests, each occurring for six times. This high frequency indicates that cybercriminals often focus on areas where individuals are likely to respond quickly, sometimes without exercising adequate caution. Family matters, package delivery, and internal revenue services ranked third, fourth, and fifth, respectively.

An in-depth analysis of spam messages indicates that cybercriminals frequently employ specific trust-building techniques to deceive their victims. Notably, 78.6% of the spam messages analysed were designed using the *future expectation* trust-building process. This process often promises future rewards or urgent actions, such as account verification or prize claims. It leverages urgency and anticipation to compel recipients to respond quickly without critically evaluating the legitimacy of the request. For instance, the following messages are classic examples of this approach:

> 'Congratulations! You've won a $500 Amazon gift card. Claim it here [Link]'
>
> and
>
> 'Your IRS tax refund is pending acceptance. Must accept within 24 hours: [Link]'

Table 1. Sample spam messages used by cybercriminals

No.	Spam messages	Trust development process
1.	Congratulations! You've won a $500 Amazon gift card. Claim it here [Link].	Future expectation
2.	ACTION REQUIRED. Please verify your Bank of America account information to avoid a hold on your account. Click here to confirm: [Link].	Future expectation
3.	You've been overcharged for your 2021 taxes. Get your IRS tax refund here: [Link].	Future expectation
4.	Get delivery updates on your USPS order [Number] here: [Link].	Future expectation
5.	Thank you for paying last month's bill. We're rewarding our very best customers with a gift for their loyalty. Click here! [Link].	Future expectation
6.	Congratulations! Your credit score entitles you to a no-interest Visa credit card. Click here to claim: [Link].	Future expectation
7.	We've received your resume and would love to set up an online interview. Click here [Link] or call us at [Phone Number] at your earliest convenience.	Relationship history
8.	There's an issue with your payment information from your recent order [Order Number]. Take action now: [Link].	Future expectation
9.	We have detected suspicious activity on your Wells Fargo account. Log in at [Link] to update your account preferences and protect your information.	Future expectation
10.	Hi Grandpa, it's me – I've been in a car accident, and my parents aren't around. Can you please send me money so I can get home? You can wire funds to me here: [Link].	Relationship history
11.	'Your 2FA settings are not up to date. To avoid account suspension, please click the following link to update your settings: [Link]'.	Future expectation
12.	'Hey, it's [Boss Name]. I'm in a meeting now and need your help with something urgent. Can you transfer $5,000 to this account ASAP? I'll explain everything later. Please keep this confidential'.	Relationship history
13.	'We're happy to inform you that you're entitled to a refund for overpayment on your AMEX account. Click on this link [Link] below to claim your refund'.	Future expectation
14.	Congratulations! You have all been selected to receive a free gift card worth $1000. Click on this link [Link] to claim your reward now. Limited time offer, so act fast! Don't miss out on this amazing opportunity.	Future expectation
15.	'Congratulations! You've won a $500 gift card to Target. Click here to claim your reward'.	Future expectation
16.	'Hello [Name], your shipment from UPS will arrive today. Click here to track your package'.	Future expectation
17.	'Your Wells Fargo account has been locked for suspicious activity. Please log in here and verify your account'.	Future expectation
18.	'Hey, this is [Name]. I'm in a meeting, but I need you to order 5 Amazon gift cards ASAP. I'll reimburse you once you send them to this e-mail address'.	Future expectation
19.	'[Name], your Verizon billing statement is ready. Please review your charges and send full payment by [date] to avoid late fees'.	Future expectation
20.	Congratulations! You've won a $1000 Walmart gift card. Go to [Link] claim now.	Future expectation

(*continues*)

Table 1. Continued.

No.	Spam messages	Trust development process
21.	Your IRS tax refund is pending acceptance. Must accept within 24 hours: [Link].	Future expectation
22.	Wells Fargo Bank: Your account is temporarily locked. Please log in at [Link] to secure your account.	Future expectation
23.	Hello, your FEDEX package with tracking code DZ-8342-FY34 is waiting for you to set delivery preference: [Link].	Future expectation
24.	Apple Notification. Your Apple iCloud ID expires today. Log in to prevent deletion [Link].	Future expectation
25.	URGENT: Your grandson was arrested last night in Mexico. Need bail money immediately Western Union Wire $9,500 [Link].	Relationship history
26.	Federal Credit Union ALERT: Your Credit Card has been temporarily LOCKED. Please call Card Services line [Tel. no].	Future expectation
27.	Thank you for your recent Amazon purchase. You've been charged $108.34. If there has been a mistake, please call [Tel. no].	Future expectation
28.	Dear [Bank Name] customer, we've detected unusual activity on your account. Please click the link to verify your transactions: [malicious link].	Future expectation
29.	Hello, this is [Courier Service]. We've attempted to deliver your package today but failed. Schedule your redelivery here: [malicious link].	Future expectation
30.	We detected a login attempt from an unfamiliar location. If this wasn't you, please secure your account here: [malicious link].	Future expectation
31.	You're the lucky winner of our grand prize! Register here to receive your reward: [malicious link].	Future expectation
32.	A family member of yours has been in an accident. Call this premium rate number for details: [malicious phone number].	Relationship history
33.	I'm your landlord. My current number is unreachable. Send the rent through this number [Tel. no].	Relationship history
34.	This is agent (name withheld) from telecom company (name withheld). Your mobile money account has insufficient funds. Deposit TSh 500,000 today, then call us back. Otherwise, we are going to close your account.	Future expectation
35.	You are speaking with someone from the telecom company (name withheld); your monthly bonus is TSh 400,000 now. Use a different mobile phone so that we can help you obtain the money.	Future expectation
36.	This is the Revenue Authority office. Why don't you use an electronic fiscal device (EFD) when conducting business? A Tsh 3 million fine is being sent to you immediately.	Future expectation
37.	After unexpectedly collapsing at school, your son was brought to the hospital. Send money right away for medical care.	Relationship history
38.	Please get in touch with us as soon as you can; your child is extremely ill. Teacher.	Relationship history
39.	You won in the draw for the best customers who use our services. Please contact the following number to learn how to collect your prize.	Future expectation

(*continues*)

Table 1. Continued.

No.	Spam messages	Trust development process
40.	You have received Tshs 50,000 from [Tel. no] – (name of sender). New balance 67,850.00 Tshs. Trans ID: [Trans. No]. [Date and time].	Future expectation
41.	I'm at the funeral; please send twenty thousand shillings at the following phone number. I will pay back your money later.	Relationship history
42.	The person received a phone call from someone pretending to be a human resource officer at an airport. The caller claimed, the recipient's job application had been received and requested Tsh 300,000 in exchange for persuading his superiors to select the recipient for the position.	Future expectation

Source: Extracted from [14, 17–21].

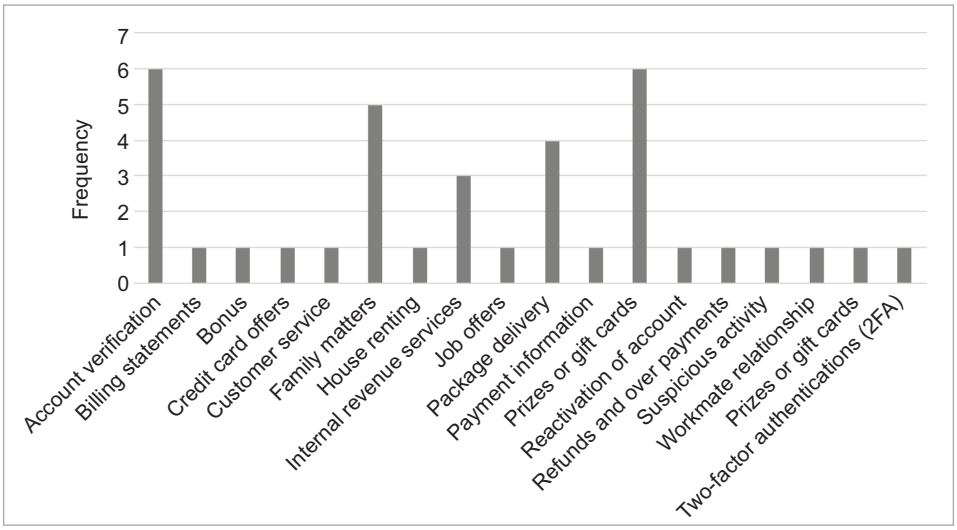

Figure 1. Categories of spam messages

These messages leverage the recipient's hope for a positive outcome or fear of missing out.

On the other hand, the relationship history trust development process relies on exploiting the existing relationships or creating fictitious ones. These messages are crafted to appear as if they come from someone the recipient knows or trusts. They may appear to come from a family member, colleague, or service provider. An example of this would be the following message:

> 'Hi Grandpa, it's me – I've been in a car accident and my parents aren't around. Can you please send me money so I can get home? You can wire funds to me here: [Link]'

The use of the relationship history trust development process emphasises the emotional connection and people's trust in their close relationships. It makes the recipient more likely to comply with the request without scepticism.

5. Discussion of the Findings

The analysis reveals key insights into cybercriminals' methods to exploit human trust in phishing attacks. One key insight concerns the prominence of the future expectation trust development process. Through future expectation, attackers target psychological triggers that urge immediate action. This approach is effective to attackers because it preys on common human behaviours. Examples of such behaviours include human tendency to seek financial gain or resolve issues quickly. Such behaviours are closely linked to the power of anticipation and urgency which cybercriminals understand and use it. Cybercriminals usually create scenarios where the victim believes they must act quickly to avoid negative consequences or secure a reward. While doing so, attackers limit the time available for critical assessment by the victim. This technique is particularly dangerous in today's fast-paced digital environment, where individuals often juggle multiple tasks and may overlook the need to scrutinise each message.

Various principles in literature underpin the dominance of future expectations as the trust-building process in phishing attacks. One is Cialdini's [16] principle of urgency and scarcity, where attackers create a sense of urgency, such as 'Your mobile money account will be closed immediately'**.** This tactic exploits the fear of missing out, pressuring victims to act quickly without assessing the message's legitimacy. This fear reinforces the deterrence-based trust developed through the future expectation process. Another principle is based on Sweller's [22] cognitive load theory, which posits that individuals under time pressure tend to rely on cognitive shortcuts (heuristics), rather than engaging in critical thinking. Attackers exploit this by leveraging the future expectation trust process. They do so by prompting victims to respond to phishing messages that promise rewards or threaten penalties. That action effectively bypasses the cognitive effort required to assess the authenticity of phishing messages.

Besides this, the findings from this study align with and extend to the existing research on phishing and social engineering in cybersecurity. For instance, an SMS phishing experiment revealed that combining urgency with either the promise of a reward or the

threat of a penalty successfully deceived 50% of participants [23]. This urgency is a key element in the future expectation trust development process discussed by Daudi [14]. Overall, these findings align with Vishwanath et al.'s [24] conclusion that superficial e-mail processing increases phishing success.

On the other hand, the use of relationship history as a trust-building process demonstrates the effectiveness of social engineering in phishing attacks. Cybercriminals bypass initial scepticism by impersonating someone the victim knows or trusts. This tactic exploits the victim's existing relationships, making it a powerful tool for attackers. It is particularly effective in urgent scenarios, such as requests for emergency funds to care for a sick child at school. The relationship history trust development process identified in this study further illustrates how cybercriminals exploit emotional connections to bypass rational scrutiny. To bypass rational scrutiny, cybercriminals often build rapport to gain trust and extract sensitive information. The success of such cybercriminals' attacks is backed up by humans' tendency to rely on familiar cues when assessing the authenticity of messages [25].

In addition to the trust-building process based on future expectations, gifts and financial incentives are often employed in the trust development process rooted in relationship history. For instance, spear-phishing e-mails that exploit a fabricated relationship history tend to achieve higher success rates than generic phishing e-mails [26]. Some of these e-mails create a sense of urgency by demanding immediate action from victims. Under such time pressure, individuals are more likely to overlook security protocols, skip essential steps, and make decisions that compromise cybersecurity [27]. Similarly, Razaq et al. [28] observed that fraudsters frequently pose as bank officials or government representatives, leveraging urgency to prompt swift compliance and establish trust with their victims.

6. Implications

The findings of this study have significant implications for cybersecurity practices, policy-making, and behavioural research. The dominance of future expectation mechanisms in phishing attacks highlights the need to address cognitive biases like urgency, anticipation, and risk perception in cybersecurity training. It must be recognised that individuals often fall victim to phishing because they are manipulated into prioritising immediate outcomes over critical evaluation. This insight necessitates integrating behavioural

and psychological theories, such as cognitive load theory [22] and temporal discounting [29], into awareness programs. This integration helps users better process suspicious messages. Moreover, the marginality of relationship history mechanisms suggests that attackers also capitalise on emotional connections. For this reason, users should exercise caution and verify communications, particularly those claiming personal relationships.

From a policy perspective, cybersecurity frameworks should incorporate behavioural training alongside technical solutions. Emphasis should be placed on vigilance and critical thinking. Additionally, the results indicate the need for adaptive cybersecurity measures that account for cultural and demographic differences in trust dynamics. Future research should explore these variations more deeply to develop region-specific strategies. Overall, this study emphasises that mitigating phishing effectively requires a holistic approach. This approach must integrate technological defences, psychological insights, and user education to create strong protection against evolving cyber threats.

7. Practical Recommendations

Organisations and individuals must implement targeted strategies to counter phishing attacks exploiting trust mechanisms. Firstly, cybersecurity training programs should focus on psychological manipulation tactics, such as urgency and anticipated rewards in phishing messages. These programs should teach individuals to recognise common phishing patterns, such as requests for immediate actions, financial rewards, or penalties. Secondly, organisations should simulate real-world phishing scenarios through controlled phishing campaigns. These exercises provide users with hands-on practice in identifying suspicious messages and offer immediate feedback. This approach effectively enhances their resilience against such attacks. Thirdly, automated e-mail and message filters should be strengthened by using appropriate tools. These tools can detect phishing-related language patterns, such as urgency cues or impersonation attempts. Verification practices should be emphasised for individuals. They should involve crosschecking of messages through alternative channels like direct calls or official websites. Lastly, organisations must develop user-specific awareness programs tailored to various demographics, such as employees, older adults, and students. This is because each group faces distinct vulnerabilities to trust-based phishing tactics. Combining these strategies will improve detection rates and minimise successful phishing exploits.

8. Conclusion

The exploitation of human trust to deceive and manipulate computer system users has become a significant concern in cybersecurity. Through social engineering and phishing, many users have fallen victim in various contexts. This study reveals that phishing attacks primarily exploit human psychological vulnerabilities through two trust development processes: future expectations and relationship history. The findings indicate that future expectations – such as promises of rewards, warnings of penalties, or urgent requests – are the most frequently used mechanisms by cybercriminals. These tactics rely on creating a sense of urgency and anticipation. Through this sense, victims are compelled to act impulsively without critically assessing the message's legitimacy. On the other hand, relationship history exploits familiarity and emotional connections. Attackers use this method to build trust by impersonating known individuals or organisations. The study highlights the need to incorporate behavioural insights into cybersecurity training and awareness programs. These programs should address cognitive biases, such as urgency and emotional triggers, to help individuals better identify and resist phishing attempts. Furthermore, mitigating phishing threats requires a multifaceted approach combining technological defences, user education, and understanding the human psychology of trust. By addressing these aspects holistically, individuals and organisations can develop more effective strategies to combat evolving phishing tactics and enhance overall cybersecurity resilience.

While this research provides valuable insights into the trust mechanisms exploited in phishing, its reliance on secondary data introduces certain limitations. Future studies should incorporate primary data collection methods, such as surveys, interviews, or experiments, to better understand user behaviours and responses to phishing attacks. Such approaches can provide richer insights into how cybercriminals exploit trust in real-world scenarios.

References

[1] R. Anderson, *Security engineering: A guide to building dependable distributed systems*. Hoboken, NJ: Wiley, 2020.

[2] A. Pollini et al., "Leveraging human factors in cybersecurity: An integrated methodological approach," *Cognition Technology & Work*, vol. 24, no. 2, pp. 371–390, 2022. doi: 10.1007/s10111-021-00683-y.

[3] N. Khan, R.J. Houghton, S. Sharples, "Understanding factors that influence unintentional insider threat: A framework to counteract unintentional risks,"

Cognition Technology & Work, vol. 24, no. 3, pp. 393–421, 2022. doi: 10.1007/s10111-021-00690-z.

[4] I. Alhasan, *Human factors in cybersecurity: A cross-cultural study on trust.* West Lafayette, IN: Purdue University, 2023.

[5] S. Chaudhary, V. Gkioulos, S. Katsikas, "Developing metrics to assess the effectiveness of cybersecurity awareness program," *Journal of Cybersecurity*, vol. 8, no. 1, pp. 1–19, 2022. doi: 10.1093/cybsec/tyac006.

[6] E.O. Yeboah-boateng, P.M. Amanor, "Phishing, SMiShing & vishing: An assessment of threats against mobile devices," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 5, no. 4, pp. 297–307, 2014.

[7] H. Kilavo, L.J. Mselle, R.I. Rais, S.I. Mrutu, "Reverse social engineering to counter social engineering in mobile money theft: A Tanzanian context," *Journal of Applied Security Research*, vol. 18, no. 3, pp. 546–558, Jul. 2023. doi: 10.1080/19361610.2022.2031702.

[8] M. Grobler, R. Gaire, S. Nepal, "User, usage and usability: Redefining human centric cyber security," *Frontiers in Big Data*, vol. 4, pp. 1–18, 2021. doi: 10.3389/fdata.2021.583723.

[9] K. Mitnick, W.L. Simon, *The art of deception: Controlling the human element of security*. Hoboken, NJ: Wiley, 2002.

[10] C. Hadnagy, *Social engineering: The art of human hacking*. Hoboken, NJ: Wiley, 2010.

[11] W.J. Triplett, "Addressing human factors in cybersecurity leadership," *Journal of Cybersecurity and Privacy*, vol. 2, no. 3, pp. 573–586, 2022. doi: 10.3390/jcp2030029.

[12] M. Daudi, *Trust in sharing resources in logistics collaboration*. Düren: Shaker Verlag GmbH, 2019.

[13] M. Laeequddin, B.S. Sahay, V. Sahay, K.A. Waheed, "Trust building in supply chain partners relationship: an integrated conceptual model," *Journal of Management Development*, vol. 31, no. 6, pp. 550–564, 2012. doi: 10.1108/02621711211230858.

[14] M. Daudi, "Trust framework on exploitation of humans as the weakest link in cybersecurity," *Applied Cybersecurity & Internet Governance*, vol. 2, no. 1, pp. 1–26, 2023. doi: 10.60097/acig/162867.

[15] N.P. Nguyen, N.T. Liem, "Inter-firm trust production: Theoretical perspectives," *International Journal of Business, Management*, vol. 8, no. 7, 2013. doi: 0.5539/ijbm.v8n7p46.

[16] R.B. Cialdini, *Influence: The psychology of persuasion*. NewYork, NY: Harper Business, 2007. doi: 10.1021/jf970693b.

[17] J. Chantel. (2022). *10 spam text message examples (& how to identify them)* [Online]. Available: https://blog.textedly.com/spam-text-message-examples [Accessed: Mar. 06, 2024].

[18] Proofpoint. (2023). *State of the phish report* [Online]. Available: https://www.proofpoint.com/us/threat-reference/smishing [Accessed: Apr. 17, 2024].

[19] SlickText. (2023). *17 Spam text statistics & spam text examples for 2024* [Online]. Available: https://www.slicktext.com/blog/2022/10/17-spam-text-statisitics-for-2022/ [Accessed: Mar. 06, 2024].

[20] R. Smith. (2023). "Stop scammers! 14 Examples of spam text messages," *Texting Base* [Online]. Available: https://blog.textingbase.com/how-to-identify-spam-text-messages [Accessed: Apr. 17, 2024].

[21] I. H. Bakar (2016). *Social engineering tactics used in mobile money theft in Tanzania*, The University of Dodoma [Online]. Available: http://repository.udom.ac.tz/handle/20.500.12661/1168 [Accessed: Jan. 09, 2024].

[22] J. Sweller, "Cognitive load theory," *Cognition Science*, vol. 12, no. 2, pp. 257–285, 1988. doi: 10.1016/0364-0213(88)90023-9.

[23] H. Siadati, T. Nguyen, P. Gupta, M. Jakobsson, N. Memon, "Mind your SMSes: Mitigating social engineering in second factor authentication," *Computers & Security*, vol. 65, pp. 14–28, 2017. doi: 10.1016/j.cose.2016.09.009.

[24] A. Vishwanath, T. Herath, R. Chen, J. Wang, H.R. Rao, "Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model," *Decision Support System*, vol. 51, no. 3, pp. 576–586, 2011. doi: 10.1016/j.dss.2011.03.002.

[25] R. Dhamija, J.D. Tygar, M. Hearst, "Why phising works," in *Proceedings of CHI-2006: Conference on human factors in computing systems, April 2006.* New York, NY: ACM, 2006, pp. 581–590. doi: 10.1145/1124772.11248.

[26] K. Dubovecka, "Vulnerability of students of Masaryk University to two different types of phishing," *Applied Cybersecurity & Internet Governance*, vol. 4, no. 2, 2024. doi: 10.60097/ACIG/190268.

[27] N.H. Chowdhury, M.T.P. Adam, T. Teubner, "Time pressure in human cybersecurity behavior: Theoretical framework and countermeasures," *Computers & Security*, vol. 97, p. 101931, Oct. 2020. doi: 10.1016/j.cose.2020.101931.

[28] L. Razaq, T. Ahmad, S. Ibtasam, U. Ramzan, S. Mare, "We even borrowed money from our neighbor," *Proceedings of ACM Human-Computer Interaction*, vol. 5, no. CSCW1, pp. 1–30, Apr. 2021. doi: 10.1145/3449115.

[29] R. Herrnstein, "Temporal discounting," *Journal of Experimental Analysis of Behavior*, vol. 4, no. 3, pp. 267–272, 1961. doi: 10.1901/jeab.1961.4-267.