Understanding Susceptibility to Phishing in Mobile Instant Messaging Applications

University of Strathclyde Glasgow

Rufai Ahmad

This dissertation is submitted for the degree of Doctor of Philosophy

March 2026

Department of Computer and Information Sciences

This thesis is dedicated to the memory of my beloved father, Ahmad Isah Sokoto, who passed away on December 3, 2023. His unwavering emotional and financial support throughout my academic journey was invaluable and deeply appreciated. I also dedicate this work to my mother, Aisha Ahmad Isah Sokoto, and my wife, Rabia Muhammad Jiddah, whose love, encouragement, and sacrifices have been a constant source of strength and support.

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.' 'The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed: Rufai Ahmad

Date:21/09/2025

Acknowledgements

My PhD journey has not been easy, with many obstacles and setbacks. However, with the support of those around me, I was able to keep pushing. I extend my profound gratitude to Allah (SWT) for granting me the strength and wisdom to complete this research. I also want to express my sincere appreciation to my first supervisor, Dr Sotirios Terzis, for supporting me throughout my PhD journey. I would also like to thank my second supervisor, Dr Karen Renaud, for her support during this journey. Lastly, I would like to thank my family and some of my friends for helping me reach this goal.

Abstract

Mobile Instant Messaging (MIM) applications, including WhatsApp, Viber, and Telegram, are among the most prevalent communication methods, utilising users' mobile devices and existing internet data plans to facilitate real-time text, voice, and video communication over the internet. Recently, concerns have been raised regarding cybercriminals' adoption of these applications to propagate phishing campaigns. Yet, little attention has been given to understanding phishing in such platforms, specifically the behaviour of users and the attackers' tactics and techniques that can increase susceptibility. The studies described in this thesis examined susceptibility to phishing in MIM applications. The first study employed an exploratory approach, using an online questionnaire to assess whether users' self-reported behaviours during MIM expose them to phishing risks. The second study employed qualitative content analysis to examine the tactics and techniques used by cybercriminals in MIM phishing messages. The third study employed a factorial vignette survey to investigate the effect of specific features of MIM applications on susceptibility to phishing.

The findings indicate that users frequently click on and share links, and seek to mitigate their susceptibility to phishing by engaging with links from individuals within their social cycles, while demonstrating a greater tendency to share links received in private rather than in public groups. The findings indicate that, unlike email phishing but similar to vishing, social proof is the predominant persuasion principle in MIM phishing, while URL construction techniques show similarities across MIM and other phishing media. Furthermore, empirical results have been found to underscore the user-sender relationship, link preview, and habitual MIM app usage as significant factors in susceptibility to phishing in MIM applications. This thesis highlights both the social and technical dimensions of MIM phishing, underscoring the need for phishing awareness programs and security features that consider link previews, habitual behaviours and social context within MIM apps.

Publications

Some of the materials in this thesis have been presented at various research venues, and one paper was published in the Journal of Information and Computer Security.

- Ahmad, R., Terzis, S. (2022). Understanding Phishing in Mobile Instant Messaging: A Study into User Behaviour Toward Shared Links. In: Clarke, N., Furnell, S. (eds) Human Aspects of Information Security and Assurance. HAISA 2022. IFIP Advances in Information and Communication Technology, vol 658. Springer, Cham. https://doi.org/10.1007/978-3-031-12172-2_15
- Ahmad, R., Terzis, S., Renaud, K. (2023). Investigating Mobile Instant Messaging Phishing: A Study into User Awareness and Preventive Measures. In: Moallem, A. (eds) HCI for Cybersecurity, Privacy and Trust. HCII 2023. Lecture Notes in Computer Science, vol 14045. Springer, Cham. https://doi.org/10.1007/978-3-031-35822-7_26
- Ahmad, R., Terzis, S., Renaud, K. (2023). Content Analysis of Persuasion Principles in Mobile Instant Message Phishing. In: Furnell, S., Clarke, N. (eds) Human Aspects of Information Security and Assurance. HAISA 2023. IFIP Advances in Information and Communication Technology, vol 674. Springer, Cham. https://doi.org/10.1007/978-3-031-38530-8_26
- Ahmad, R., Terzis, S. and Renaud, K. (2024), "Getting users to click: a content analysis of phishers' tactics and techniques in mobile instant messaging phishing", Information and Computer Security, Vol. 32 No. 4, pp. 420-435. https://doi.org/10.1108/ICS-11-2023-0206

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter presents a comprehensive overview of the research, encompassing the study's motivation, research problems and questions, methodological approach, research outcomes, contributions, and the dissertation structure.

1.1 Motivation

Phishing is a cybercrime in which malicious individuals attempt to obtain sensitive information from internet users by deceiving them into accessing fake websites, disclosing personal data, or downloading harmful software (NCSC, 2018). According to a recent report from the Anti-Phishing Working Group (APWG), 1,003,924 phishing attacks were observed in the first quarter of 2025, marking the largest number since the fourth quarter of 2023 (APWG, 2025). Although email continues to be the primary vector through which cybercriminals conduct phishing attacks (Sjouwerman, 2025), the 2024 Phishing Threat Trends Report from the email security provider Egress highlighted that cybercriminals currently do not restrict themselves to one mode of communication but instead use other media such as Microsoft Teams, Slack and SMS (Egress, 2024a).

In recent years, Mobile Instant Messaging (MIM) applications (apps) such as WhatsApp, Viber, Telegram, and Facebook Messenger have become popular means of communication (Dixon, 2024). They allow users to communicate with each other in real-time. According to recent data from Statista, the number of users utilising MIM apps for communication in 2023 was 3.9B (Laura Ceci, 2024). It has been shown that the frequency of usage of such platforms supersedes that

of SMS (Seufert *et al.*, 2015). Their popularity has resulted in their adoption in electronic commerce, a phenomenon known as instant messaging (IM) social commerce (Cao *et al.*, 2021).

The widespread use and popularity of MIM apps have drawn the attention of cybercriminals, who use them to propagate phishing campaigns (Kaspersky, 2021). A 2021 report from Kaspersky reveals that attackers use MIM apps to spread phishing attempts, with the majority of phishing links being shared on WhatsApp (89.6%), followed by Telegram (5.6%) and Viber (4.7%) (Kaspersky, 2021). These attacks involve the impersonation of familiar contacts or reputable organisations through the use of compromised or falsified identities, thereby generating a misleading perception of legitimacy (Dror, 2024).

The use of MIM apps for propagating phishing has been identified as a significant concern by cybersecurity experts (Cuddeford, 2018). Of particular concern is that, compared to the email, MIM apps have far less protection in place. These apps currently lack automated safeguards to filter phishing messages and protect users against phishing (Stivala and Pellegrino, 2020). Another concern among security experts is the exploitation of victims to propagate phishing campaigns by encouraging them to share malicious content with their contacts (Cuddeford, 2018). This approach further extends the reach of the attackers, contributing to the campaign's virality (Cuddeford, 2018).

The points made above demonstrate that MIM phishing has unique characteristics that make it highly attractive to cybercriminals. Firstly, the lack of automated safeguards against phishing, which is attributed to the end-to-end encryption feature of MIM apps (Dror, 2024), enables cybercriminals to carry out phishing campaigns without immediate detection. Secondly, the exploitation of victims to propagate phishing campaigns by encouraging them to share malicious content may allow cybercriminals to strategically exploit the inherent credibility of interpersonal and professional networks to enhance the effectiveness of the attack.

1.2 Research Problem

MIM phishing represents an emerging threat that deserves attention. However, little attention has been given to understanding phishing on such platforms, specifically the behaviour of users and the tactics and techniques employed by attackers that can increase susceptibility. Unlike conventional platforms, MIM applications possess distinctive contextual and technical characteristics that can fundamentally reshape the phishing strategies and techniques used by attackers and the vulnerabilities experienced by users. The unique contextual characteristics of MIM apps justify further investigation into how phishing tactics and user behaviours are shaped and influenced within this environment. Such research is crucial for informing the development of effective technological solutions and user awareness strategies to mitigate this growing risk. This thesis aims to contribute to the literature by examining the phishing problem in MIM apps from both the phisher's and user's perspectives using self-reported data from study participants about their behaviours and intentions, and secondary data to explore and analyse the phishing strategies used during mobile instant messaging phishing attacks. This approach will provide a comprehensive understanding of how the communication medium influences phishing tactics and user behaviour. To do this, the thesis addressed the following research questions:

RQ1: Do users of MIM apps exhibit behaviours that could expose them to the risk of falling victim to a phishing attack?

RQ2: What are the tactics and techniques used in the content of phishing messages targeting the users of MIM apps?

RQ3: Do specific characteristics of MIM apps influence user susceptibility to phishing within these platforms?

The above questions have been further refined into sub-questions based on the literature in each study. Details regarding the rationale for the refinement can be

found in the corresponding chapters. However, Table 1-1 shows both the questions and sub-questions.

Table 1-1: Research Questions with Sub Questions

Main Questions	**Sub-Questions**	**Research Methods**
RQ1: Do users of MIM apps exhibit behaviours that could expose them to the risk of falling victim to a phishing attack?	RQ1.1 Are users of MIM apps aware of the phishing attacks using MIM messages to target them? RQ1.2 What strategies do MIM app users employ to safeguard themselves from phishing attempts? RQ1.3 How often do people click on and forward links shared in various MIM apps? RQ1.4 Does the relationship between the recipient and sender of a message, or the type of group, impact users' behaviours towards links shared in MIM apps? RQ1.5 What other factors influence MIM app users' behaviours towards links	Exploratory study using online questionnaire

	shared during instant messaging?	
RQ2: What are the tactics and techniques used in the content of phishing messages targeting the users of MIM apps?	RQ2.1 What persuasion principles are employed by phishers in MIM phishing, and what was the frequency of their usage? RQ2.2 Do phishers combine multiple persuasion principles in MIM phishing messages? RQ2.3 How are the different principles of persuasion implemented? RQ2.4 What URL construction methods do phishers employ to create deceptive URLs during MIM phishing? RQ 2.5 Does the prevalence of URL construction methods vary according to the type of organisation being impersonated?	Qualitative content analysis (QCA)
RQ3: Do specific characteristics of MIM apps influence user susceptibility to phishing within these platforms?	RQ.3.1 Does susceptibility to phishing in WhatsApp differ by the user's relationship with the sender of a message?	factorial survey (FS) methodology

	RQ 3.2 Are users of WhatsApp more susceptible to phishing with a link preview than those without a link preview? RQ 3.3 Does habituation to WhatsApp influence susceptibility to phishing within it?	

1.3 Methodological Approach

The research in this thesis was conducted in several stages, utilising various methodological approaches carefully selected to address the research questions. The methods used include a questionnaire, qualitative content analysis, and a questionnaire with vignettes. The following outlines the research methodology.

- Background literature review: Initially, a comprehensive literature review was conducted to gain an understanding of phishing and its various forms. The review examined the evolution of phishing, with a particular focus on its transition from desktop computers to mobile devices, specifically smartphones. The review examined phishing based on the technology and transmission medium used. Also, it highlighted the unique features of these media that facilitated phishing attacks. Through the literature, the researcher gained insights into the specific contextual attributes of MIM apps that can potentially enable phishing activities. This understanding guides the design of the studies conducted in this thesis.
- An exploratory study of user behaviours during mobile instant messaging (RQ1): To answer RQ1, a study was conducted to understand the behaviours of MIM apps users during instant messaging and examine whether such behaviours expose them to phishing risks. In particular, the study focused on their self-reported behaviour towards links shared

during communication in MIM apps, which has not been explored in the literature. The study utilised an online questionnaire using both Likert scale and multiple-choice questions. A total of 129 participants accessed the survey. However, 18 participants were discarded for failing to meet the screening criteria.

- Qualitative content analysis to understand the tactics and techniques used in the content of MIM phishing attacks (RQ2): To answer RQ2, a qualitative content analysis of 67 MIM phishing messages was carried out. The study aimed to understand the tactics and techniques used by cybercriminals during MIM phishing.
- A factorial survey to understand the effect of specific characteristics of MIM apps on user susceptibility to phishing in such apps(RQ3): To answer RQ3, an online factorial survey, also known as a vignette survey, was conducted with MIM app users to understand the influence of specific characteristics of MIM apps on their decision to trust and respond to phishing messages. The survey utilised both Likert scale, multiple-choice, and open-ended questions. A total of 814 participants responded to the survey. However, after applying a stringent data cleaning strategy, only 127 participants qualified for analysis.

1.4 Research Outcome

This thesis provides valuable insights into phishing within mobile instant messaging (MIM) environments, highlighting how the unique features of these applications both increase user susceptibility and inform attacker strategies.

Firstly, the findings reveal a gap between users' perceived ability to detect phishing and their self-reported behaviour in MIM environments. Despite high self-assessed competence, the participants frequently engaged in actions that increased their susceptibility to phishing. The findings from self-reported data indicate that MIM apps users frequently engage in behaviours that increase their

susceptibility to phishing. For instance, participants frequently click on links shared by others and tend to forward such links further, exposing themselves to the risk of phishing. These findings suggest that some features of MIM apps, such as users' ability to share and forward links, while facilitating efficient communication, also create conditions that attackers can exploit.

The findings also show how current vulnerabilities in the link preview generation mechanism have allowed phishers to create deceptive previews, which empirical evidence from this thesis has confirmed increases user susceptibility to phishing. The findings show that link previews, whose aim is to inform users about websites, are exploited by phishers to display organisational logos and create a false sense of authority, thereby deceiving users into perceiving malicious content as legitimate. These findings suggest that certain features of MIM apps, when combined with typical usage patterns, can increase user susceptibility to phishing.

Secondly, the first and third studies of this thesis revealed that MIM app users appear more inclined to trust and act upon messages and links received from individuals with whom they share social ties, thereby directing their attention away from potentially malicious content. Whilst the self-reported behaviour of users shows evidence of reduced risk awareness and reliance on social heuristics, the second study of this thesis indicates that cybercriminals are aware of these vulnerabilities, with evidence showing that they systematically weaponise interpersonal trust in MIM phishing campaigns. The findings show that cybercriminals exploit trust and habitual interaction to implement the persuasion principles of liking and social proof during MIM phishing, leveraging the informal nature of MIM platforms and interpersonal connections to disseminate their campaigns.

Thirdly, the findings of this thesis demonstrate that while phishing in MIM apps employs familiar URL manipulation techniques found in email phishing, its uniqueness lies in the consistent combination of complex URLs with deceptive

link previews. This approach makes parsing URLs difficult for users while instilling a false sense of legitimacy.

Overall, the research outcomes revealed the unique contextual and technical characteristics of MIM apps that together influence user susceptibility to phishing in such platforms. The outcomes of this thesis also underscore the need for phishing awareness programs and security features that consider link previews, habitual behaviours and social context within MIM apps.

1.5 Research Scope

The development of effective phishing prevention techniques is inherently challenging due to the semantic nature of phishing attacks, which rely on deceptive strategies to persuade users to disclose sensitive information. As a result, current efforts to mitigate phishing attacks extend beyond technical measures and encompass aspects such as understanding the human factors involved in phishing attacks, including user behaviours that may increase their risk of being phished, as well as the psychological manipulation used in phishing attacks. The scope of this research is limited to understanding the factors that make users of MIM apps susceptible to phishing attacks in such apps, as well as the characteristics of the apps that aid attackers in crafting deceptive phishing messages. The first study of the thesis focused on users of WhatsApp, Signal, Slack, Line, Telegram, and Viber. The choice of these platforms was due to their widespread usage and the range of functionalities they offer, including group communication, link previews, link sharing, forwarding messages/links, and the capability to allow users to join public groups through links shared by group administrators. The second study focused on MIM phishing attacks targeting the users of Telegram, WhatsApp, and Viber. These platforms were chosen because of their widespread use in phishing campaigns by malicious actors (Kaspersky, 2021). The third study focused on WhatsApp users. WhatsApp was chosen because it is currently the most popular instant messaging application (Dixon, 2024) and is the most frequently targeted by phishers (Kaspersky, 2021).

Furthermore, this programme of research does not include any technical approach or solution for detecting/preventing phishing in mobile instant messaging applications, but a series of quantitative and qualitative studies to understand how the mobile instant messaging communication medium influences phishing tactics and user behaviour. The Quantitative studies used online surveys to collect self-reported data from participants about their behaviours and intentions, whilst the qualitative study used secondary data to explore and analyse the phishing strategies used in mobile instant messaging phishing attacks.

1.6 Research Contributions

This research makes contributions to the field of information security by advancing our understanding of phishing in the context of MIM apps.

Despite MIM phishing being an emergent threat, there is a paucity of studies that examine how the distinctive contextual and technical characteristics of MIM apps influence the phishing strategies and techniques used by attackers and the vulnerabilities experienced by users. This thesis filled this gap by providing insights into how the distinctive affordances of MIM apps heighten users' susceptibility to phishing and also how attackers strategically exploit these same features to design and disseminate phishing campaigns in ways that differentiate them from other communication media. Other contributions of this thesis include demonstrating that participants' susceptibility to phishing in MIM apps is shaped by the relational context in which messages are received. This finding advances current understandings of phishing susceptibility by highlighting the role of interpersonal trust and social network dynamics, thereby offering a novel perspective that complements existing research. The research also contributes to the literature on the impact of habituation on susceptibility to phishing by highlighting that habitual WhatsApp usage increases susceptibility to phishing.

1.7 Thesis Outline

The structure of the thesis is as follows:

- Chapter 2 presents the background and the related work of this thesis.
- Chapter 3 discusses research methodology and the research strategies adopted in this thesis. The chapter justifies why each method was deemed appropriate and discusses the data analysis strategy used.
- Chapter 4 presents the exploratory study investigating MIM apps' users' behaviours towards links shared during instant messaging. This chapter also describes the data collection method and data analysis strategy, and highlights the key findings from the study.
- Chapter 5 presents the qualitative content analysis (QCA) study that investigates the tactics and techniques used by cybercriminals to lure MIM apps users into visiting fraudulent pages. This chapter also describes the data collection method, data analysis strategy, and highlights the key findings from the study.
- Chapter 6 presents a factorial survey (FS) that investigates the impact of specific characteristics of MIM apps on user susceptibility to phishing in such apps, using WhatsApp as a case study. This chapter also describes the data collection method and data analysis strategy, and highlights the key findings from the study.
- Chapter 7 concludes the key findings of this thesis, provides recommendations for MIM platforms, software engineers, and security awareness professionals, and provides avenues for future work.

Chapter 2

Background and Related Work

The purpose of this chapter is to provide a comprehensive overview of the existing literature on phishing and to identify gaps in current knowledge that underpin the research objectives of this PhD thesis. The chapter begins with a specification of the literature search and review process. It then goes on to provide a general discussion of phishing attacks by briefly introducing and defining the term. It then discusses the life cycle of a phishing attack, its history, and evolution, paying particular attention to its transition from desktop computers to smartphones. The chapter proceeds to classify phishing based on the technology and transmission medium used. The chapter then discusses current efforts to counter the problem of phishing, highlighting the human weaknesses and medium characteristics that may aid phishing, while also identifying gaps in the current literature which are worthy of further study in this thesis.

2.1 Literature review method

The literature review for this thesis adopted a narrative approach, which enables the synthesis of diverse aspects of the research topic to provide a broad overview of the field. Unlike a systematic literature review, this approach is wider in scope and less structured, offering an initial conceptual understanding rather than an exhaustive analysis. However, the structure of the literature and the decision on which areas of the research topic to review were agreed upon by discussing with academic advisors. Particularly, it was agreed to divide the literature review into four sections, with the first section reviewing the fundamentals of phishing, covering definition, types, classification, life cycle and history. The second aspect

of the review examined academic literature that shed light on our understanding of phishing, looking at studies on investigating the characteristics of users that make them susceptible to phishing attacks, those focused on understanding external factors such as features of phishing materials that increase user susceptibility and studies that highlight phisher tactics and techniques that contribute to susceptibility. The third part of the review focused on technical solutions developed to address the problem of phishing, whilst the last part reviewed the literature on phishing awareness training. The actual process started with the identification of relevant literature via non-systematic sources (Google Scholar), but also other databases, such as Web of Science and Science Direct, were searched. The review also employed backward citation chasing, in which the reference lists of the latest journal articles on the topic of interest were examined to identify additional relevant sources. Boolean operators were used to create a search string combining "phishing" or "social engineering" with the relevant topics of interest to find initial articles. For example, "phishing AND user characteristics" was used to identify articles discussing user characteristics that make them susceptible to phishing attacks. For each article, the title, abstract, and conclusions were used to decide if it was relevant. The gaps and issues identified in the existing literature informed the focus of this thesis.

2.2 What is phishing?

Phishing is a type of cyberattack in which cybercriminals deceive individuals into disclosing sensitive information, such as usernames, passwords, or other personal details. It is currently considered the most common attack, utilising social engineering techniques (Nicholson, Coventry and Briggs, 2017)

The literature's definition of phishing is inconsistent, and there is no single acceptable definition. Various definitions appear in the literature, but the most popular ones are presented in Table 2-1

Table 2-1: Various definitions of Phishing

Organization or Author(s)	**Definition**
Anti-Phishing Working Group (APWG)	"*a crime employing both social engineering and technical subterfuge to steal consumers' personal identity data and financial account credentials. Social engineering schemes prey on unwary victims by fooling them into believing they are dealing with a trusted, legitimate party, such as by using deceptive email addresses and email messages. These are designed to lead consumers to counterfeit Web sites that trick recipients into divulging financial data such as usernames and passwords. Technical subterfuge schemes plant malware onto computers to steal credentials directly, often using systems that intercept consumers' account user names and passwords or misdirect consumers to counterfeit Web sites*" (APWG, 2020).
Cisco	*"the practice of sending fraudulent communications that appear to come from a reputable source. It is usually done through email. The goal is to steal sensitive data like credit card and login information or to install malware on the victim's machine*"(CISCO, 2021).

PhishTank	*"fraudulent attempt, usually made through email, to steal your personal information"*(PhishTank, 2021).
UK National Cyber Security Centre	*"when attackers attempt to trick users into doing 'the wrong thing', such as clicking a bad link that will download malware, or direct them to a dodgy website."* (NCSC, 2018).
Krombholz *et al.*, 2015	*"is the attempt to acquire sensitive information or to make somebody act in a desired way by masquerading as a trustworthy entity in an electronic communication medium"*(Krombholz *et al.*, 2015).
Khonji, Iraqi and Jones (2013)	*"a type of computer attack that communicates socially engineered messages to humans via electronic communication channels in order to persuade them to perform certain actions for the attacker's benefit"*

The definitions in Table 2-1 show the lack of a generic definition for phishing. However, analysing these definitions reveals that they fall into two groups, with one group (i.e. Krombholz et al., PhishTank) defining phishing as an attack focusing solely on stealing personal information from users through electronic communications and another group (i.e. NCSC, APWG and Cisco) that defines phishing as both the stealing of users' data through fraudulent websites and the planting of malware on users' machines. Despite being very broad, the definition by Khonji, Iraqi and Jones (2013) tends to offer the most complete definition of phishing as it covers all the elements from other definitions. In addition, Khonji, Iraqi and Jones (2013) also add other media through which phishing attacks occur to cover all other situations (i.e. instant messaging applications).

Nevertheless, the literature broadly classifies phishing attacks into two categories: deceptive phishing and malware-based phishing attacks(Gupta *et al.*, 2017).

Deceptive phishing tricks victims into revealing confidential information to fake websites (Kumaraguru, 2009). Malware phishing involves phishers using phishing emails and websites to plant and execute malicious code into a victim's machine. This thesis defines phishing as an attempt by cybercriminals to obtain sensitive details from internet users by masquerading as trustworthy entities and tricking them into visiting fraudulent websites or downloading malware.

2.2.1 The life cycle of phishing attacks

Phishing attacks involve six stages, from planning to removing all evidence (Gupta *et al.*, 2017). These stages are described below:

Planning: In this stage, the phisher identifies the target. A target could be an individual, an organisation, or even a nation. At this stage, the phisher researches the target, intending to get more information about them. The information can be obtained by physically visiting the organisation, utilising techniques such as dumpster diving to extract information from trash or conducting open-source intelligence research to gather information from online platforms, including social media.

Set-up: In the set-up stage, the phisher sets up the attack material using a suitable means, e.g., an email with a malicious link to direct the victim to a spoofed website.

Sending attack materials: In this stage, the phisher selects a suitable vector for transmitting the phishing material. Many vectors can be used for phishing attacks, including emails, instant messengers, SMS, chat rooms, blogs and social networking sites (Kumaraguru, 2009; Gupta *et al.*, 2017). Regardless of the vector used, the primary goal of phishers at this stage is to prompt users to click on the

link provided in the phishing material. In email phishing, phishers collect targets' emails using open-source intelligence tools (OSINT). These tools enable phishers to scrape information from platforms like LinkedIn. LinkedIn is a popular professional networking site which allows professionals to network and interact with one another. For smishing and instant messaging phishing, phishers can extract the phone numbers of targets using OSINT tools or by joining public instant messaging groups (Garimella and Tyson, 2018).

Infiltration/Break-in: The attacker infiltrates the user's system once the target responds to the phishing message. Infiltration can occur through malware installed on the user's machine, following the user's click on a link, which grants the attacker privileged permission to modify the system's configuration or access rights (Gupta *et al.*, 2017). For attacks involving a credential harvester through fake websites, the break-in stage starts when the attacker intercepts user details from phishing pages.

Data Collection: Once the phisher gains access to the target's system and collects the user's credentials, they can extract as much data as necessary from the user's account or system. They can save this information in a file to collect later or forward it directly to their email (Kumaraguru, 2009). Using this information, the attacker can access the victim's account, leading to financial loss or identity theft. Suppose malware is installed on the victim's machine. In that case, the attacker can have remote access to the victim's system and collect any interesting data. In some instances, phishers sell this data to other cybercriminals, who may use it to conduct various forms of fraud, such as identity theft.

Breakout/exfiltration: This is the final stage of the phishing life cycle. It involves removing all evidence that could lead to tracing the phishers. Such evidence includes the phishing website, which phishers take down once the attack is completed. Phishers also use this stage to reflect on the attack by analysing what made it successful. They then use this knowledge to improve future attacks.

2.2.2 History and evolution of phishing

The history of phishing can be traced back to the 1990s, during the time of America Online (AOL) (Phishing, 2021). The motives behind attacks in those days were to gain free access to the internet using stolen credentials. During that period, hackers created accounts using fake credit card details and screen names that imitated AOL administrators (Cofense, 2021). They used these accounts to phish real AOL users' credit card and account login details. Phishing was carried out during this period using AOL's email messaging functionality. Cybercriminals exploited this functionality to send messages to users, requesting that they verify their login credentials or confirm their billing information. In 1994, this process was simplified by a software called AOHell, which aided hackers in crafting phishing messages and automatically sending them to AOL users (Rekouche, 2011).

In the 2000s, email was still the most popular medium for conducting phishing. For example, in 2001, phishers targeted online payment systems with an attack on E-Gold, a digital gold currency company. Although the attack on E-Gold was considered unsuccessful, it paved the way for more sophisticated attacks to be carried out (Phishing, 2021). Similarly, in 2003, phishers registered numerous fake domains that imitated legitimate sites, such as eBay and PayPal. eBay is a global online marketplace that allows individuals and businesses to buy and sell goods online, whilst PayPal is an online payment platform that enables individuals and businesses to send, receive, and manage payments electronically. Phishers utilised malicious worm programs to send fake emails to customers of both eBay and PayPal, requesting that they update their credit card details, which were later stolen.

In 2006, phishers started targeting mobile phone users through the Short Message Service (SMS). SMS is a ubiquitous capability integrated into the wireless standard GSM. It allows individuals to send or receive messages using their smartphones. Phishers used this capability to send users fraudulent messages

with phishing links. This new form of phishing was first discovered in 2006 by security researchers at McAfee (CIO, 2006a). The move to mobile phones demonstrates cybercriminals' tendency to adapt to technological advances, as 2006 marked a significant increase in mobile phone adoption worldwide (Cofense, 2023).

VoIP enables the transmission of voice communications as data packets over the Internet Protocol (IP) (Varshney *et al.*, 2002). With the rise of Voice over Internet Protocol (VoIP) in the mid-2000s, phishers shifted their attention to conducting phone-based scams, often utilising VoIP features such as caller ID spoofing to defraud unsuspecting users (CIO, 2006b).

Quick Response (QR) codes, invented in 1994 by a Denso Wave employee named Masahiro Hara (Microsoft, 2023a), are two-dimensional matrix barcodes used for encoding information. QR codes allow users to embed a unique identifier, which can be used to verify resources against a backend database. They enable the embedding of Universal Resource Locators (URLs) that grant immediate access to web information when scanned by users' smartphone cameras (Sharevski *et al.*, 2022). The year 2020 witnessed exponential growth in the use of QR codes to access online resources, largely driven by the COVID-19 pandemic, which necessitated contactless access as a precaution against the spread of the disease. With QR codes becoming very popular among internet users, cybercriminals started to exploit this technology by embedding malicious URLs in QR codes and inducing users to scan them using their mobile phones.

Recent technological advances and the increasing use of smartphones are transforming the way people communicate, particularly with the rise of social media platforms. Social media platforms, including Facebook, LinkedIn, and X, allow individuals to connect and communicate with friends and family. These platforms also enable businesses to inform customers about new services and products (Trend Micro, 2024). The popularity and use of these platforms have led to phishing moving to these platforms (Gillin, 2023), with cybercriminals

exploiting weaknesses present in these platforms, such as the ability to create fake pages and accounts to propagate phishing campaigns.

More technological advancements led to the introduction of mobile instant messaging (MIM) applications in 2009 (Taivaskero, 2017). Using users' existing data plans, these applications enable users to instantly transmit and receive text, location-based messaging, photos, audio, and video messages to individuals or groups of friends. The real-time transmission of messages by these apps makes the interaction between users natural and fast compared to other messaging systems that rely on cellular networks. However, as communication moves to instant messaging, phishing follows. Cybercriminals now use MIM apps to spread phishing attacks (Kaspersky, 2021). They utilise the functionalities of these apps to propagate their campaigns, relying on the apps' features, such as users' ability to share messages to create a 'chain scheme', a practice where a cybercriminal asks a user to share malicious links with their connections (Kaspersky, 2021).

The move to MIM by cybercriminals has shown that as technology advances and becomes more ubiquitous, cybercriminals continue to adapt. Phishing attacks have transitioned fluidly between email and mobile instant messaging, one of our most reliable services.

2.2.3 Types of phishing

At a very high level, phishing can be classified into two broad categories: general and spear phishing. These categories are explained below:

General phishing

General phishing is a type of phishing where the phisher targets multiple users by sending large volumes of fake phishing messages to them, requesting that they click on links within those messages (Gupta *et al.*, 2017). These messages can be sent through various mediums, including emails, SMS and instant messaging platforms. The attacker follows the typical life cycle of phishing, as highlighted in

2.1.1. Generally, this involves setting up a phishing website and disseminating many fraudulent messages to many users. These messages prompt recipients to click on a link and provide personal information, potentially compromising their identity and other sensitive data. The information the victims offer is transmitted to the phisher via a phishing server. The phisher can use this information for identity theft to obtain illicit financial advantages or alternative objectives. A key point about general phishing is that it targets a vast number of recipients.

Spear Phishing

In 2007, spear phishing —a new form of phishing —emerged and remains active today (Smadi, 2017). Unlike the previous approach, where phishers indiscriminately send messages to many users, this approach is targeted, with phishing emails customised to appear as if they come from someone the target knows. For instance, phishers can send an email to an employee of an organisation, pretending that the IT department is making an upgrade and they need the employee to update their details. It is believed that technological advancements and the rise of social media have contributed to the success rate of spear phishing attacks (Cofense, 2021). These platforms now make it easier for cybercriminals to extract information relating to organisations and employees, which they can use to conduct more targeted attacks on users. Spear phishing cases increased dramatically from 2009 to 2011 (Caputo *et al.*, 2014). In some instances, phishers may target executives with high-level access to organisational resources and information. This form of spear phishing is referred to as Whaling (Goenka, Chawla and Tiwari, 2024). Phishers exert considerable effort in creating whaling attacks, frequently incorporating business terminology and encompassing personal details regarding the targeted individual, aiming to mitigate any suspicion. As a result, the success rate of this form of attack tends to be very high, as evidenced by an incident in 2019 where a cybercriminal stole $100 million from Google and Facebook (Jr., 2019).

2.2.4 Phishing classification based on the medium of circulation

All phishing attacks occur through some medium. These media provide phishers with a platform to target unsuspecting users**.** Section 2.1.2 provided the history and evolution of phishing, highlighting that phishers follow technological advances by leveraging popular communication media to propagate their campaigns. Figure 2-1 shows the phishing classification suggested in this thesis. At the first layer, the classification reveals the technology being abused by cybercriminals to conduct phishing campaigns, aligning with the information presented in Section 2.1.2. The second layer displays the various phishing types, categorised by the medium of distribution. These are discussed in the coming paragraphs.

Email Phishing: Email phishing is a type of phishing where cybercriminals target users via the email communication medium. They aim to get users to click on fraudulent links and divulge personal information, such as account login credentials, or download malware on their systems. Email is the most prevalent phishing attack vector (Gold, 2023). Email phishing typically occurs through two primary methods: email spoofing and the use of compromised accounts. Email spoofing involves the transmission of phishing messages that appear to originate from a legitimate sender address (Maroofi *et al.*, 2021); spoofed emails are designed to imitate trusted senders such as business partners, financial institutions and educational institutions. This approach enhances the success of such attacks, as the victims may believe they are interacting with a legitimate entity. For example, cybercriminals can send a phishing email from a spoofed business partner, instructing the recipient to click a link that will ultimately redirect to a fake Microsoft SharePoint page.

Another example is emails warning users that their credit cards have been blocked due to the detection of fraudulent activity on their accounts, or emails from the IT department suggesting that the user's account will be deactivated if they fail to change their password. The content of these emails often appears real

in terms of language and authenticity cues. Additionally, cybercriminals may employ strategies such as fear appeals, urgency, and greed to elicit a response from the target.

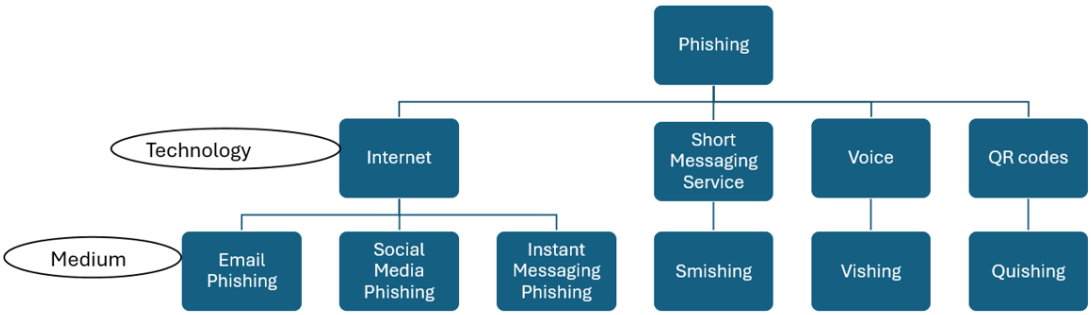


Figure 2-1: Classification of phishing based on technology and medium of circulation

Social media phishing is often carried out through popular social networking sites, including Facebook, Twitter, Instagram, and LinkedIn. As noted in Section 2.1.2, Social media platforms play a key role in the lives of individuals and organisations. As of February 2025, there were 5.24 billion social media users, equivalent to 63.9% of the global population (Statista, 2025). Cybercriminals frequently utilise these platforms to disseminate phishing scams through impersonation, credential theft, and data collection (Goenka, Chawla and Tiwari, 2024).

Some popular phishing attacks on social media have been documented (Trend Micro, 2024). Examples of these attacks include cybercriminals creating a fake Instagram page and sharing links with unsuspecting users, who may give out their login details, believing they are submitting them to a real Instagram account. In another instance, cybercriminals create fake Facebook pages that impersonate prominent brands and personalities, using these pages to spread fake free prizes and giveaways. They seek to elicit sensitive information from users, such as credit

card details and other personal information. There have been specific documented instances of phishers conducting employment scams on LinkedIn, soliciting information such as bank account numbers, mailing/physical addresses, and Social Security numbers from unsuspecting users (Smith, 2020). Similar instances of Facebook phishing show phishers creating fake profiles on Facebook and sharing malicious links (Arsen, 2024).

Instant messaging phishing occurs via instant messaging platforms**,** such as WhatsApp, Telegram, and Viber. Instant messaging phishing can occur on both desktop and mobile versions of these platforms. However, recent data suggests that most instant messaging phishing attacks target users through MIM apps (Kaspersky, 2021). Notwithstanding, an example of an attack using the desktop version of these apps was documented by the Google Threat Analysis Group, where cybercriminals target individuals in the security domain (Sheridan, 2021) via the Telegram Messenger app. The threat actors used Telegram to establish trust with security researchers through constant interaction and by sharing links to their blogs and videos of claimed exploits. The attacker then asks if the target wants to work on a security project. Once this request is accepted, they will share malicious code and links that could compromise the target system if clicked.

On the other hand, in 2021, Security researchers at CheckPoint discovered a malicious app on the Google Play Store called FlexiOnline, which spread via the mobile version of WhatsApp (Checkpoint, 2021). The app is a fake service that claims to give users free access to Netflix content. However, instead, it monitors users' WhatsApp incoming message notifications and sends automatic replies using content it receives from a command-and-control server (Checkpoint, 2021).

Cybercriminals also targeted internet users, using phishing campaigns that exploited reputable brands. Cybercriminals host web content offering high-value incentives, such as free tickets, low-cost products or discounted gift cards. (Cuddeford, 2018). The attack commenced with the hackers transmitting

messages containing links to their victims using the target's Mobile instant messaging platform. In these campaigns, phishers inform their target of the limited time to complete a task before the offer attached to that task is withdrawn, invoking a sense of urgency in the target and making them act without considering the consequences of their actions. According to Windels (2018), in most cases, surveys and forms are embedded in the pages used in these new attacks, and, therefore, they aim to extract information from users. In addition, they instruct users to automatically share content with other users in their instant messaging contact list.

Utilising MIM apps can aid cybercriminals in propagating their campaign to many users. This approach may also exploit users' trust in their contacts, as the messages appear to come directly from the user's contacts.

Smishing: Smishing is a type of phishing where cybercriminals target users via their mobile phones' Short Message Service (SMS) capabilities. Like email phishing, a smishing attack begins with cybercriminals sending text messages with links to their targets. These texts claim to come from a known and trusted entity, such as the user's bank or an online merchant with whom the user has recently interacted (Yeboah-Boateng and Amanor, 2014). The messages contain persuasive content that aims to manipulate users into clicking on a link, replying to the attacker, or calling back a number. Clicking on links within such messages will lead to fake websites that request sensitive information from users. Such information may include credit card details and login credentials. The links sometimes contain malicious code or .apk files. These files or code may contain malware that performs malicious activities on the victim's phone (Mishra and Soni, 2020). Sometimes, smishing messages contain self-replying SMS where users are asked to agree or disagree with a supposed subscription (Goenka, Chawla and Tiwari, 2024). This form of attack aims to confirm if the number is valid so that the criminals can follow up with a more targeted attack later.

Vishing: Voice messaging is another medium phishers use to defraud users. The term “Vishing" refers to this type of attack. It entails using IP-based voice messaging technologies, i.e., voice over Internet protocol (VoIP), to lure victims into providing their personal, financial, or other sensitive information to attackers. These often involved calls from automated systems informing victims of alleged credit card misuse and requesting they call fraudulent numbers and divulge personal information (CIO, 2006b). Unsuspecting users sometimes receive calls from a live person or recorded messages (FBI, 2007). These messages or calls direct users to take specific actions, which ultimately lead to the compromise of their accounts. Vishing attacks may also mislead users into installing malware, primarily conducted from fraudulent call centres (Microsoft, 2023b).

One of the most notable real-world vishing attacks was the incident on Twitter in July 2020. In what was termed a phone spear phishing attack, cybercriminals contacted Twitter employees via phone calls. They employed fake identities to deceive them into providing credentials, which they later used to access an internal corporate tool, enabling them to reset passwords and two-factor authentication configurations of some targeted user accounts, including those of high-profile individuals and celebrities (Wired, 2020).

QR code phishing, also known as quishing**,** is a type of phishing where cybercriminals embed malicious URLs in QR codes and induce users to scan them using their mobile phones. Upon scanning, the user is redirected to a fake web page soliciting users' personal information. Quishing can occur in two ways: open space QR code fraud and Email with QR codes (NCSC, 2024). In the former, cybercriminals placed fraudulent QR codes in public spaces (such as stations and car parks) to trick unsuspecting and curious users. Documented evidence reported a victim losing £13,000 in a railway station QR code scam (BBC, 2023). In email phishing with QR codes, cybercriminals send fraudulent emails containing QR codes that, when scanned, direct users to phishing websites (Egress, 2024b). Most QR code readers display a notification stating the URL to

which the user will be redirected; however, URL shorteners limit the user's ability to review the URL before opening it.

2.3 Understanding phishing

Developing phishing prevention techniques is challenging due to the semantic nature of phishing, which employs deception to entice users into disclosing sensitive information. Section 2.1.2 highlighted the evolving nature of phishing attacks, demonstrating how phishing tactics adapt to technological advances. Since the 1990s, when phishing emerged as a significant cybersecurity threat, extensive research has been conducted to improve our understanding of phishing. This type of work usually aims to facilitate the development of novel mitigation techniques and enhance existing protective measures. Therefore, this section highlights and discusses the literature that sheds light on phishing by looking at studies investigating the characteristics of users that make them susceptible to phishing attacks (Vishwanath *et al.*, 2011), those focused on understanding external factors such as features of phishing materials that increase user susceptibility (Dhamija, Tygar and Hearst, 2006; Alsharnouby, Alaca and Chiasson, 2015), and studies that highlight phisher tactics and techniques that contribute to susceptibility. This section describes the complex relationship between human elements, such as personality traits and demographics, and external factors, including phishing strategies and medium/context features, that affect phishing susceptibility.

2.3.1 User characteristics

Phishers exploit human cognitive and behavioural traits to deceive and manipulate victims (Abroshan *et al.*, 2021). Although some technical solutions can detect and block phishing before it reaches the user's inbox. However, once the phishing message bypasses technical solutions, its success depends on human decision, trust, and other cognitive factors (Abroshan *et al.*, 2017). Consequently,

understanding user susceptibility to phishing requires an understanding of user behaviours that may increase their risk of being phished. The behaviour of individuals in terms of how they process online information has been linked to phishing susceptibility (Harrison, Svetieva and Vishwanath, 2016). The literature outlined two forms of information processing: systematic and heuristic information processing (Chaiken, 1980). Systematic processing refers to a cognitive approach in which individuals form judgments through careful and deliberate evaluation of the quality and strength of arguments presented within a persuasive context, whilst heuristic processing is a cognitive processing approach that relies on simple decision-making rules, often triggered by peripheral cues within the context, to facilitate rapid judgment formation (Vishwanath, Harrison and Ng, 2018). In an email phishing simulation study, the authors (Vishwanath, Harrison and Ng, 2018) tested the effect of information processing on individuals' suspicion about the authenticity of a phishing email. The study's results showed that systematically processing phishing emails increases user suspicion, meaning that those who carefully and deliberately evaluate phishing emails are more likely to detect them. A similar result was also found for social media phishing, as the findings from an online questionnaire study show that heuristic processing significantly influenced susceptibility to phishing on social media (Frauenstein and Flowerday, 2020).

In more recent research, the authors (Gan, Lee and Liew, 2024) tested the effect of message involvement, as a systematic information cue, on susceptibility to instant messaging phishing. The authors found that phishing messages with a desirable or relevant topic (high message involvement) were positively related to the instant messaging phishing susceptibility.

While the above findings highlight the effects of information processing on phishing susceptibility in different media, it is unclear whether these findings are similar for smishing and quishing, since such results are lacking in the literature.

In another online questionnaire study, the researchers (Alohali *et al.*, 2018) used the Big Five Inventory Scales (BFI) test to investigate the relationship between personality traits — namely, extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience — and security-related behaviours. The results show that personality traits can predict security-related behaviours, including actions such as clicking on links in emails from unknown sources, opening attachments from unknown sources, forwarding chain emails, clicking on links from known sources without verifying their legitimacy, and deleting dubious emails.

Another online survey study (Frauenstein and Flowerday, 2020) shows that the personality trait measures from the BFI significantly influence both heuristic and systematic processing, potentially leading to increased susceptibility to phishing on social networking sites (Frauenstein and Flowerday, 2020). The findings specifically show that the personality traits of agreeableness, openness and neuroticism are associated with heuristic processing, whilst neuroticism is significantly positively related to systematic processing.

In another study, the researchers (Abroshan *et al.*, 2021) used a simulated phishing exercise with the Balloon Analog Risk Task (BART) to investigate how participants' risk-taking behaviour influences their susceptibility to phishing. The results show that a high level of general risk-taking can increase the possibility of clicking on a phishing link.

Habit, defined as the degree to which users automatically perform behaviours (Limayem, Hirt and Cheung, 2007), can also affect individuals' susceptibility to phishing. Existing literature has shown that when individuals cultivate a habit of specific behaviour, they tend to do it automatically and unconsciously, without conscious consideration (Orbell *et al.*, 2001). Interestingly, existing literature has shown that individuals' email and social media habits affect their susceptibility to phishing (Vishwanath, 2015; Williams and Polage, 2019). A study (Williams and Polage, 2019) found that individuals whose responses to email

communication become habitual and cannot control their behaviour are more likely to respond to phishing emails. Similarly, Vishwanath (2015) found that individuals who frequently use Facebook, those with an extensive social network, and those who cannot control such behaviours are more susceptible to phishing on Facebook. Similarly, research on smishing has shown that participants who used their mobile phones for 11-20 hours per week were more likely to fall for smishing attacks than those who used them for 6-10 hours (Rahman *et al.*, 2023). However, researchers have yet to investigate the relationship between the frequency or habitual usage of mobile instant messaging apps and susceptibility to instant messaging phishing. Given the increasing prevalence of instant messaging in people's lives, the frequent use of these applications may develop into a habit, potentially impacting users' susceptibility to phishing attempts. Thus, investigating whether the frequency with which one uses MIM apps impacts one's susceptibility to instant messaging phishing will add to our understanding of the problem of phishing in MIM apps**.**

In addition to the above behavioural factors, researchers have found that knowledge, demographics, perception, and beliefs are other factors that influence susceptibility to phishing (Zhuo *et al.*, 2023). In terms of knowledge, individuals can acquire knowledge through learning and direct training. For example, they can develop knowledge about phishing through training or other awareness programs. Individuals can also gain knowledge through experience (i.e. when they fall for phishing). Evidence from the literature has shown that knowledge can impact how users respond to phishing messages (Zhuo *et al.*, 2023). In examining the correlation between knowledge and phishing, researchers typically emphasise either the user's technical proficiency or their awareness of phishing threats (Zhuo *et al.*, 2023), as either of these can influence how users respond to phishing attacks.

More than a decade ago, researchers employed a role-based scenario and an online questionnaire to evaluate the impact of users' computer knowledge on their susceptibility to phishing (Pattinson *et al.*, 2012). Findings from that study

revealed that familiarity with computers positively affected users' ability to manage phishing messages effectively. Similarly, another study, based on a simulated phishing experiment, found that individuals with higher knowledge and experience with email were less likely to fall for phishing (Harrison, Svetieva and Vishwanath, 2016). The same study also found that a lack of attention to email elements was associated with a high likelihood of falling for phishing attacks. In another study (Graham and Triplett, 2017), an online questionnaire was used to investigate the effect of security and privacy knowledge on phishing susceptibility. The researchers found that individuals with higher knowledge of internet security and privacy were less likely to respond to phishing emails. More recently, the researchers (Sturman *et al.*, 2024) used an online email sorting experiment to investigate the influence of phishing knowledge, cue utilisation, and decision on phishing susceptibility. The results showed that knowledge of phishing leads to increased phishing detection but is also associated with a tendency to label all emails as phishing.

Whilst research on the influence of users' technology capabilities on smartphone phishing or smishing is limited, a recent questionnaire-based study highlighted that users with greater awareness and understanding of smishing were better equipped to identify smishing scams (Edwards *et al.*, 2023). Similar results were found for social media phishing, with a questionnaire study revealing that participants with greater security knowledge exhibited less susceptibility to phishing on social media (Algarni, Xu and Chan, 2015). These findings were also similar to quishing, as a questionnaire-based study shows that those with knowledge and awareness of quishing were less likely to fall for quishing attacks (Singkeruang, Susanto and Saeni, 2025). In a recent online questionnaire study, the authors (Gan, Lee and Liew, 2024) found that knowledge about phishing does not directly affect Gen-Z's vulnerability to instant messaging phishing; however, it significantly encourages them to implement appropriate online security management practices on instant messaging applications. There is currently no

literature outlining the influence of knowledge/technical competence or awareness of vishing attacks on susceptibility to vishing.

The findings from the above studies demonstrate the link between security and technology-related knowledge and phishing susceptibility, justifying why organisations adopt anti-phishing training to increase user knowledge of phishing.

In addition to knowledge, other researchers have studied the relationship between user demographic variables and phishing susceptibility, with most studies focusing on age and gender (Oliveira *et al.*, 2017; Taib *et al.*, 2019). However, the literature currently provides inconsistent results regarding the importance of these variables on phishing susceptibility. For example, the simulated email phishing experiment (Jagatic *et al.*, 2007) shows that females were more likely to fall for phishing than males. Similarly, the role-play email phishing experiment conducted by Sheng *et al.* (2010) shows that females were more susceptible to email phishing than males. A more recent study also shows similar results, with older women found to be more susceptible to a simulated spear phishing attack (Lin *et al.*, 2019). Whilst these studies show that females are more vulnerable to email phishing, other studies reported that males were more susceptible. For example, the two large-scale national studies conducted in Tjostheim and Waterworth (2020) show that males were more susceptible to phishing than females. However, other email-based phishing studies reported that gender does not contribute to susceptibility to email phishing (Taib *et al.*, 2019; Lain, Kostiainen and Čapkun, 2022; Ribeiro, Guedes and Cardoso, 2024). However, the effect of gender was found in social media phishing, as existing literature found that females are more susceptible to social media phishing than men (Algarni, Xu and Chan, 2015). Gender was also not associated with an increased likelihood of falling for smishing attacks, according to a recent simulated SMS phishing experiment (Rahman *et al.*, 2023). The literature currently lacks empirical findings regarding the effects of gender on an individual's susceptibility to quishing, vishing and instant messaging phishing.

Research on the effect of age on phishing susceptibility has also found inconsistent results. For example, a role-play scenario experiment study (Parsons *et al.*, 2013) found no evidence of a relationship between age and the ability to manage phishing emails correctly. Similarly, a more recent lab experiment study (Sarno *et al.*, 2020) found that age did not influence users' classification of phishing emails. However, older participants were more careful when classifying emails. Studies such as those by Sheng *et al.* (2010) and Sarno *et al.* (2017) have found that younger users are more susceptible to phishing, while others have found that older users are more vulnerable to phishing (Greitzer *et al.*, 2021; Grilli *et al.*, 2021). This finding is similar to those for social media phishing, which show that younger participants are more likely to fall victim to social media phishing attacks (Algarni, Xu and Chan, 2015). However, a study on the influence of age on susceptibility to smishing attacks revealed that individuals aged 45-54 exhibit a greater propensity for falling victim to smishing scams compared to other age demographics, with the 18-24 age group being the least susceptible to smishing attacks (Rahman *et al.*, 2023). Thus, similar to the results on gender, these findings reveal the inconsistency of the age effect on susceptibility to phishing across email, SMS, and social media. The literature currently lacks empirical findings regarding the effects of age on an individual's susceptibility to quishing, vishing and instant messaging phishing.

Users' occupation or academic major is another demographic factor that has been studied extensively in the literature. The idea is that since phishing is carried out via technology-enabled media, users with academic training in STEM subjects, such as IT, should be less susceptible to phishing scams. In their simulated email phishing experiment, Diaz, Sherman and Joshi (2020) found empirical evidence linking academic training and cybersecurity knowledge to reduced susceptibility to email phishing. Similarly, another simulated phishing study (Taib *et al.*, 2019) found that employees with fewer years of experience, specifically those in their first year of employment, were more vulnerable to email phishing than those with many years of experience (i.e. 5 to 10 years). On the other hand, employers who

used computers more frequently in their jobs were more susceptible to email phishing (Lain, Kostiainen and Čapkun, 2022). However, job/profession didn't influence the likelihood of individuals responding to smishing attacks (Rahman *et al.*, 2023). The literature currently lacks empirical findings regarding the effects of occupation/profession on susceptibility to quishing, vishing and instant messaging phishing.

Our perceptions of threats and beliefs about our capability to safeguard ourselves from such threats can also affect how we respond to phishing. Perception of threat refers to an individual's assessment of the threat in a particular situation (Witte, 1992). Perceived threat refers to both how individuals perceive the consequences of falling for phishing (Canfield, Fischhoff and Davis, 2019) and the concerns they have about falling victim to such threats (i.e., concerns about becoming a phishing attack victim)(Wang, Li and Rao, 2017). On the other hand, phishing self-efficacy relates to our self-confidence and ability to recognise phishing scams (House and Raja, 2020). The impact of these factors on phishing susceptibility has been well documented in the literature. For example, an email management study found that participants who believed phishing attacks had more severe negative consequences were more effective at detecting phishing emails(Canfield, Fischhoff and Davis, 2019). However, such groups of participants were also more likely to classify legitimate emails as phishing. This suggests that while our perception of the negative consequences of phishing can reduce our susceptibility to phishing, it can also lead to an increased number of false positive classifications. Similarly, findings from a questionnaire study show that the more concerned individuals are about becoming victims of phishing, the more likely they are to take precautionary measures against it (Lei, Hu and Hsu, 2023).

Regarding phishing self-efficacy, a questionnaire-based study on email phishing has shown that higher perceived phishing self-efficacy or individuals' self-confidence in recognising phishing leads to higher motivation for performing anti-phishing behaviour (Sun *et al.*, 2016). A lab-based study with Facebook users

found that the higher a participant's phishing self-efficacy, the lower their chances of clicking on a phishing link on Facebook (Seng, Al-Ameen and Wright, 2018). However, a recent questionnaire study on instant messaging phishing reveals that the higher the phishing self-efficacy of participants, the greater their chances of becoming victims of instant messaging phishing attacks (Lee, Gan and Liew, 2023a), which contradicts prior studies on email phishing. Thus, further research is needed to confirm whether this is the case, as the findings in Lee, Gan and Liew (2023a) were based on a Malaysian sample**.** Researchers have yet to investigate the relationship between phishing self-efficacy and susceptibility to quishing, vishing, or smishing. The studies highlighted in this section, along with the variables and research methods used in each paper, are presented in **Table 2-2**.

Table 2-2: Current literature on user characteristics and phishing susceptibility

Literature	**User characteristics studied**						**Media**						**Method used**
	Knowledge	**Gender**	**Age**	**Occupation**	**Perception and beliefs**	**Behaviour & Habits**	**Email**	**SMS**	**Social Media**	**QR codes**	**Voice**	**Mobile Instant Messaging**	
(Pattinson *et al*., 2012)	✓						✓						Role-play and Questionnaire
Harrison, Svetieva and Vishwanath, 2016	✓						✓						Simulated phishing experiment

Graham and Triplett, 2017	✔						✔						Online questionnaire
Sturman *et al.*, 2024	✔						✔						Online email sorting experiment
Edwards *et al.*, 2023	✔							✔					Online questionnaire
Algarni, Xu and Chan, 2015	✔	✔	✔						✔				Online role-play experimental questionnaire
Singkeruang, Susanto and Saeni, 2025	✔									✔			Online questionnaire
Gan, Lee and Liew, 2024	✔											✔	Online questionnaire
Jagatic *et al.*, 2007		✔					✔						Simulated phishing experiment
Sheng *et al.*, 2010		✔	✔				✔						Role-play experiment
Lin *et al.*, 2019		✔					✔						Simulated phishing experiment
Tjostheim and Waterworth, 2020		✔					✔						Online questionnaire
Ribeiro, Guedes and Cardoso, 2024		✔					✔						Online questionnaire
Rahman *et al.*, 2023		✔	✔	✔		✔		✔					Simulated phishing experiment

Parsons *et al.*, 2013			✔				✔						Role-play scenario experiment
Sarno *et al.*, 2020			✔				✔						Lab experiment
Sarno *et al.*, 2017			✔				✔						Online questionnaire
Diaz, Sherman and Joshi, 2020				✔			✔						Simulated phishing experiment
Taib *et al.*, 2019				✔			✔						Simulated phishing experiment
Lain, Kostiainen and Čapkun, 2022				✔			✔						Simulated phishing experiment
Canfield, Fischhoff and Davis, 2019					✔		✔						Online role-play scenario experiment
Lei, Hu and Hsu, 2023					✔		✔						Online questionnaire
Sun *et al.*, 2016					✔		✔						Online questionnaire
Seng, Al-Ameen and Wright, 2018					✔				✔				Role-play experiment
Lee, Gan and Liew, 2023a					✔							✔	Online questionnaire

Williams and Polage, 2019						✔	✔						Online questionnaire
Vishwanath, 2015							✔		✔				Simulated phishing experiment
Vishwanath, Harrison and Ng, 2018						✔	✔						Simulated phishing experiment
Frauenstein and Flowerday, 2020						✔			✔				Online questionnaire

2.3.2 Medium Characteristics

The factors that affect susceptibility to phishing are multifaceted, comprising both the user characteristics discussed in 2.2.1, and also the characteristics of the communication medium. In this section of the thesis, a detailed examination of the characteristics of the communication medium and their role in facilitating phishing activities is provided.

Email: Email, or electronic mail, is one of the world's most popular methods of communication. Email is an electronic communication method utilising devices to transmit messages over a computer network (Cloudflare, 2025). Emails are sent using email addresses or accounts, which are strings of characters that identify an email account and allow messages to be sent and received. In addition to having an email account, users also need an email client. The email client is a software program that allows users to send emails from their accounts. Anyone can have an email address, including those with malicious intentions.

In 2023, the worldwide e-mail user base reached 4.37 billion and is projected to increase to 4.89 billion (Statista, 2024). In 2023, approximately 347 billion emails

were transmitted and received daily worldwide (Statista, 2024). Organisations utilise email more than any other mode of communication(Bujang and Hussin, 2013). It is believed that in the absence of emails, businesses can potentially cease to operate effectively (Saleem Khawaja, 2022) because most communication within organisational settings happens via email. Businesses rely on email for advertising goods and services. However, email is inherently an insecure method despite its frequent usage for transmitting personal information. This makes it a preferred vector for conducting phishing attacks. As mentioned in section 2.1.4, phishers can deceive users by sending phishing messages that appear to originate from a legitimate sender address. This tactic is feasible because, in the absence of an appropriate configuration, standard email protocols lack the inherent capability to authenticate the true origin of an email message. Consequently, it is trivial for phishers to alter the header of an email, making the protocols perceive the email as originating from an authentic sender (fortinet, 2024). Email spoofing is possible because email is not significantly different from traditional mail systems. Like the conventional mail system, an email has three main elements: the envelope, the header, and the body.

When conducting email spoofing, phishers can modify critical parts of email headers, including the 'From', 'Reply to', 'From', 'Subject', 'Date', and 'To' fields (Cloudflare, 2025). Because the recipient only sees the information in the email's header and body, they may believe the message is from a legitimate sender. The literature has confirmed that emails from known sources significantly increase users' response likelihood (Moody, Galletta and Dunn, 2017). While modifying the email header is sometimes enough to deceive email users, phishers sometimes impersonate organisations by sending emails with domains that closely resemble those of the organisation being impersonated. The risk of this happening is associated with improper configuration of Sender Policy Framework (SPF), DomainKeys Identified Mail (DKIM) and Domain-based Message Authentication Reporting and Conformance (DMARC). These three

authentication protocols, if rightly configured, can prevent phishers from sending emails on behalf of a domain.

As mentioned earlier, email phishing can also occur from compromised accounts. These are accounts from individuals who have already provided their details to the threat actor by either clicking a phishing link and submitting their account details or clicking a malicious attachment. Spoofing may have led the individual to fall for the attacks. The victims may have responded to the phishing attacks due to a lack of awareness and education, or perhaps the user fell for the attack because Email clients allow users to attach documents or logos and share links with other users, making it possible for cybercriminals to share malicious links and attachments with them. Thus, organisations can only mitigate the risks of phishing by deploying both technical phishing protection tools and user awareness training.

Most email clients generate link previews for email messages that contain URLs or links (Microsoft, 2025). This approach provides users with a quick view of the websites they are about to visit by summarising the content of a shared webpage using its title, description, and images. These fields are retrieved from the HTML code of the web page, either from standard HTML elements or custom meta tags such as Open Graph or Twitter Cards. Whilst this feature may improve the aesthetic aspect of the email environment, it is also vulnerable to exploitation, as evidenced by recent research (Stivala and Pellegrino, 2020).

Short Message Service (SMS): SMS is a ubiquitous capability integrated into the wireless standard GSM, allowing individuals to send and receive messages using their smartphones. SMS allows individuals to send or receive messages using any mobile phone, irrespective of the cellular carrier, implying that cybercriminals can send fraudulent messages to anyone if they have the user's number.

In addition to sending text messages, SMS allows users to share links. Such links can be sent individually or as part of a text or multimedia message. Links are sent

in both media as hyperlinks, meaning they go directly to their corresponding web pages when clicked. When sending a link, users can choose between two formats. First, they can send it as a long URL containing all its components, including the scheme, hostname, path and query string. The second approach is to send it as a shortened URL, which can be created using URL shorteners such as Bitly and Ow.ly. Such shorteners use random letters and characters to obscure the actual URL. Due to the character limitation of SMS, URL shortening is often the preferred approach for SMS marketing, as it enables marketers to truncate long URLs to just a few characters (Tatango, 2020). However, this technique is currently used by phishers to hide the true identity of phishing URLs (Chhabra *et al.*, 2011). Though a recent SMS phishing study found that the participants easily detected smishing messages with shortened URLs compared to those with long URLs (Clasen, Li and Williams, 2021).

Similar to email clients, SMS can present links to users in the form of link previews. Generating a preview requires that the messaging app read the link's landing page, which means mobile phones must be connected to the internet for the technique to work. Due to the aesthetic aspects of link previews, marketing research has shown that such previews increase user click rates (McLachlan, 2021). Thus, this feature could be abused by cybercriminals to trick users into visiting malicious web pages. However**,** there are no empirical findings on the influence of link previews on susceptibility to SMS phishing.

SMS is also vulnerable to caller ID spoofing, also known as number spoofing. Caller ID or Number spoofing is changing the caller ID to a number other than the actual calling number. This is possible because existing caller ID protocols do not offer real authentication. As a result, they cannot be relied upon to authenticate the identities or locations of callers (Mustafa *et al.*, 2016). This vulnerability enables cybercriminals to disguise the number from which they make calls or send texts. With readily available software such as BurnerApp and SpoofCard, spoofing a number is trivial. Many cases of these forms of attacks have been highlighted in the literature (Mustafa *et al.*, 2016).

Instant messaging: Instant messaging is a form of online communication that facilitates the real-time exchange of text, photos, video, and audio via the Internet while also enabling the transmission of emotions using emoticons (Piwek and Joinson, 2016). Instant messaging can occur through desktop computers and mobile devices, with the latter being referred to as **mobile instant messaging**.

Whilst phishing can occur in either the desktop or mobile form of instant messaging, recent data suggests that most instant messaging phishing attacks target users through MIM apps (Kaspersky, 2021). These apps include WhatsApp, Viber and Telegram. Using mobile instant messaging (MIM) applications necessitates registration with a valid mobile phone number, which serves as the user's unique identifier. This enables users to initiate contact with others, including individuals they may not personally know. The only requirement is that they have the same application as those they are contacting.

A significant advantage of MIM platforms is their cost-effectiveness; these services are generally free, requiring only an active internet connection for message transmission. Consequently, users incur expenses solely for internet access rather than per-message charges. These applications facilitate global communication without geographic or quantitative limitations. However, this accessibility also presents security vulnerabilities, as cybercriminals can exploit the platform to disseminate phishing messages to a broad audience at minimal cost, unlike traditional SMS, which typically involves charges for each message sent.

The limit for MIM apps' messages depends on the application. For example, WhatsApp users can send up to 65,536 characters in a single message (Nieto, 2018), while a single Telegram message can contain up to 4,096 characters (Paganini, 2016), and a Viber message can contain up to 1,000 characters (Bulkgate, 2021). The ability to send longer messages in MIM apps means cybercriminals can craft more convincing messages than they would in SMS.

MIM apps provide functionality that allows multiple participants to be added to a dedicated sequence of conversations, forming what is known as a group or channel. Each member of a MIM group can send a message to other members, except in cases where the ability to send a message is limited to the group administrators. Communication in MIM groups is often referred to as group chats. A study by Seufert *et al.* (2015) has shown that users of MIM apps frequently engage in group chats. Groups in MIM apps are of two types: private and public groups.

Private groups typically have members who are familiar with one another. Private groups are often formed by friends or individuals who have regular interactions at work or school. Membership in private groups is by invitation only. While public groups are openly accessible, their invitation links are published online by their creators for anyone interested in joining (Garimella and Tyson, 2018). Users find public groups by searching on search engines. Members of public groups discuss topics as diverse as politics, technology, and sports (Wadhwa, 2018). Groups in MIM apps are similar to small communities that individuals form in real life because they offer their members a sense of collective presence (Dixon, 2018). They also offer their members numerous benefits. Although the benefits users derive from such groups tend to differ and often depend on the group. For example, the study by Cheung et al. (2017) found that such groups were useful platforms for peer support and information sharing, enabling individuals to quit smoking. Another study also found that individuals used such groups for selling, buying, comparing, and sharing information about products and services (Cao *et al.*, 2021).

The social group functionality of MIM apps, specifically the public group, has implications because members of such groups can view the contact details of other group members and privately message them. This feature enables cybercriminals to contact users individually and potentially socially engineer them into divulging personal information. Members of such groups can view and copy the contact numbers of others, providing those with malicious intent,

including phishers, the opportunity to harvest the contact information of others. These groups also provide an opportunity for cybercriminals to send phishing messages to multiple users simultaneously.

MIM apps enable users to click and share/forward links to each other, including links posted in public groups by strangers, which can potentially expose users to phishing attacks. Furthermore, the message-forwarding feature of MIM apps enables users to instantly share received messages with a large audience, which can be done through both one-on-one and group-based communication. This tendency can facilitate the quick dissemination of fraudulent messages throughout the network. Evidence from the literature has shown that cybercriminals exploit this feature to expand their reach by instructing users to forward fraudulent links to several others (Dassanayake, 2021). Therefore, it is crucial for users to consistently assess whether the individuals sending them messages can distinguish between deceptive and authentic links. Users prioritising security would likely assess and scrutinise links before sharing them. Ultimately, the likelihood that a user is exposed to phishing in such apps depends mainly on their behaviour towards links shared on such platforms. Moreover, the fact that users can join public groups means they can be exposed to fraudulent links shared by those with malicious intent. These issues necessitate the conduct of user behaviour studies that specifically focus on the behaviours of users of these apps towards links shared during communication in such apps. However, despite evidence of phishing moving to MIM apps, no study has investigated how frequently users engage in these behaviours. Understanding the frequency at which users engage in these behaviours may help us understand the extent to which users are susceptible to phishing in MIM apps and whether users of such apps and the research community should be concerned about the phishing problem.

Similar to SMS, MIM apps use a link preview mechanism to summarise the content of a shared webpage by displaying the shared page's title, description, logo/images, and domain name. Figure 2-2 shows an example of a benign link

preview on WhatsApp. However, findings from Stivala and Pellegrino (2020) showed that individuals with malicious intent can manipulate such previews to create a benign-looking link preview that redirects users to malicious links when clicked. The study found that MIM apps, including WhatsApp, Telegram, Viber, Messenger, and Snapchat, allow attackers to change a link preview's title, site description, Image/logo, domain and even the shared URL. These changes can be achieved by manipulating the HTML tags of the fraudulent web page or custom meta tags (i.e. Open Graph or Twitter Cards). One key concern is that cybercriminals can embed the logos of impersonated organisations in link previews by either inserting the image into the standard HTML tags of the phishing page or into the "og:image" meta tag within the HTML source code of the phishing site. This is concerning, considering evidence from a social marketing study that showed that the link preview increases the click-through rate of a URL (Cuddeford, 2018; McLachlan, 2021). This raises questions of whether link previews impact users' susceptibility to phishing in MIM apps by increasing their likelihood of clicking. It is currently unclear in the literature if this is the case.

Figure 2-2: Example link preview on WhatsApp

Prior research has demonstrated that communication within mobile instant messaging (MIM) applications tends to be predominantly informal and typically occurs between individuals who share an existing offline relationship (Quan-Haase and Young, 2010; O'Hara *et al.*, 2014; Piwek and Joinson, 2016). This

finding aligns with the foundational design principle of these platforms, which is to foster a sense of closeness and sustained connection among users and the individuals they consider most significant. However, evidence from the literature indicates that users assess the credibility of online resources using a social confirmation or social proof approach (Metzger, Flanagin and Medders, 2010), perceiving information as credible because others have deemed it credible. Given that users of MIM apps are likely to receive messages and links from individuals in their social circle, the question arises whether the user's relationship with the sender affects their susceptibility to phishing in MIM apps. This is crucial, given evidence from an online article showing that cybercriminals attempt to exploit the trust users have in their MIM contacts by asking them to forward phishing messages to them (Cuddeford, 2018).

Finally, although cybercriminals use MIM apps to propagate phishing campaigns, findings from Stivala and Pellegrino (2020) indicate that MIM apps often lack or have ineffective automatic technical countermeasures to protect users from phishing scams. The absence of automatic anti-phishing solutions in these apps may be attributed to end-to-end encryption, which prevents reading or analysing user messages. However, WhatsApp indicates that they perform on-device phishing detection to uphold the principle of end-to-end encryption. However, it is unclear how this was implemented. Furthermore, it is not apparent when WhatsApp first implemented these safeguards, as the findings of Stivala and Pellegrino (2020) indicate that most messaging applications, including WhatsApp, either lack automatic anti-phishing safeguards or have inadequate security measures.

Social media: The term "social media" refers to interactive platforms that utilise mobile and web-based technologies to enable individuals and communities to create, discuss, share, and modify user-generated content (Kietzmann *et al.*, 2011). There are numerous social media platforms available nowadays, and they vary in scope and functionality. Some social media platforms, such as Facebook, Twitter, and LinkedIn, allow users to connect with people they know and

strangers. Today, social media platforms are integral to the lives of individuals and organisations. People use these platforms for various purposes, including staying in touch with family and friends, making new connections, advertising goods and services, purchasing products, and learning new skills. Companies also utilise these platforms to introduce new products and maintain contact with their customers. Unfortunately, the popularity and rapid growth of these platforms have drawn the attention of cybercriminals, who use them for phishing attacks. Social media platforms, with their inherent properties that enable users to share, interact, converse, and form groups (Kietzmann *et al.*, 2011), provide a perfect medium for social engineers to deliver their lures and payloads to numerous users. Moreover, extending social media to smartphone applications may further expose users to a wider range of threats (Frauenstein and Flowerday, 2016).

Much like email phishing, phishers on social media also employ a range of social engineering techniques to attack users. URL shorteners, chain letters, hidden charges, impersonation, credential theft, and propagation are some of the most common social media scams (Elliot Volkman, 2019). The features of these platforms often facilitate these attacks. First, creating fake profiles on social media platforms is very easy. Cybercriminals can easily impersonate legitimate organisations and individuals to gain trust or share malicious content with users. This is evident in a recent article, where the police warned users about phishing scams originating from fake Facebook accounts that imitate the target's friends/family (ctinsider, 2025). Moreover, a phishing simulation experiment involving Facebook users has demonstrated that individuals are susceptible to social media phishing attacks initiated by fraudulent profiles. The study further revealed that participants often assess the authenticity of such profiles based on the number of friends associated with them (Vishwanath, 2017). Second, due to the limited text capacity on social platforms, users can utilise URL shorteners to convert long, complex URLs into shorter, more manageable links. However, this feature has been exploited by cybercriminals to obscure the final destination of hyperlinks (Chhabra *et al.*, 2011). While it is important to acknowledge that other

forms of media also permit URL shorteners, the character limitations imposed by social networking platforms make these tools a preferred method for sharing links in such contexts.

Third, social media platforms enable users to share/repost content, including links to external websites, with one another. Reposting/resharing content during social media communication creates what is known as viral social media communication (Borges-Tiago, Tiago and Cosme, 2019). Viral social media communication is a crucial aspect of social media marketing; however, it can also aid phishers in disguising malicious messages as sensational news, giveaways, or urgent alerts to a broad audience through shares, retweets, or reposts. Cybercriminals often compromise high-value accounts and use those accounts to share malicious content, aiming to reach the followers/friends of those accounts. These tactics can be seen widely used across the social media space (techradar, 2025).

Fourth, by default, user profile information such as addresses, phone numbers, and other sensitive data is often publicly accessible to other users, including individuals outside a user's direct network. This presents a significant security concern, as malicious actors can exploit this publicly available information to facilitate phishing attacks. However, it is worth noting that users can prevent this by ensuring the correct privacy settings are in place. However, a study conducted with social media users found that 43% of those surveyed didn't apply the existing privacy settings provided by social media (Ali *et al.*, 2018).

QR codes: Section 2.1.2 briefly explains QR codes and how they enable the embedding of Universal Resource Locators (URLs) that, when scanned by users' smartphone cameras, grant immediate access to web information (Sharevski *et al.*, 2022). This feature of QR codes is susceptible to exploitation by those with malicious intentions, allowing them to embed and distribute malicious URLs to unsuspecting users. Furthermore, URLs embedded in QR codes can't be visually examined, making it impossible for users to differentiate between legitimate and

phishing URLs before navigating to a site (Methods, 2024). The ease of reproducing and deploying QR codes in both physical and digital contexts renders them a preferable instrument for malicious actors seeking to mislead consumers and extract confidential information (Sharevski *et al.*, 2022). Furthermore, when users scan QR codes, the URL is visible briefly, typically for only a few seconds before establishing a connection, which complicates users' ability to verify whether the URL meets their expectations. Even if consumers can observe the URL before connecting, phishers may redirect them to harmful websites.

Voice: Voice communication, whether through Voice over Internet Protocol (VoIP) or traditional mobile phone networks, is a popular way for individuals and organisations to communicate. However, this medium has also attracted cybercriminals, who use it to conduct phishing attacks known as Vishing attacks. The definition of vishing is provided in section 2.1.4 of this thesis. However, for recall, vishing is when cybercriminals use IP-based voice messaging technologies, i.e., voice over Internet protocol (VoIP) or traditional mobile phone networks, to lure victims into providing their personal, financial, or other sensitive information. Features of voice communication technology that enable phishing include affordable mobile calling plans, which provide phishers an accessible and scalable means to execute extensive phishing calls (Gupta *et al.*, 2015).

Furthermore, spoofing caller ID is easy due to VoIP, allowing attackers to alter caller identifying details to mimic reputable companies such as financial institutions, governmental bodies, or service providers (Gupta *et al.*, 2015). This helps cybercriminals establish a false sense of legitimacy, increasing their chances of defrauding unsuspecting users. A recent attack on a healthcare provider demonstrates how cybercriminals utilised caller ID spoofing to impersonate internal IT staff, leading to an employee approving MFA that ultimately allowed the criminals to gain access to the victim's network (keepnetlabs, 2024b).

2.3.3 Persuasion techniques

Persuasion techniques are methods of communication employed by individuals to influence others into executing acts that the persuader desires. Persuasion techniques originate in marketing, where sales professionals use these methods to encourage consumers to purchase goods and services (Cialdini, 2006).

During phishing attacks, the primary goal of cybercriminals is to trick users into believing they are communicating with a legitimate entity. As such, cybercriminals spend much effort crafting phishing materials to make them sufficiently believable. Therefore, they employ the same persuasion approaches used in marketing to persuade victims to download a malicious attachment or click on a link, ultimately leading to malware being downloaded on their devices. Psychologists have investigated the reasons for and methods of persuasion. In this line of work, three taxonomies are often referenced: Cialdini (2006) six principles of persuasion, Gragg's psychological triggers (Gragg, 2003), and the principles of system security from the work of Stajano and Wilson (2011). Table 2-3 shows the elements of each of these principles. These principles tend to share some similarities, as found by Ferreira, Coventry, and Lenzini (2015), as illustrated in Figure 2-3.

Table 2-3: Cialdini's, Gragg and Stajano *et al.*'s principles

(Cialdini, 2006) six principles of persuasion	**(Gragg,2003) psychological triggers**	**(Stajano and Wilson, 2011)Principles of system security**
Reciprocation	Reciprocation	Kindness
Commitment and Consistency	Integrity and Consistency	Need and Greed
Social Proof	Diffusion and Social Responsibility	Social Compliance
Authority	Authority	Dishonesty

Liking	Deceptive relationship	Herd principle
Scarcity	Overloading	Time
	Strong Affect	

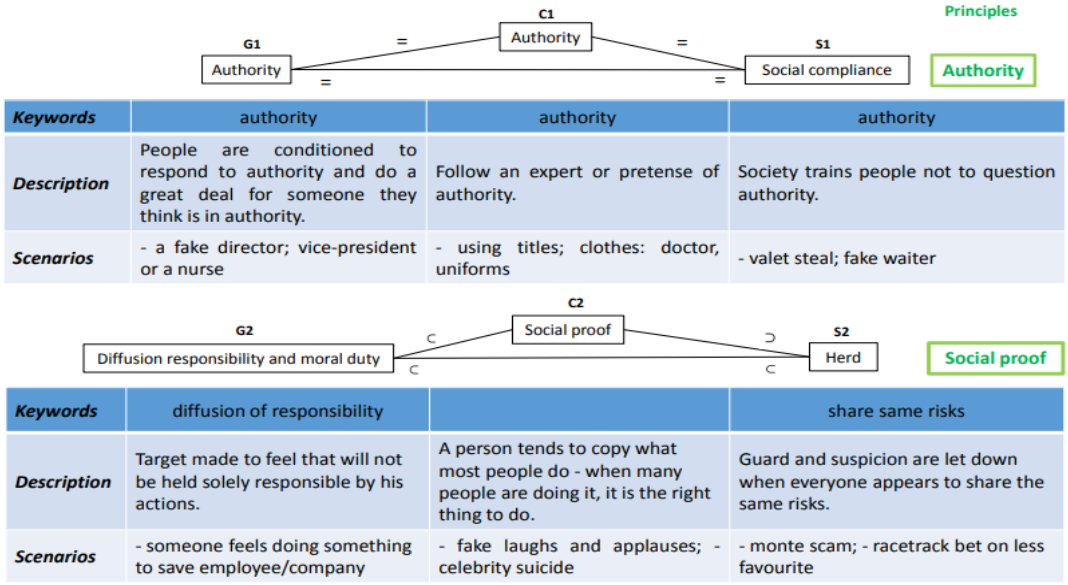


Figure 2-3: Similarities between Cialdini's, Gragg and Stajano *et al.*'s principles (Ferreira, Coventry and Lenzini, 2015)

The six principles of persuasion identified by Cialdini (2006) are currently the most popular reference point in the social engineering literature (Butavicius *et al.*, 2016). Additionally, these principles have been widely observed in phishing attacks, as cybercriminals use them to deceive internet users into falling for their tricks (Ferreira, Coventry and Lenzini, 2015).

Below is a brief description of Cialdini's six principles of persuasion.

Authority: The authority principle states that people may fall for social engineering scams due to fear of losing privileges and social condemnation. For

example, adversaries can impersonate an organisation's CEO and request the victim to transfer funds to an account.

Commitment and Consistency: This principle states that people may fall for persuasion tricks because they want to stay committed and consistent with their previous stand. Adversaries can exploit commitment and consistency by crafting phishing emails that reflect what people care about and are passionate about.

Liking: This principle is based on the notion that people believe and trust those they like or share similarities with. For example, a user may respond to a phishing email because it appears to be sent by a friend or an online/offline group member with whom the user is associated.

Scarcity: This principle is based on the idea that the value of items rises when they are in demand. Often seen in marketing, the scarcity principle can be exploited by luring internet users to click on a harmful link, often out of fear of missing out on promotions such as discounts on goods and services.

Social Proof: This principle leverages the fact that when individuals are afraid of making mistakes, they tend to conform to the actions of most others. Thus, people are more likely to click on malicious links if they are told several people have already clicked on those links and benefited from the scammer's request.

Reciprocity principle: This principle is based on humans' belief that they owe a debt to people who show them kindness or provide an enjoyable experience. An adversary can exploit this idea to persuade victims into installing malware on their computers in exchange for a gift or a free service.

Researchers have investigated how cybercriminals apply these principles during phishing attacks across communication media. For example, in Akbar (2014), the application of Cialdini's six principles of persuasion in phishing emails directed at individuals in the Netherlands was examined using a qualitative content analysis method. Utilising a deductive content analysis approach, the researcher

created a codebook using categories derived from prior research (Akdemir and Yenal, 2021). The codebook was then used to analyse a corpus of 207 unique English-language phishing emails from a Dutch database. The findings indicated that the principle of authority was the most frequently employed persuasive technique, followed by the principle of scarcity. In another qualitative content analysis study (Ferreira, Coventry and Lenzini, 2015), 52 English-language emails were analysed, obtained from personal email inboxes and publicly available Internet sources. The results revealed that the principle of liking was the most frequently employed, followed by the principles of scarcity and authority.

Zielinska *et al.* (2016) used qualitative content analysis to analyse 887 phishing emails targeting US universities between January 2010 and June 2015. Similar to Akbar (2014), this study employed a deductive content analysis approach, using a codebook with categories derived from prior research. This study found a temporal increase in the use of the principles of commitment and consistency, as well as the principle of scarcity. At the same time, the application of reciprocation and social proof showed a declining trend over the same period (Zielinska *et al.*, 2016). More recently, the researchers (Akdemir and Yenal, 2021) analysed persuasive principles employed in 208 coronavirus-themed phishing emails. These emails were collected in image format from three major search engines—Google, Bing, and Yandex —between April 1 and April 16, 2020. They were analysed using a deductive content analysis approach with a codebook that used categories derived from prior research. The findings indicated that the principle of authority was the most frequently utilised, followed by commitment/consistency and liking principles.

The results from the above studies have shown that phishers frequently apply the principles of authority and scarcity during email phishing, while other principles, such as social proof, are less commonly used in this context. Thus, it is likely that the context or media of communication determines which persuasion technique is used. This assertion is supported by a qualitative content analysis of vishing

attacks, which shows that the social proof principle was commonly used in Vishing attacks (Jones *et al.*, 2020). However, more studies are needed to determine if this is the case, as the literature currently lacks evidence on how cybercriminals apply persuasion techniques in Quishing, smishing, social media phishing, and mobile instant messaging phishing. A proposal was presented at the Information Systems and Neuroscience conference for an eye-tracking study to investigate the use of persuasion techniques in smishing attacks (Shamy, Xie and Zhong, 2024). However, the findings of this study have yet to be published. It is vital to investigate the use of these principles across different media, given that current evidence suggests that the type of persuasion principles phishers employ depends on the attack vector.

The way a particular medium of communication is used may also determine what persuasion principle the phisher uses. For instance, prior research has shown that communication in MIM apps is primarily informal and occurs between those with some offline connection(O'Hara *et al.*, 2014). Therefore, it is likely that MIM app users will receive phishing messages from acquaintances with whom they are familiar. These contacts could be those who have been compromised or those who lack knowledge of phishing scams and have already fallen for the attacks. This raises the question of the nature of MIM-based phishing scams. Thus, given the informal nature of communication in MIM apps, it is unclear whether phishers utilise persuasion concepts that exploit users' connection with the message's sender.

In addition to analysing how phishers apply persuasion principles during phishing, some studies have investigated the impact of Cialdini's principles on users' susceptibility to phishing scams. Wright *et al* (2014) first studied the impact of these principles on phishing susceptibility (Wright *et al.*, 2014). Since then, several studies have investigated these principles, with recent ones being (Lin *et al.*, 2019; Parsons *et al.*, 2019; Williams and Polage, 2019; Jones *et al.*, 2020; Akdemir and Yenal, 2021). Other studies, such as those by Lin *et al.* (2019), who examined the impact of internet user age, principles of influence, and life

domains on spear-phishing susceptibility using a simulated phishing study, found that the effectiveness of phishing scams varied by principles of influence and life domains, with age-group variability. More specifically, the study found that all age groups were most susceptible to emails containing scarcity and authority principles. Although younger adults were most susceptible to scarcity, older adults were most vulnerable to reciprocation. Users were also susceptible to emails about legal domains. However, the role-play study (Parsons *et al.*, 2019) showed the opposite, with emails containing the scarcity and social proof principles being the least successful, and those applying consistency and reciprocity principles being the most successful.

In a recent online email management study, Auton and Sturman (2025) investigated the extent to which persuasive principles, namely authority, scarcity, reciprocity, and social proof, influence users' ability to distinguish between phishing and legitimate emails, as well as whether this relationship is moderated by time pressure. Participants were asked to evaluate 60 emails, comprising 50 legitimate messages and 10 phishing emails. The stimuli were systematically manipulated to include both commonly employed persuasion principles (authority and scarcity) and less frequently used principles (reciprocity and social proof). Time pressure was experimentally varied by allocating either 7 or 15 seconds for participants to review each email. The results indicate that participants demonstrated the highest detection accuracy for phishing emails and unsafe links when messages incorporated scarcity and social proof cues. Moreover, phishing detection performance was significantly higher under the more restrictive time condition of 7 seconds. However, no significant interaction effect was observed between persuasion principles and time pressure.

2.3.4 Source credibility features

The existing literature has demonstrated that design cues or features related to the aesthetic presentation of emails or websites influence individuals' trust

perceptions of these platforms. For example, findings from Sillence *et al.* (2006) and Fimberg and Sousa (2020) show that design cues influence users' trust perceptions of websites. Specifically, these cues were instrumental in whether users continued to navigate to other parts of the sites.

Messages are more persuasive when they present themselves as credible (Kim and Kim, 2013). Cybercriminals employ various tactics to enhance the credibility of their messages, misleading their targets into believing they are communicating with legitimate, known entities such as banks and online retailers. Most phishing campaigns include crafted domain names that imitate the target organisations, official company logos, telephone numbers and email addresses of credible sources (Kim and Kim, 2013; Akdemir and Yenal, 2021). These features enhance believability and encourage Internet users to visit fraudulent pages (Williams and Polage, 2019). The work of Kim and Kim (2013) is one of the first studies to analyse the source credibility features of phishing emails. Applying a qualitative content analysis approach to 285 phishing emails, the researchers found that phishers used legitimate sender prefixes, legitimate domain names and company logos to enhance source credibility. Similarly, authors in Zielinska *et al.* (2016) analysed 887 phishing messages. They found that phishers used logos to increase credibility in only (n=44, 5%) of the messages.

In a recent eye-tracking experiment, Oh et al. (2025) investigated which features of phishing emails are most frequently overlooked by users during email interaction. The study involved twenty undergraduate participants who were exposed to phishing emails designed to mimic communications from Netflix and Adobe Creative Cloud. Each stimulus incorporated six commonly recognised deceptive cues: a spoofed sender address, generic salutation, urgent language, fraudulent hyperlink, brand impersonation, and false source attribution. The results indicate that brand impersonation, particularly the use of authentic-looking logos associated with well-known organisations, significantly impairs users' ability to distinguish phishing emails from legitimate communications.

Academic literature lacks in-depth studies on the most frequently used source credibility features in smishing messages. However, online articles suggest that smishing messages do not typically contain logos but tend to feature other credibility indicators, such as crafted domain names that imitate the target organisations (Buxton, 2024). This is in line with traditional text messaging platforms, where logos aren't supported unless the message contains a link and the corresponding web page contains the appropriate HTML tags that allow the SMS platform to generate a link preview with a logo. The process itself is not straightforward because the link must be placed at the end or beginning of a message, followed by a full stop, which must be placed at the end of the link. It must also contain either HTTP:// or HTTPS:// protocols and must be the only link in the message. In iOS, the message's sender must be a contact in the user's phonebook before such reviews are generated.

Like Smishing, academic literature lacks in-depth studies on the most frequently used source credibility features in social media and mobile instant messaging phishing. However, evidence from an online article shows that mobile instant messaging phishing messages can contain logos and spoofed domains of the mimicked organisations (Cuddeford, 2018). However, the online article was based on one sample attack. Therefore, further research is necessary to determine whether this technique is frequently used.

Due to the frequent use of source credibility features in phishing, researchers have conducted studies to understand their impact on user susceptibility to phishing scams. Findings from such studies show that when deciding the legitimacy of phishing webpages or emails, individuals make decisions based on the source credibility features, such as the logos of known brands and other design features (Dhamija, Tygar and Hearst, 2006; Furnell, Bryant and Phippen, 2007; Alsharnouby, Alaca and Chiasson, 2015; Jayatilaka, Arachchilage and Babar, 2021). Blythe, Petrie and Clark (2011) found that users were more susceptible to phishing emails containing a logo than those without one. The findings in Williams and Polage (2019) show a significant main effect between

participants' rating of the trustworthiness of emails and the presence of authentic design cues, including logos of known brands and copyright statements. However, a recent interview-based study of smartphone users shows that users didn't focus on visual cues when evaluating smishing messages (Tabassum, Faklaris and Lipford, 2024). However, in some instances, the use of excessive emojis, exclamation marks and currency symbols ignites suspicion.

The literature currently lacks evidence on the relationship between source credibility elements and susceptibility to phishing on either social media or mobile instant messaging, signalling a gap that needs to be addressed.

2.3.5 URL obfuscation techniques

Cybercriminals use URL obfuscation techniques to deceive victims into believing they are interacting with the correct URL. The results of the analysis of real-world phishing emails conducted by Lin *et al.* (2011) and Fernando and Arachchilage (2020) show that cybercriminals obfuscate phishing URLs to make them appear legitimate. They do this by 1) creating domain names that closely resemble genuine ones and inserting them either as a subdomain or within the URL's path component, 2) by displaying cryptic IP addresses as the domain rather than the actual domain, 3) by replacing one or more characters in the domain with characters that appear similar to those in actual domains; 4) by extending the length of the URL with random characters that make it difficult for users to read the URL; 5) by using HTTPs to give users a false sense of security; 6) using homoglyphs or Unicode characters similar to ASCII characters in the URL and 7) using tiny URLs to hide the actual URL.

Cybercriminals use the above techniques to either enhance source credibility by deceiving users into believing they are interacting with a legitimate entity or make URL parsing difficult for users. Indeed, evidence from the literature shows that including the name of a legitimate organisation as a subdomain in phishing URLs contributes to user susceptibility by fooling users into believing that the

URL redirects to a legitimate website (Alsharnouby, Alaca and Chiasson, 2015; Albakry, Vaniea and Wolters, 2020). Using HTTPs in phishing URLs is an intentionally deceptive approach, considering that online anti-phishing tips tend to advise users to look for HTTPS while transmitting critical information over the internet (Roberts *et al.*, 2019). However, the presence of HTTPs only signifies that a third party cannot intercept communication between two parties on the internet and doesn't signify the reliability and credibility of a webpage. However, evidence from a large-scale survey study shows that 30% of 1,888 participants who responded to the study believe the presence of HTTPS means a site is protected from phishing (von Zezschwitz, Chen and Stark, 2022). Albakry, Vaniea and Wolters (2020) show that it is difficult for users to predict the destination of complex URLs or phishing URLs when the URL length is extended with random characters. In a survey-based study, the authors (Thao *et al.*, 2020) found that using homoglyphs or Unicode characters similar to ASCII characters in phishing URLS makes it difficult for users to identify phishing URLS.

Previous paragraphs demonstrate how cybercriminals utilise the above techniques to obfuscate phishing URLs and how these deceptive techniques affect user susceptibility to phishing. However, it is currently unclear whether these techniques are uniformly used across the different communication media discussed in this thesis, or whether some are more pronounced in certain media or communication contexts than others. The limited display size of smartphones, which has already been acknowledged as an obstacle for users to accurately interpret a URL (Goel and Jain, 2018), may impact how phishers create deceptive URLs during phishing in smartphone applications such as SMS and MIM. For example, given the small screen size of smartphones, phishers can extend the length of the URL with random characters, making it difficult for users to read. Furthermore, the use of URL previews may influence how phishers create deceptive URLs during MIM phishing since, as highlighted in section 2.2.2, phishers can manipulate the key components of the link preview by including the details of the impersonated organisation in the title, site description, Image/logo,

and domain name. This approach can impact what URL obfuscation techniques the phisher uses during MIM phishing. As a result, further research is necessary to determine whether the URL obfuscation techniques used during MIM phishing differ from those used in other media.

2.3.6 Context of use

Susceptibility to phishing is a multifaceted phenomenon shaped by individual and contextual factors (Williams, Hinds and Joinson, 2018). Bradley and Dunlop, (2005) define context as "anything that influences the process in which focal user actions are undertaken". Context is widely recognised as a critical determinant of users' experiences, as it significantly influences how situations are experienced (Distler, 2023). Consequently, most user experience (UX) models emphasise context as a central factor in shaping user interactions and responses (Distler, 2023). One important aspect of context related to technology use is the technical context or IT context (Bradley and Dunlop, 2005; Lallemand and Koenig, 2020; Distler, 2023). Distler (2023) defines IT context as "the technologies used by participants as well as any technical issues". Thus, regarding phishing, the IT context encompasses the devices users use to access communication materials, which can influence how information is presented to users and their judgment and decision-making (Zhuo *et al.*, 2023). For example, how users respond to phishing on desktop computers may differ from how they respond to such messages on smartphones, primarily due to differences in the layout and interface design of email clients and web browsers. However, there are currently limited studies exploring the impact of IT context on phishing susceptibility. These studies include those by Loxdal *et al.* (2021), who investigated smartphone users to determine how they determine the legitimacy of a webpage. The results showed that smartphone users rely on webpages' design cues, content, and functionality, similar to desktop studies (Alsharnouby, Alaca and Chiasson, 2015). Interestingly, those who rely on URLs and other browser security indicators performed better than those who didn't. Unlike most desktop-based

studies, the results show no correlation between participants' susceptibility and age, technical proficiency, or time spent on a smartphone. This again highlights the contradictory results regarding the relationship between socio-demographic factors and phishing susceptibility.

Another study (Vishwanath, 2016) simulated a phishing attack on undergraduate students to test the effect of device type on the heuristic processing of phishing emails and whether device type influences email habits. Two types of phishing emails were used, one with only textual content and another containing a logo of the impersonated organisation and the sender's signature. This experiment showed that the type of device doesn't influence the heuristic processing of phishing emails. However, the study found that individuals with stronger email habits were more likely to utilise a smartphone than a PC.

While the above studies provide a starting point for discussing the impact of the type of device a user uses on susceptibility to phishing, further studies are still needed to back these results.

2.4 Anti-phishing software solutions

The knowledge derived from our understanding of the factors that make users susceptible to phishing and the tactics used by cybercriminals to phish users informs further efforts to curb the problem of phishing by supporting the development of automated anti-phishing solutions that detect and block phishing URLs with minimal or no user intervention. The concept behind automated solutions is that people can only fall for phishing attempts when the attacks reach them. Thus, ensuring that the attacks do not get to the user is the goal of an automated solution. Automatic phishing detection tools can be deployed at different levels, including mail servers, online browser tools, clients, and internet service providers (ISPs) (Alsharnouby, Alaca and Chiasson, 2015). These solutions are categorised into list-based, URL classification and analysis

techniques, and content-based (Alsharnouby, Alaca and Chiasson, 2015; Rao and Pais, 2019b).

2.4.1 List-based techniques

List-based anti-phishing solutions block URLs based on a list, either an allow or deny list. A deny list is a record of previously used malicious resources (e.g., URLs and IP addresses, spoofed websites) (Park, 2018). Techniques using a deny list check the legitimacy of unknown content by querying this list. PhishTank, a free website for users to submit, track, and verify phishing data, is among the best-known anti-phishing deny lists (Hong, 2012). PhishTank provides an API for developers and researchers to access and integrate phishing data into their applications, making it the favourite for many applications relying on a deny list (Hong, 2012). Other deny list databases include ATLAS from Arbour Networks, OpenPhish, Scumware.org, the MalwareURL list, and Artists Against 419 (ZELTSER, 2021). These lists are created by either users, as in the case of PhishTank, or software companies. Some web browsers maintain their own deny lists and heuristics, using these to alert users when they arrive at a known phishing page (Alsharnouby, Alaca and Chiasson, 2015). Most Browser security toolbars rely on a deny list, which they use to display security-related information about websites, helping users detect phishing and other security attacks (Wu, Miller and Garfinkel, 2006). In addition to showing security-related information, some security toolbars block users from accessing malicious websites. Some current security toolbars include Google Safe Browsing (Valdez, 2016), which is used for both the desktop and mobile versions of the Chrome browser. Other desktop-based browser toolbars include Netcraft Toolbar (Netcraft, 2020), McAfee WebAdvisor (McAfeeSiteAdvisor, 2023), Web of Trust (WOT) (WOT, 2020), SmartScreen Filter (Microsoft, 2023c).

An advantage of list-based techniques is that they produce a low false-positive rate, mainly because they compare websites with a database of known malicious

URLs. However, these techniques cannot detect newly created URLs that have not been added to the list. Typically, these lists obtain their malicious URLs and emails from various data sources, including emails detected by spam filters and verified phishing links reported by users (Sheng *et al.*, 2009). Frequent and timely updates of blacklist databases are crucial, considering that phishing websites are often short-lived and attackers frequently create new sites to lure users. In their analysis of phishing denylists, Bell and Komisarczuk (2020) found that Google Safe Browsing retains phishing URLs for a maximum of 60 days, whilst PhishTanks keeps them for a maximum of 65 days. This is a concern because some of the removed URLs may still be serving malicious content at the time of removal.

Sheng *et al.* (2009), in their experiment with eight anti-phishing toolbars, have shown that anti-phishing solutions based on denylist are ineffective in detecting zero-hour phishing attacks. These are attacks that are less than an hour old. Their findings show these techniques could not detect 20% of 191 fresh phishing attacks. They also found that the speed at which the different denylist methods are updated varies, with 47-83% of phishing URLs appearing in some techniques within 12 hours of the initial test. Security toolbars that combine denylist and heuristics (i.e. Internet Explorer 7 and Symantec Norton) caught more fresh attacks than the rest. Specifically, Symantec heuristics detected 70% of phishing at hour 0, and Internet Explorer 7 detected 41%.

Contrary to a denylist, an allowlist lists all legitimate websites that are frequently visited. This list can be compiled by the user or an organisation's system administrator. Allowlists work in conjunction with denylists, allowing users to visit only legitimate URLs that are part of the list. Approaches based on an allowlist are highly effective in detecting zero-day attacks and tend to yield zero false-positive results (Smadi, Aslam and Zhang, 2018), primarily because they rely on a list of known legitimate URLs for their decision. The main disadvantage of this approach is that users can only access the URLs in the list, which means other URLs, even if legitimate, can be blocked from being accessed by the user,

increasing false negatives and making this approach ineffective and less popular for phishing detection.

2.4.2 URL classification techniques

URL classification techniques utilise features of URLs, such as count-based, binary, and a known list of suspicious words, to classify URLs as legitimate or suspicious (Rao, Vaishnavi and Pais, 2020). Examples of these studies include Feng *et al.* (2018), which used the number of characters in URLs. Some studies used binary features, including IP addresses, suspicious words, special characters and brand names (Abutair, Belghith and AlAhmadi, 2019; Aung and Yamana, 2019), whilst others, such as (Basnet and Doleck, 2015; Alani and Tawfik, 2022) combine binary features and third-party features such as WHOIS, page rank, search results and DNS records to detect suspicious domains.

Most URL classification techniques employ tailored algorithms developed by their creators or machine learning algorithms to classify URLs as legitimate or phishing (Alsharnouby, Alaca and Chiasson, 2015). However, conventional machine learning techniques require considerable time for feature engineering and domain expertise to design feature extractors that can transform the raw data into a suitable representation for the classifier (Lecun, Bengio and Hinton, 2015). Researchers employing feature engineering during model development often rely on dimensionality reduction techniques, particularly principal component analysis (PCA) and minimum redundancy–maximum relevance (mRMR), to identify and select relevant features for their models (Kustiawan and Ghauth, 2025).

The effectiveness of machine learning models can also be limited by large datasets, which often lead to considerably longer training durations. However, the learning time can be reduced through a technique known as instance selection, which refines the training dataset by selecting a smaller yet representative subset, thereby significantly conserving computational resources.

In a recent phishing URL classification study, Abad, Gholamy and Aslani (2023) addressed the problem of computational inefficiency in machine learning model training by applying instance selection methods, including locality-sensitive hashing (DRLSH) and border point extraction based on locality-sensitive hashing (BPLSH). The results show that models trained on data selected using the BPLSH method generally achieved strong performance; however, a decline was observed for KNNs.

Researchers in the phishing detection domain have started adopting deep learning techniques in their anti-phishing solutions. Deep learning techniques are learning methods based on representation learning. They can automatically extract the suitable features needed for classification from input data, thus overcoming the limitations of traditional ML techniques in feature engineering. It has been emphasised by Kritika (2025) that the key factor for the adoption of deep learning models in phishing URL detection is their ability to automatically learn discriminative features from raw URL data, thereby eliminating the need for extensive feature engineering. This strength of deep learning has been demonstrated in recent work by Prabakaran, Meenakshi Sundaram, and Chandrasekar (2023), who proposed an enhanced deep learning–based phishing detection framework for the effective identification of malicious URLs. The authors employed variational autoencoders (VAEs) and deep neural networks (DNNs). They used the VAE to extract abstract high-level feature representations from raw URL data, while the DNN was used for URL classification. By utilising the VAE for feature extraction, the approach reduced the number of features required for classifier training, thereby lowering training-time overhead. Experimental results demonstrated that the proposed model achieved a maximum accuracy of 97.45% and a response time of 1.9 seconds, outperforming all other evaluated models.

Recognising the reliance on historical data and the time-consuming pre-processing nature of early machine learning models, Su and Su (2023) introduced a BERT-based (Bidirectional Encoder Representations from Transformers) deep

learning approach for detecting malicious URLs, utilising both URL strings and features as data input. The researchers used BERT for both tokenisation and classification tasks, leveraging the self-attention mechanism of the deep learning model to grasp the semantic meaning of input data. The approach was evaluated using three publicly available datasets from Kaggle, GitHub, and ISCX 2016. The model demonstrated strong performance across all experiments, outperforming existing approaches in terms of accuracy. The model achieved accuracies of 98.78% on the Kaggle dataset, 96.71% on the GitHub dataset, and up to 99.98% on the ISCX 2016 dataset for binary classification, with 99.78% for multiclass classification. Similarly, (Meléndez, Ptaszynski and Masui, 2024) in their paper show that the deep learning model roBERTa (Robustly optimised BERT approach) achieved a maximum accuracy of 0.9943, outperforming traditional models such as Support Vector Machine SVM, which scored 0.9876 in an email detection task.

2.4.3 Content-based techniques

Content-based techniques extract features from the source code of webpages, such as those from hyperlinks, images, and tags, to classify webpages as suspicious (Medvet, Kirda and Kruegel, 2008; Rao, Vaishnavi and Pais, 2020). They also utilise network-related information, such as the payload of various communication protocols, including HTTP and SMTP (Khonji, Iraqi and Jones, 2013). Similar to URL-based techniques, these methods utilise tailored algorithms or machine learning approaches to classify phishing materials. Typically, these tools construct a search string from the extracted text to query third-party servers or search engines, obtaining more information on the suspected site. This approach is illustrated in Dunlop, Groat and Shelly (2010), where the authors extracted logos and brand names from suspected phishing web pages and used them to query the search engine. They then compare the top-level and second-level domains of the first four results returned by Google with the top-level and second-level domains of the URL a user enters. If a match is

found, the site the user enters is considered legitimate; otherwise, the site is classified as suspicious. However, relying on external third-party services, such as search engines, WHOIS, and page ranking, makes these approaches ineffective because their accuracy depends on the content of these services (Shirazi, Bezawada and Ray, 2018). In addition, relying on third-party services may compromise user privacy, as user web surfing history can be recorded if proper mitigating techniques are not in place (Shirazi, Bezawada and Ray, 2018). Further checks also reveal that approaches relying on external third-party services often fail to detect phishing sites hosted on compromised servers and sometimes incorrectly classify a newly created webpage as phishing (Rao and Pais, 2019a). Responding to the inefficiency of approaches relying on third-party services, other content-based methods have been proposed that detect phishing links and pages based on the ratio of foreign links and broken links. For example, Shirazi, Bezawada, and Ray (2018) and Rao and Pais (2019a) used the percentage of links on a page that refer to the same domain displayed in the browser relative to all the links on the page. The idea is based on the intuition that phishers will likely include hyperlinks on the phishing landing page, leading to legitimate pages to reduce user suspicion.

Some approaches have combined features from content, URLs, and third-party tools, such as WHOIS, to detect phishing (Zhang, Hong and Cranor, 2007; Rao and Pais, 2019b).

2.4.4 Effectiveness of software-based techniques

Software-based anti-phishing solutions aim to prevent users from falling victim to phishing by detecting and blocking phishing emails and URLs before they reach the user. Although these approaches have been highly effective in reducing the number of phishing attacks that reach users, the fact that they are not foolproof means that users sometimes receive phishing emails. Some studies have evaluated the effectiveness of these tools (Wu, Miller and Garfinkel, 2006; Sheng

et al., 2009). They concluded that these tools could provide some help. However, they should not be relied on to protect users from all phishing scams. Cybersecurity professionals have suggested complementing software-based anti-phishing solutions with user awareness/training to equip users with the necessary skills to detect phishing that bypasses automated solutions.

2.5 Phishing awareness/training

Phishers target human vulnerabilities by exploiting their lack of knowledge and understanding of phishing attacks. An apparent solution, therefore, is to educate and train users to recognise these attacks (Khonji, Iraqi and Jones, 2013). User awareness/training approaches target users, aiming to increase their awareness of phishing attacks by providing them with online and offline risk information regarding phishing threats. The primary objective of user training is to educate users about phishing and equip them with the necessary knowledge to prevent falling victim to such attacks (Kumaraguru, 2009).

At the basic level, organisations and government agencies use online articles and tips to enlighten users about phishing and help them detect such attacks themselves. For instance, the Federal Trade Commission (FTC, 2019) and the Anti-Phishing Working Group frequently publish tips and articles on phishing, as well as information on how users can avoid such attacks, including what to look for when interacting with phishing content. An example of such tips for MIM apps can be found in Kaspersky (2021), where MIM app users were advised to look for misspellings or other irregularities in links. Other recommendations that cybersecurity experts often provide to internet users regarding how to respond to phishing attempts include not clicking links from unknown senders, searching Google for further information on the website before giving out their details (Orunsolu et al., 2017).

Other approaches include classroom-based approaches that use classroom sessions with content that introduces participants to phishing campaigns and how to detect them (Robila and Ragucci, 2006; Tschakert and Ngamsuriyaroj, 2019); embedded training, where training materials are embedded in the user's daily work activities (Alnajim and Munro, 2009; Tschakert and Ngamsuriyaroj, 2019), and a game-based approach that uses games to educate users about phishing (Kumaraguru *et al.*, 2010; Arachchilage and Cole, 2011). Recently, approaches that train users based on context have been proposed. Contextual training involves presenting users with concise information in a setting where it is immediately applicable (i.e., when they open their emails or hover over a link) (Volkamer, Renaud and Reinheimer, 2016; Kävrestad *et al.*, 2022). Compared to alternative training methods, contextual and game-based training tend to be more engaging and provide better results (Kävrestad *et al.*, 2022).

Before conducting phishing awareness training, organisations conduct a phishing simulation exercise. A simulated phishing exercise involves an organisation sending realistic phishing emails to its staff to assess their phishing awareness and examine how they respond to phishing emails when they receive them (Mimecast, 2025). Consequently, the organisation directs instructional resources towards those who mistakenly click or input their credentials (Gordon *et al.*, 2019). Phishing simulations can help organisations identify weaknesses and risks in their human defence layer (Rizzoni *et al.*, 2022). Initially, these simulation exercises were focused on the email vector; however, more recently, companies have built platforms that allow organisations to test their staff awareness across different media, including email, SMS, voice, and QR (keepnetlabs, 2024a). However, as (Schiller *et al.*, 2024) put it, "Phishing simulations are quite controversial in the scientific literature". This is because some researchers, such as Volkamer, Sasse and Boehm (2020), believe that attacking staff in the name of phishing simulation can have a negative impact on employees' self-efficacy and reduce inter-personal trust and trust in the organisation, consequently impacting organisational goals. A clear example is when, in 2020, a phishing simulation at a

US company caused outrage and anger among employees (Picchi, 2020). According to the article, one of the simulated emails promised a $10,000 bonus to staff at a time when they were experiencing the impact of a pay cut. However, despite the controversial nature of phishing simulations, recent mixed-study research indicates that phishing simulations have a positive effect on employees' knowledge and situational awareness, making them more confident in their ability to detect phishing and feeling more protected by their company (Schiller *et al.*, 2024). In addition, contrary to Volkamer, Sasse and Boehm (2020), the findings in Schiller *et al.* (2024) show that employees have a positive attitude towards phishing simulations and don't feel attacked by their organisation.

2.5.1 Effectiveness of user awareness/training

While user awareness/training has been widely discussed in the literature and used in practice, several studies suggest that these methods fail to encourage users to perform the required secure behaviour (Hatfield, 2018; Kävrestad *et al.*, 2022), leading to many cyber incidents (Wetzig, 2022). Some of the reasons why these methods don't always have the desired effect include 1) deterioration of the acquired knowledge (Reinheimer *et al.*, 2020), 2) training materials are too long, resulting in avoidance or being left until the last minute (Gjertsen *et al.*, 2017), and 3) training tend to focus more on knowledge and awareness and less on embedding positive cybersecurity behaviour (Bada, Sasse and Nurse, 2019). Furthermore, the complexity of phishing attacks and their evolving nature mean that users must be frequently trained on new forms of attacks, which can be both time-consuming and costly for many organisations (Smadi, Aslam and Zhang, 2018).

2.6 Influence of Large Language Models (LLMs) and Generative AI on Phishing

The advanced capabilities of large language models (LLMs) have made them invaluable in various applications, including content creation. The ability of large language models to generate responses that closely emulate human-like conversation across an increasingly diverse array of subjects has improved the field of natural language processing, allowing individuals to effortlessly generate content that mimics real world human conversations. However, such capabilities have also caught the attention of those with malicious intent, exploiting them to conduct phishing attacks (SecureOps, 2023). Interaction with LLMs is usually via natural language prompts where users ask the models questions or instruct them to carry out some tasks. This technique of crafting queries to interact with the LLM is called prompt engineering (Shenoy and Mbaziira, 2024). Whilst LLMs have safeguards to identify and reject potentially harmful or misleading prompts, malicious actors can still bypass these measures (Roy *et al*., 2024). This has led to the generation of malicious content, including phishing emails (SecureOps, 2023). FraudGPT and WormGP are malicious LLM tools that provide attackers with the capability to create convincing phishing emails for business-compromise attacks (SecureOps 2023). They can also create phishing pages and websites, write malicious code, find leaks and vulnerabilities in networks or Apps. This relies on the use of effective prompt engineering to formulate and launch attacks. Afane *et al.* (2024) highlight the potential of LLM-generated phishing content to bypass conventional phishing detection tools, including Gmail's spam filter, Apache SpamAssassin, and Proofpoint. Their findings indicate that traditional phishing detectors such as SpamAssassin and Proofpoint perform effectively on original phishing emails; however, their accuracy and recall decline significantly when evaluating LLM-rephrased versions of the same emails.

Nonetheless, the ability of LLM-generated phishing attacks to evade detection means that current anti-phishing solutions must match the scale of the problem.

i.e., the defensive strategies must also apply a similar capability. As a result, there have been recent research efforts in this direction. For example, Onwuegbuche *et al*. (2025) evaluated the effectiveness of large language models for identifying affective states and emotional cues deliberately exploited by cybercriminals during phishing attacks, using a dataset comprising more than 5,000 phishing emails directed at six universities. The results indicate that LLMs, such as GPT4, Llama 3 and GeminiPro, can detect affective states, including curiosity, trust, and urgency in phishing emails. The researchers argue that the ability of LLMs to detect affective states in phishing emails means such models could be used to alert users to exercise extra caution when engaging with such emails. In another research, Desolda, Greco, and Viganò (2025) proposed APOLLO, an LLM-based phishing detection tool built on OpenAI's GPT-4o, designed to both identify phishing emails and generate explanatory messages that inform users about the associated risks. The study reported that GPT-4o achieved a phishing detection accuracy of 97%, which was further enhanced through the integration of data from third-party services, yielding a near-perfect classification rate of 99%. Moreover, user evaluations indicated that the explanations generated by the model were perceived as more understandable, engaging, and of higher quality compared to manually crafted explanations. Overall, the findings demonstrate the potential of LLMs to function simultaneously as effective technological safeguards and as human-centred defensive mechanisms against phishing attacks.

Overall, the studies reviewed demonstrate that LLMs and generative AI are reshaping attacker methodologies while simultaneously offering significant potential to enhance defensive mechanisms against phishing attacks.

2.7 Conclusions

Since the 1990s, phishing has remained a significant cybersecurity issue, prompting extensive research to deepen the understanding of its mechanisms and develop effective countermeasures to safeguard individuals and

organisations. Section 2.1.2 highlights how phishing has evolved in response to technological advances, revealing how cybercriminals have exploited vulnerabilities in new communication media to propagate their campaigns. Thus, the fight against phishing attacks is ongoing, as new forms of phishing attacks continually require fresh efforts from the research community.

The literature review highlights that MIM phishing represents an emerging threat that deserves attention. Its unique contextual characteristics justify further investigation into how phishing tactics and user behaviours are shaped and influenced within this environment, as well as the interaction between them. Such research is crucial for informing the development of effective technological solutions and educational strategies to mitigate this growing risk. Although some research work has already begun in this area, it has been limited in scope, leaving many questions unanswered. For example, whilst the current literature has investigated the impact of knowledge and phishing self-efficacy on users' susceptibility to mobile instant messaging phishing, It is essential to investigate the problem from both the phisher and user perspectives to gain a comprehensive understanding of how the communication medium influences phishing tactics and user behaviours, as well as how both parties adapt their strategies in response to the specific affordances and constraints of the medium. Moreover, early evidence has revealed that cybercriminals rely on their victims to reach a vast number of targets by asking them to forward phishing scams to their contacts (Cuddeford, 2018), raising an essential question of the impact the relationship users have with the sender of a message on their susceptibility to phishing in MIM apps. Furthermore, the literature has shown that the link preview, which MIM apps use to render URLs, can be manipulated to redirect users to fraudulent pages. However, the impact of link previews on users' susceptibility to phishing in mobile instant messaging (MIM) apps is yet to be established.

This thesis aims to contribute to the literature by examining the phishing problem in MIM apps from both the phisher's and user's perspectives using self-

reported data from participants about their behaviours and intentions, and secondary data to explore and analyse the phishing strategies used during mobile instant messaging phishing attacks. This approach will provide a comprehensive understanding of how the communication medium influences phishing tactics and user behaviour.

Chapter 3

Methodology

3.1 Introduction

This chapter outlines the methodological approaches employed in this thesis. The focus of this research is on MIM phishing. Thus, data collection was centred around users of MIM apps. The thesis aims to investigate the phishing problem in MIM apps, examining it from both the phisher's and user's perspectives to provide a comprehensive understanding of how the communication medium influences phishing tactics and user behaviour. As noted in section 2.2.2, MIM apps lack effective anti-phishing countermeasures to protect users from phishing attacks. Consequently, users must serve as the first and last line of defence. Thus, only security-aware users can safeguard themselves from these attacks by remaining vigilant and adhering to secure behaviours whilst communicating in these apps. However, section 2.2.2 also highlighted that some features of MIM can increase user susceptibility to phishing in MIM apps. Thus, the behaviours of users regarding these features can serve as an important determinant of user susceptibility to phishing in such apps. However, it is currently unknown how the users of such apps behave concerning phishing. These issues necessitate the conduct of user behaviour studies that specifically focus on the behaviours of users during communication in such apps. As a result, the first question of this thesis is:

RQ1: Do users of MIM apps exhibit behaviours that could expose them to the risk of falling victim to a phishing attack?

Whilst **RQ1** aims to understand how users behave during communication in MIM, specifically focusing on their use of the features of MIM apps and whether such behaviours expose them to the risk of falling victim to a phishing attack, another key aspect to understanding phishing in MIM apps is understanding how cybercriminals use these features to target unsuspecting users. Section 2.2.3 has demonstrated that cybercriminals employ persuasion techniques to persuade targets into clicking on phishing links. Additionally, Section 2.2.3 shows that perpetrators of phishing employ different persuasion principles depending on the attack channel, with differences found across various communication media. However, it is currently unclear whether the features offered by MIM apps have any influence on persuasion principles that cybercriminals employ during MIM phishing. Phishers may employ distinct methods when carrying out MIM phishing, as opposed to other attack channels, due to the medium's unique features, which are highlighted in Section 2.2.2. Furthermore, section 2.2.5 has highlighted that phishers use URL obfuscation techniques during phishing attacks. However, the limited screen size of mobile devices and the use of link previews may influence how phishers construct deceptive URLs during MIM phishing. Therefore, the second research question of this thesis is:

RQ2: What are the tactics and techniques used in the content of phishing messages targeting the users of MIM apps?

Section 2.2.2 highlighted the unique features of MIM apps that can potentially affect susceptibility to phishing. However, it is currently unclear in the literature whether these features actually affect user susceptibility to phishing in MIM apps. Therefore, a more extensive investigation is required to assess how these features influence users' responses to actual phishing messages that simulate real-world MIM phishing attacks. Such a study can be utilised to examine further whether the unique features of MIM influence susceptibility to phishing in such apps**.** As a result, the third question of this thesis is:

RQ3: Do specific characteristics of MIM apps influence user susceptibility to phishing within these platforms?

3.2 Research Philosophy

Research is guided by a set of beliefs or a worldview. These beliefs or worldviews are referred to as a paradigm (Killam, 2013). A paradigm refers to a particular way of understanding or interpreting the world. According to Guba (1990), a research paradigm consists of core beliefs that guide actions and shape the philosophical assumptions about how knowledge is generated. In research, paradigms serve as foundational frameworks that guide researchers in all aspects of their work (Killam, 2013). Guba and Lincoln ((1994) define paradigms as "basic belief systems based on ontological, epistemological and methodological assumptions". Axiology is also regarded as a key component of a research paradigm (Klakegg and Tvedt, 2024). Axiology refers to what the researcher believes is valuable and ethical. These beliefs guide the researcher's decision. Ontology refers to the researcher's assumptions about what constitutes reality and focuses on the fundamental nature or essence of the social phenomenon under investigation (Scotland, 2012). Ontology is concerned with the "what is" and implores researchers to adopt a position regarding their perception of how things really are and how things work (Scotland, 2012). An example of an ontological question is: Does the researcher view reality as a single, context-independent entity that can be objectively discovered, or as multiple, context-dependent mental constructions? Epistemology explores the relationship between the researcher and the knowledge being investigated. It therefore refers to how we come to know what we know. An individual's ontological beliefs determine the degree of objectivity in the relationship between the researcher and the knowledge being studied. In other words, a researcher's ontological assumptions inform their epistemological assumptions. An example epistemology question is: Does the researcher believe that the relationship with participants should be objective or subjective? (Killam, 2013). Finally,

methodology refers to the approach researchers use to investigate and generate knowledge, and is concerned with why, what, from where, when, and how data is collected and analysed (Scotland, 2012). According to Guba and Lincoln (1994), the ontological, epistemological, and methodological assumptions within research paradigms are interconnected, so that addressing one question influences how the others are approached. For instance, if a researcher believes that reality can be measured, they are likely to seek an objective relationship with the subjects. Consequently, the chosen methodology would typically be experimental in nature. Although the literature discusses numerous research paradigms, there is no consensus on a definitive number or classification system (Ugwu, Ekere and Onoh, 2021). However, some popular research paradigms discussed in the literature include positivism and post-positivism (Maksimović and Evtimov, 2023).

The positivism paradigm is based on realism. Realism is the belief that reality exists and is driven by natural laws (Killam, 2013). Positivism believes there is one reality that can be objectively measured and determined (Maksimović and Evtimov, 2023). Della Porta and Keating, (2008) stated the positivism paradigm believes that " the world exists as an objective entity, outside of the mind of the observer, and in principle it is knowable in its entirety. The task of the researcher is to describe and analyse this reality. Positivist approaches share the assumption that, in both natural and social sciences, the researcher can be separated from the object of his/her research and therefore observe it neutrally and without affecting the observed object". The axiological assumption of the positivism paradigm advocates for honesty and integrity of the researcher. However, due to the belief that reality can be objectively measured, the ontological assumption of the positivism paradigm is based on realism, whilst its epistemological assumption is that the researcher and the researched are independent entities and have no influence on each other. The methodology of the positivism paradigm is often experimental or manipulative in nature, aiming to test and verify hypotheses through quantitative approaches. These assumptions allow the

positivism paradigm to be grounded in the principle of verification, where its proponents believe that knowledge can be distinguished from personal opinion only when there is a method for empirically verifying the truth of a statement (Maksimović and Evtimov, 2023). Consequently, only knowledge that has been scientifically tested, verified, and supported by observable evidence is considered valid. Positivism's emphasis on objectivity has been criticised in the social sciences, as many social and human phenomena are complex, context-dependent, and not easily measurable. As a result, the rigid objectivity associated with positivism is often viewed as unsuitable for social and humanistic research (Maksimović and Evtimov, 2023).

The post-positivism paradigm emerged in response to the limitations of positivism. According to (Scotland, 2012), "post-positivism has similar ontological and epistemological beliefs as positivism; however, it differs in several ways". For example, post-positivism believes that "knowledge is not neutral and that all knowledge is socially constructed" (Henderson, 2011). Post-positivists recognise that the world cannot be viewed in a fully objective way and that subjectivity shapes our understanding of reality (Maksimović and Evtimov, 2023). In essence, whilst the axiological assumption of the post-positivism advocates for the conduct of quality research, through the principles of beneficence, respect and justice, the ontological assumption of post-positivism is founded on critical realist ontology, where it is believed that reality exists, but it cannot be perfectly detected due to limitations in our methods of inquiry and the complex nature of phenomena (Guba and Lincoln , 1994). The post-positivists maintained a similar epistemological assumption to the positivists, valuing objectivity. However, unlike positivism, post-positivists argue that total separation between the researcher and the researched is not possible (Killam, 2013). Post-positivists are aware of the potential impact of researchers' background knowledge on observation and make efforts to minimise its effects (Killam, 2013). The methodology of the post-positivism paradigm involves a combination of quantitative and qualitative methods. Researchers adopt the

post-positivist paradigm because it allows the integration of multiple methodologies, enabling each approach to offset individual limitations while drawing on their respective strengths (Kankam, 2019).

The present study adopts a post-positivist paradigm, drawing on its axiological, ontological, epistemological, and methodological assumptions to address the research questions of this thesis. This paradigm enables the use of both quantitative and qualitative methods to examine phishing in MIM applications from the perspectives of both phishers and users. Employing a mixed-methods approach enhances the value of the research findings, as qualitative insights can be used to explain and enrich the quantitative results. However, the researcher acknowledges that their prior experiences, beliefs, assumptions, education, and expectations can influence how data is observed, interpreted, and understood. As a result, rigorous methods such as inter-coder reliability checks, pilot testing of the survey and appropriate statistical techniques are used to minimise these biases.

3.3 Research Methods

Research methods can be classified into three main categories: qualitative, quantitative, and mixed methods. Each methodology presents a distinct perspective for assessing and explaining the study's outcomes and discoveries. The qualitative research approach utilises textual data to conduct a comprehensive investigation and provide a thorough depiction of a phenomenon (Bryman, 2016). It is commonly employed when there is a scarcity of information about the issue being investigated (Taherdoost, 2022). It enables researchers to examine participants' experiences, interpretations, and perspectives, typically from their standpoint (Hammarberg, Kirkman,. Qualitative research is inherently exploratory. Consequently, it aids researchers in uncovering novel ideas and formulating fresh theories (Taherdoost, 2022).

Researchers employ qualitative research methods when questions cannot be adequately answered by quantitative research methods (Loraine, Wick and Christoph, 2020). For instance, questions that necessitate uncovering the cause of an observable pattern require a qualitative approach (Loraine, Wick and Christoph, 2020). For example, qualitative research methods can be used to understand users' decision-making processes when encountering phishing.

One of the most popular approaches to conducting qualitative research is qualitative descriptive research (Merriam, 2002). Qualitative descriptive research is a type of research in which researchers investigate a phenomenon in its natural state, primarily from a naturalistic viewpoint (Sandelowski, 2000). As the focus of qualitative descriptive research is on learning about certain events as they are, researchers collect data using techniques such as observations and analysis of recordings, reports, images, and documents (Lambert and Lambert, 2012). This form of research is recommended when a clear understanding of a phenomenon is required (Lambert and Lambert, 2012). Two notable qualitative descriptive research approaches are qualitative content analysis (QCA) and thematic analysis (Vaismoradi, Turunen and Bondas, 2013). While QCA focuses on describing the characteristics of the document's content, the thematic analysis focuses on analysing, identifying, and reporting patterns or themes within the data (Vaismoradi, Turunen and Bondas, 2013). QCA is an analytical procedure where researchers systematically categorise textual or visual data into categories, patterns, and themes for description and interpretation (Akdemir and Yenal, 2021).

Furthermore, QCA is a highly adaptable methodology applicable to diverse research designs (Kyngäs, 2020). For example, researchers can analyse data inductively or deductively depending on the research question. The inductive approach is applied when the study design aims to describe the outcome of a phenomenon. The method is suited for conditions where existing theory or research on a phenomenon is lacking (Hsieh and Shannon, 2005). In inductive content analysis, categories and names are derived from the data, thus

preventing researchers from using preconceived categories (Kondracki, Wellman and Amundson, 2002). Deductive content analysis, on the other hand, requires researchers to use existing knowledge or prior research to establish the categories before analysis (Stemler, 2000). This approach is based on the fact that existing theories or previous research may exist about a phenomenon that can be extended (Hsieh and Shannon, 2005).

Whilst QCA has numerous advantages, researchers have pointed out that the reliability and trustworthiness of the process are sometimes impacted by the lack of standard procedures that can be applied throughout the analysis process (Elo *et al.*, 2014). As a result, several scholars have proposed standard procedures to enhance the trustworthiness of this method (Elo *et al.*, 2014; McKibben *et al.*, 2020). An example of these procedures is interrater reliability, which measures the extent to which individual coders agree that a particular aspect of a text or data belongs to a specific category (McKibben *et al.*, 2020).

Despite the numerous benefits of qualitative research, a key disadvantage is that it is time-consuming, requires skilled personnel, and the analysis can be labour-intensive (Mohajan, 2018). Furthermore, when there is a need for social interaction between the researcher and the respondents, such as during an interview, respondents can conceal their intentions, facts, and important information, and even invent fictitious events (Kelle, 2006). Additionally, respondents may be concerned about their privacy, especially when the research involves sensitive topics.

Quantitative research methods, on the other hand, are appropriate for situations in which 'factual' data is required to answer the research questions (Hammarberg, Kirkman and De Lacey, 2016). This research method utilises numerical values derived from observations to explain and describe a specific phenomenon (Taherdoost, 2022). Quantitative research methods allow researchers to answer questions such as how many and what percentage. For instance, they can provide insights into the percentage of the sample that

accessed a phishing page and their demographic distribution based on age, marital status, and educational qualification. Additionally, researchers can use these methods to track any changes between two consecutive surveys (Kovacs *et al.*, 2012). For instance, researchers can track changes in user susceptibility to phishing before and after security awareness training by comparing statistics of individuals who access a phishing page before and after the training.

Quantitative research methods are suitable when researchers can readily identify the variables, separate them and link them to formulate hypotheses before data collection. Thus, these methods are ideal for research with clearly defined questions (Hammarberg, Kirkman and De Lacey, 2016). Examples of quantitative research methods include surveys, descriptive research, correlation studies, and experimental research (Taherdoost, 2022). Quantitative research enables researchers to collect data from larger samples more efficiently, resulting in greater accuracy when generalising research findings. Furthermore, the fact that quantitative surveys often allow participants some days to complete them and also require no close observation of the participants means participants have sufficient time and privacy to complete their responses. However, quantitative research methods can suffer from methodological problems when researchers have limited or lack knowledge of the sociocultural background of the research subjects (Kelle, 2006). This issue can lead to difficulties in theory building and hypothesis construction, resulting in the incorrect specification of statistical models (Kelle, 2006). Furthermore, a lack of knowledge about the investigated social world means researchers may misinterpret respondents' answers in questionnaires and can ask the wrong questions (Kelle, 2006).

Sometimes, neither of the methods above can sufficiently address research enquiries. Consequently, researchers frequently integrate qualitative and quantitative methodologies, creating what is known as a mixed-methods approach. These methods are used to improve understanding of a subject (Taherdoost, 2022). The mixed method is “research in which the investigator collects and analyses data, integrates the findings, and draws inferences using

both qualitative and quantitative approaches or methods in a single study or a program of inquiry" (Tashakkori and Creswell, 2007). The idea is to combine the strengths of qualitative and quantitative methods to effectively address the constraints and resolve issues associated with using either approach (Kelle, 2006). For example, researchers can use a quantitative approach to identify problem areas and research questions, which they can further investigate using qualitative methods. Furthermore, findings from the quantitative part of a mixed-method method can help researchers corroborate those of a qualitative study and transfer them to other domains (Kelle, 2006).

Due to the nature of the research questions in this thesis, a mixed-methods research approach utilising both quantitative and qualitative methods was employed. Using a mixed-methods approach can triangulate the findings of this research, providing more insights by combining the strengths of quantitative and qualitative data (Creswell and Clark, 2007). This aligns well with RQ1 and RQ2, as these questions aim to examine different issues. Another justification for adopting a mixed-methods design is that it enables researchers to utilise both qualitative and quantitative data to address questions within a study that is largely quantitative or qualitative. This approach is also known as embedded mixed methods design (Yu and Khazanchi, 2017) and is suitable for answering RQ3 as it focuses on examining the characteristics of MIM apps that affect susceptibility to phishing. Mixed-methods research is beneficial when researchers plan to use qualitative data findings to explain or elaborate on quantitative results (Creswell and Clark, 2007). The following section will outline how the research in this thesis was conducted.

3.4 Research Design

A research design is a plan for a piece of research and consists of a framework for efficiently conducting research. It contains the tools and procedures the researcher will use to collect and analyse the research data (Punch, 2013). According to Punch (2013), the research design should include all research

procedures, from defining the research problem to presenting the findings. A good research design should specify the type of research design adopted, the data collection approach, and the data analysis method (Sreejesh *et al.*, 2014). Regarding the type of research design, the three most popular ones are exploratory, qualitative descriptive and causal/experimental (Sreejesh *et al.*, 2014). Exploratory research is conducted to gain a deeper comprehension of a topic or evaluate a phenomenon (Swedberg, 2020). Qualitative descriptive research is a type of research in which researchers investigate a phenomenon in its natural state, primarily from a naturalistic viewpoint (Sandelowski, 2000). Causal/experimental seeks to identify the relationship between variables and whether a change in one variable leads to a change in another (Sreejesh *et al.*, 2014).

The research in this thesis was conducted in several stages, utilising various methodological approaches carefully selected to address the research questions. Therefore, three studies were conducted to address the research questions of the thesis. The first was exploratory and aimed to investigate the behaviours of MIM app users towards shared links and whether these behaviours expose them to security risks, such as phishing attacks. The second study was a qualitative descriptive study that employed qualitative content analysis to investigate the tactics and techniques employed by cybercriminals during MIM phishing attacks, aiming to determine if they differed from those used in other media. In the third study, the thesis employed an experimental research approach to investigate the impact of key characteristics of MIM apps on susceptibility to phishing in such apps, using WhatsApp as a case study. Consequently, this thesis employed an exploratory, descriptive, and experimental research design to answer the research questions. Further details about these methods and justification for their adoption are provided in the following sections.

3.4.1 Understanding user behaviour during MIM (RQ1)

This thesis's first research question (RQ1) aims to understand whether MIM app users engage in behaviours that could expose them to security risks, such as phishing attacks. This approach is exploratory, as it aims to provide a deeper comprehension of a topic or evaluate a phenomenon (Swedberg, 2020). Indeed, exploratory research is valuable when the goal is to generate ideas, provide insights, establish priorities for future research, and identify variables for descriptive research (Stevens and Wrenn, 2013).

Given that MIM phishing is a new phenomenon and that current literature on phishing fails to provide insights into users' behaviours in such apps and whether such behaviours make them susceptible to phishing attacks, this research deemed that an exploratory study was the right approach to provide the needed understanding of this problem and develop the proper foundation for conducting a detailed study in the later stages of the research.

When conducting an exploratory study, researchers can select from various methods, including surveys/questionnaires, interviews, and focus groups. The exploratory study employed an online survey to gather information from users of MIM apps. The choice of a quantitative method over a qualitative one was due to the research question, which aims to describe the relationship between self-reported user behaviour and their risk of phishing in MIM apps. This type of research design is quantitative, as highlighted in Gelo, Braakmann and Benetka (2008). Furthermore, the study also aims to derive statistical conclusion validity, which seeks to draw inferences from the sample to the population—a characteristic associated with a quantitative research approach Gelo, Braakmann and Benetka (2008). Online surveys are currently among the most used techniques for collecting data about user perceptions and behaviours. This data collection method is fast and inexpensive (Van Selm and Jankowski, 2006) and is widely used in phishing susceptibility studies as highlighted in **Table 2-2**. Furthermore, the choice of online surveys to investigate RQ1 allows for data

collection across different segments of society (Sreejesh *et al.*, 2014), as this is essential for answering RQ1, given that this question targets MIM app users worldwide. Online surveys are generally suitable when research questions require large sample sizes or access to participants with varied demographics (Schmidt, 1997; Voit *et al.*, 2019).

Additionally, online surveys and questionnaires are helpful when the study aims to acquire an initial understanding of a phenomenon, a setting, or a population, particularly in areas where limited research has been conducted (Anderson and Lightfoot, 2022). Thus, this research approach is suitable for answering RQ1 since, at the time of the study, there was a lack of studies relating to the relationship between the behaviour of MIM app users and their exposure to phishing.

However, despite their advantages, researchers have pointed out that participants may not fully engage in the research when they are not being observed. For instance, they may guess answers or consult external sources (Clifford and Jerit, 2014). Yet, online surveys continue to be used for studies on susceptibility to phishing (Parsons *et al.*, 2019).

Details about the online survey/questionnaire and how other design issues, such as piloting and the recruiting process, were addressed are presented in Chapter 4.

3.4.2 Understanding the tactics and techniques used in the content of MIM phishing (RQ2)

The second research question aims to understand how cybercriminals exploit the features of MIM apps to conduct phishing attacks. This requires analysis of the content of MIM phishing messages to understand the strategies and tactics employed by cybercriminals during MIM phishing campaigns.

The objective is to examine the characteristics of these attacks and ascertain whether cybercriminals exploit the features of MIM to conduct phishing. The justification for conducting this study stems from the findings in the literature, which indicate that phishing techniques vary according to the communication medium, as highlighted in sections 2.2.3 and 2.2.5. Therefore, the second research question of this thesis (RQ2) aims to investigate the techniques and tactics used during phishing in MIM apps, thereby enhancing our understanding of the problem.

In section 3.2, it was stated that one of the notable qualitative descriptive research approaches is qualitative content analysis (QCA), which is used by researchers when the intention is to describe the characteristics of the document's content or when they aim to systematically categorise textual or visual data into categories, patterns, and themes for description and interpretation (Akdemir and Yenal, 2021). Thus, according to the aim of RQ2, QCA is deemed appropriate, as it can provide insights into the characteristics of phishing messages targeting MIM app users. Investigating these characteristics within a qualitative framework provides insights into cybercriminals' tactics and techniques, hence facilitating the identification of necessary technology and user awareness measures to curb the problem of phishing in MIM apps. Such insights may not be obtained through dependence on quantitative methods (Gillis and Jacksin, 2002) since quantitative approaches rely on numerical data to test and validate hypotheses. QCA is widely used in studies where little is known about the phenomenon under investigation (Green and Thorogood, 2018). Indeed, MIM phishing is a new phenomenon; thus, QCA is an ideal method for examining the tactics and techniques used in MIM phishing attacks. In Section 3.2, it was highlighted that researchers can apply QCA either inductively or deductively, depending on the research question, with the deductive approach being most commonly applied when existing theories or previous research on a phenomenon can be extended. While phishing in MIM apps may be a new phenomenon, existing research on phishing in other media makes the deductive content analysis

approach an ideal choice for answering RQ2. Furthermore, using deductive content analysis enables the comparison of findings from RQ2 with the current literature on phishing in other media, thereby helping to achieve the overall aim of RQ2.

Deductive QCA has been employed in studies examining the tactics and techniques used by cybercriminals in email-based phishing (Ferreira, Coventry and Lenzini, 2015; Akdemir and Yenal, 2021), as well as in SMS phishing and vishing (Jones *et al.*, 2020). More details about how these studies employed QCA to answer their questions can be found in section 2.2.3.

Whilst deductive QCA has been widely applied in phishing research, Section 3.2 highlights the limitations of this approach, including issues related to the reliability and trustworthiness of the process. As a result, the study, which utilised deductive QCA, applied the recommended strategies from the literature to enhance the trustworthiness and reliability of the findings. Chapter 5 presents further details about this and other stages of the QCA.

3.4.3 Investigating the effect of specific characteristics of MIM apps on susceptibility to phishing in MIM apps (RQ3)

The third research question for this thesis aims to investigate whether specific features of MIM apps influence susceptibility to phishing in these apps. When researchers study the effect of features or characteristics of communication media on phishing susceptibility, they often adopt methods from Human-Centred Security & Privacy (HCSP). HCPS is a field of research at the intersection of Human-Computer Interaction and computer security and privacy research (Garfinkel and Lipford, 2014; Renaud and Flowerday, 2017). Researchers use various methods when conducting security-related user studies, the most notable being lab-based, in-situ and online surveys (Voit *et al.*, 2019). Each technique has weaknesses and strengths, as highlighted in the subsequent paragraphs. Thus,

deciding which method to use depends on the study and the level of control researchers want over the study settings and environments (Voit *et al.*, 2019)

Online surveys are very popular in security-related user studies (Voit *et al.*, 2019; Distler *et al.*, 2021). However, a major disadvantage of this method is that participants may not fully participate in the research. For instance, they may guess answers or consult external sources (Clifford and Jerit, 2014). However, implementing mitigation strategies such as attention-check questions could prevent such problems (Paas and Morren, 2018).

Lab studies allow researchers to study participants' behaviours in controlled settings. They are prevalent in security-related user studies because they enable researchers to separate the effect of certain variables (Distler *et al.*, 2021). However, such types of studies do not represent the natural setting. Thus, results from such studies can have low external validity (Mathis, Vaniea and Khamis, 2021). Additionally, elements of lab studies, such as participant briefings, instructions, and research framing, could introduce biases in the study. For example, participants may claim to care more about the measured behaviour than they do in real life (Distler *et al.*, 2021).

An in-situ study will enable the researcher to observe participants in their natural settings, capturing their responses to phishing messages as a secondary activity embedded within primary activities (Franz and Croitor, 2021). However, they are costly and time-consuming (Kjeldskov *et al.*, 2004). In-situ studies require researchers to collect user activity and logs for an extended period. However, some researchers have argued that such an approach presents logistical and privacy concerns (Distler *et al.*, 2021). For example, the researchers (Lin *et al.*, 2019) monitored participants' internet activity for 21 days as part of an in-situ study on user susceptibility to phishing.

In-situ studies require researchers to introduce simulated attacks into a participant's real-world activities. However, researchers have argued that people

may acclimate to these attacks for research purposes and disregard actual attacks. Despite this, advocates of in-situ studies argue that such studies provide ecologically valid data (Franz and Croitor, 2021), as they require researchers to send simulated phishing attacks to users in their natural settings and record their responses.

Due to the low external validity of lab studies, difficulty and logistical and privacy concerns of in-situ studies, and limitations of online surveys, security-related user studies have used the vignette methodology to investigate how specific features/characteristics of the medium of communication affect susceptibility to phishing (Seng *et al.*, 2019; Klütsch *et al.*, 2024). A vignette study is a set of carefully constructed hypothetical scenarios proposed to respondents to assess dependent variables such as intentions, attitudes and behaviours (Aguinis and Bradley, 2014). The vignette study is structured according to a factorial survey design. As a result, it is sometimes referred to as a factorial survey experiment (Auspurg and Hinz, 2014).

Factorial surveys are described as quasi-experiments (Wallander, 2009) primarily because they blend experimental research concepts with conventional social surveys (Di Stasio, 2014). In a factorial survey, the researcher identifies key characteristics, also called dimensions or factors, that are expected to influence an individual's judgment of the outcome of interest. The researcher then assigns levels to each of the factors. However, the levels must be theoretically relevant (Di Stasio, 2014). After assigning levels to the dimensions, the researcher combines the levels at random to form a hypothetical description of an event or situation, known as a vignette. A vignette universe is established by crossing all factor levels to create all possible combinations. Depending on the number of combinations, the researcher can present the entire vignette universe to users or a subset.

In a vignette study, participants' intentions, attitudes and behaviours are elicited through survey-based evaluations of the hypothetical scenarios (Aguinis and

Bradley, 2014). A unique characteristic of a factorial survey is that scenarios can be text-based, images, video or other media (Hughes and Huby, 2002). A factorial survey enables researchers to manipulate and control independent variables, thereby enhancing the internal and external validity of experiments (Aguinis and Bradley, 2014). Factorial surveys have been used in human-computer interaction and cybersecurity studies (Harborth and Pape, 2021; van de Weijer, Leukfeldt and Zee, 2021). For example, Seng *et al.* (2019) conducted a vignette study to investigate users' decisions to respond to phishing messages on Facebook across different hypothetical scenarios. The study was administered as a survey using the Qualtrics platform. However, data collection took place in a controlled laboratory environment with 20 participants. The participants completed the survey under the supervision of the researchers. Participants were presented with a series of text-based vignettes. They were asked to indicate their likelihood of clicking on the links embedded within each scenario. According to the researchers, using vignettes enabled them to explore more scenarios, thereby helping to identify the potential attributes of Facebook posts that influence phishing susceptibility (Seng *et al.*, 2019). In another vignette study (Klütsch et al., 2024), the influence of users' relationship with the message sender and Fear of Missing Out (FoMO) on susceptibility to phishing on Instagram was examined. The online vignette study was hosted on the Pavlovia platform and was completed by 192 participants. Participants were instructed to imagine themselves in specific scenarios and respond to a series of simulated Instagram chat interfaces containing social media messages and potential phishing links. According to the researchers, using vignettes enabled them to systematically investigate their research objective within scenarios that more closely reflect real-world contexts than traditional questionnaire items (Klütsch *et al.*, 2024).

In van de Weijer, Leukfeldt and Zee (2021), the researchers used vignettes to examine the factors that influence the reporting of cybercrime victimisation by small and medium-sized enterprises in the Netherlands. The study used an online survey with vignettes. Participants were presented with three vignettes that

depicted hypothetical scenarios of cybercrime victimisation. They were instructed to imagine themselves as the victim in each situation and reflect on how they would respond under such circumstances. Like previous studies, the researchers employed the vignette method to measure the influence of four crime features on participants' likelihood of reporting victimisation to relevant authorities.

This thesis's third research question (RQ3) aims to investigate the effect of specific features of MIM apps on susceptibility to phishing within these apps. Therefore, the researcher has chosen factorial surveys as the appropriate method for this study because 1) this method allows the researcher to systematically investigate the effect of specific characteristics of MIM apps using scenarios that more closely reflect real-world contexts than traditional questionnaire items, 2) the vignette study also allow the study to be conducted using an online survey allowing for responses from a diverse set of participants.

Details about the factorial online survey/questionnaire and how other design issues, such as piloting and the recruiting process, were addressed are presented in Chapter 6.

3.4.4 Ethical considerations

When conducting research, researchers must consider ethical factors for all research activities, regardless of the participants or the subject being researched (Zikmund *et al.*, 2000). The current research involves human subjects as participants. Therefore, this research must adhere to the university's ethical standards, the Data Protection Act, and the General Data Protection Regulations. Among the three studies conducted by the researcher, two required ethical approval, whilst one didn't require ethical approval due to its reliance on publicly available data. The first study seeks to understand whether MIM app users engage in behaviours that could expose them to security risks, such as phishing attacks. As highlighted in 3.3.1, this research employed an online survey to

examine user behaviours during communication in MIM apps. Thus, due to the involvement of human subjects, the research applied for and granted ethical approval by the Department of Computer and Information Sciences at the University of Strathclyde. For reference purposes, the ethics approval ID is 1587. Similarly, the third study aims to investigate the effect of specific characteristics of MIM apps on susceptibility to phishing using a factorial survey design. Since the research involves human subjects, ethics approval was sought and granted by the Department of Computer and Information Sciences at the University of Strathclyde. For reference purposes, the ethics approval ID is 2398. The ethics approval addressed key issues regarding participant recruitment, including voluntary participation, consent, the tasks participants are expected to perform, the data that will be collected, how the data will be processed and stored, the retention period, and data disposal.

The second study, which analysed the content of MIM phishing messages to understand the strategies and tactics employed by cybercriminals during MIM phishing campaigns, didn't require ethics approval because it was based on publicly available data.

3.4.5 Data Collection Procedure

To answer RQ1 and RQ3, a web-based survey has been developed as an evaluative instrument utilising the Qualtrics online survey platform. For RQ1, the survey was used to examine whether MIM app users engage in behaviours that could expose them to security risks. For RQ3, the survey instrument was used to investigate the effect of some specific characteristics of MIM apps on phishing susceptibility in MIM apps. Qualtrics was chosen because it was made available by the researcher's institution. Furthermore, Qualtrics also offers researchers features such as the ability to randomise questions, implement participant screening questions, and design complex survey flows. For RQ2, sample images were collected from the Google search engine, following the approach used in

similar studies (Akdemir and Yenal, 2021). Details about the data collection process are presented in the respective chapters of the studies.

3.4.6 Data Analysis

As highlighted in Section 3.2, this research employed a mixed-methods approach to answer the research questions. Thus, the analysis employed reflects this choice by combining both quantitative and qualitative analysis methods. Quantitative data analysis comprises cleaning, screening, coding, and selecting the most suitable statistical analysis method. For each of the two quantitative studies, RQ1 and RQ3, respectively, the data from each study were analysed using a two-phase approach: an initial or preliminary analysis and a detailed analysis using statistical techniques. The initial/preliminary analysis familiarised the researchers with the data, whilst the in-depth analysis was used to answer the research questions of the respective studies.

The preliminary analysis of the data collected from questionnaires was performed using SPSS statistical software. The data was exported from Qualtrics, where it was collected. The data were then imported into SPSS, cleaned, and analysed preliminarily. Specifically, the cleaning phase checks for incomplete responses and evidence of straight-lining, which occurs when participants provide similar responses to a series of questions using the same response scale, potentially compromising data quality. The cleaning phase also checks for respondents who failed attention checks. Further details about these activities are provided in the respective study Chapters. Furthermore, frequency graphs were used to illustrate the frequencies and percentages of each response in cases where the response was nominal or ordinal.

An in-depth analysis of the first quantitative study (RQ1) began with identifying the proper statistical tests that consider the measurement scale used. Non-parametric tests, including the Wilcoxon signed-rank test, the Friedman test, and Spearman's correlation coefficient, were used to assess distributional differences

among groups or correlations between variables. These tests are appropriate for data on an ordinal measurement scale (Jamieson, 2004). Nevertheless, the researcher acknowledges that the discourse regarding the suitability of parametric versus non-parametric tests for ordinal data remains unresolved (Norman, 2010). Furthermore, the researcher agrees with the views of some researchers that descriptive statistics, such as means and standard deviations, have unclear meanings when applied to Likert scale responses, necessitating the use of Median as a central tendency measure (Sullivan and Artino Jr, 2013). This group of researchers argues that whilst the responses of ordinal scales can be ordered or ranked, the magnitude of difference between response categories cannot be quantified. Consequently, the intervals between categories such as "always," "often," and "sometimes" on a frequency-based Likert scale cannot be assumed to be equal. Although numerical values may be assigned to these categories for analytical purposes, these values do not represent equidistant intervals. This distinguishes ordinal data from interval-level data, where differences between values are meaningful, measurable, and reflect a consistent underlying quantity (Sullivan and Artino Jr, 2013).

In addition to the preliminary analysis of the factorial survey experiment (RQ3), an in-depth analysis was conducted. The factorial survey collected both quantitative and qualitative data from participants. The quantitative analysis tested the study's hypotheses using ordinal logistic regression with Cluster-robust standard errors, as the dependent variables were measured on ordinal scales (Jamieson, 2004).

For the qualitative content analysis (QCA) of RQ2, the guidelines from McKibben *et al.* (2020) were followed. The data from the coding sheet were imported into SPSS, where initial analysis was conducted, generating frequency counts and percentages as recommended by McKibben *et al*. (2020). The descriptive data tells the reader how frequently a particular category appears in the sampled unit.

An in-depth analysis of qualitative content analysis (QCA) results was conducted using the Python programming language. This approach enabled the researcher to loop through the data, searching for specific combinations of tactics and techniques as required by some of the study's questions. For this task, the researcher employed several Python data analysis packages, including Pandas, SciPy, and NumPy.

The qualitative data from the factorial survey were analysed using deductive thematic analysis. Deductive thematic analysis was employed to identify the themes that participants reported as influencing their decisions. This approach relies on generating themes of interest from the literature before data analysis (Braun and Clarke, 2006). The approach is suitable in the current context, as it enables comparison with previous studies that have explored similar themes. The thematic analysis was based on Braun and Clarke's recommended six-step (Braun and Clarke, 2006), which are: 1) data familiarisation, 2) coding according to the previously known themes, 3) creating new themes inductively where data didn't match, 4) reviewing the naming of the new themes, 5) defining the themes and 6) and writing report.

3.5 Chapter Summary

This chapter provides a detailed description of the research methodologies used in this thesis. The chapter discusses the methods employed in this thesis and justifies the reasons for their adoption. This thesis employed three research methods, beginning with an exploratory study that utilised an online survey to collect data on the behaviours of MIM app users. The qualitative and descriptive research employed qualitative content analysis (QCA) to understand the tactics and techniques used in MIM phishing attacks. Finally, a factorial survey, a quasi-experiment design, was used to investigate whether specific characteristics of MIM apps affect susceptibility to phishing in MIM apps. This was followed by a detailed explanation of the data analysis strategy employed and a justification of why they were considered appropriate for the data collected in this thesis.

Chapter 4

An exploratory study of user behaviours during Mobile Instant Messaging

This chapter presents a study that investigates the behaviours of MIM app users towards links shared during communication in MIM apps with the overarching goal of understanding whether their behaviours expose them to the threat of phishing. The chapter also discusses the study design, including questionnaire design and selection of target participants. Finally, the chapter presents the study's results, along with a discussion of the findings.

4.1 Introduction

Section 2.2.2 highlighted that some features of MIM apps, such as the ability to share links, forward links and join private and public groups, can expose unsuspecting users to phishing risks. Furthermore, these apps currently lack effective anti-phishing countermeasures, raising concerns among researchers regarding user safety. As highlighted in Section 2.2.1, one of the key factors in understanding user susceptibility to phishing is understanding the key user behaviours that may expose them to phishing risks. Consequently, a key factor in understanding user susceptibility to MIM phishing is understanding how users behave while communicating in these apps, specifically whether they exercise caution while using the features of the apps that are likely to expose them to phishing. Prior work on susceptibility to phishing in MIM apps fails to provide insights into users' behaviours in such apps and whether such behaviours make them susceptible to phishing attacks. Through an exploratory study, the first

study of this thesis investigates the degree to which users participate in behaviours within MIM apps that put them at risk of being exposed to phishing. Specifically, it focuses on the self-reported behaviours of users of these apps towards links shared during communication in such apps.

The study aims to provide insights into users' attitudes towards links shared on such platforms and the factors influencing their decision to perform specific actions, such as clicking and forwarding links. The study contributes to the knowledge of this topic by identifying potential gaps in the strategies users adopt to protect themselves from phishing in MIM apps.

4.2 Study Aim and Research Questions

This study aims to understand if users of MIM apps engage in behaviours that could expose them to security risks, such as falling victim to phishing attacks. The study in this chapter aims to address RQ1 of the thesis, which is restated below:

RQ1: Do users of MIM apps exhibit behaviours that could expose them to the risk of falling victim to a phishing attack?

The above question aims to understand how susceptible users of MIM apps are to phishing attacks targeting them by examining their self-reported behaviours to understand if such behaviours expose them to phishing attacks. To successfully answer this question, it was further refined to the following five sub-questions. The refinement was supported by the related work in section 2.3.1. Specifically, section 2.3.1 highlighted that users' awareness of phishing attacks can impact how they respond to phishing messages. As a result, sub-question one was introduced to explore if this is the case in MIM apps. Section 2.5 showed that users are frequently provided with tips on how to protect themselves from phishing threats. As a result, sub-question two was introduced to explore if MIM app users follow the recommended strategies. Section 2.3.1 revealed that habits or the degree to which users automatically perform behaviours can influence their susceptibility to phishing. As a result, sub-question three was introduced to

investigate users' habits regarding clicking and forwarding links. Furthermore, the discussion on the characteristics of mobile instant messaging apps in section 2.3.2 has highlighted how communication in MIM apps is likely to be between known contacts but has also raised concern regarding evidence from the literature that indicates that users assess the credibility of online resources using a social confirmation or social proof approach, perceiving information as credible because others have deemed it credible. As a result, sub-question four was introduced to investigate whether the relationship between the recipient and sender of a message, or the type of group, impacts users' behaviours towards links shared in terms of whether they would click or share the links.

Finally, section 2.3.1 shows that other factors, such as users' phishing self-efficacy and their concern about the consequences of falling for such attacks, can affect their susceptibility. Thus, sub-question five was introduced to explore if this is the case in MIM apps.

RQ1.1: Are users of MIM apps aware of the phishing attacks using MIM messages to target them?

RQ1.2: What strategies do MIM app users employ to safeguard themselves from phishing attempts?

RQ1.3: How often do people click on and forward links shared in various MIM apps?

RQ1.4: Does the relationship between the recipient and sender of a message, or the type of group, impact users' behaviours towards links shared in MIM apps?

RQ1.5: What other factors influence MIM app users' behaviours towards links shared during instant messaging?

4.3 Study Design

As discussed in the methodology section of this thesis, the study in this chapter employed a quantitative research approach using a web-based survey tool, Qualtrics, to gather data from users of MIM apps. The choice of a Quantitative approach for this study was justified in 3.2.2. Furthermore, justification for using Qualtrics was provided in section 3.4. The survey was conducted in English. The survey primarily examined the self-reported behaviour of MIM Apps users when interacting with shared links in one-to-one and group-based conversations. The study focused on links because existing email phishing studies (Alohali *et al.*, 2018; Abroshan *et al.*, 2021) have primarily focused on users' behaviours towards links during email communication, as highlighted in Section 2.2.1.

Additionally, the study aims to investigate whether the self-reported behaviour of users differs based on whether the communication is one-to-one or group, since Section 2.2.2 has highlighted the potential of group-based communication to expose users to phishing risks. Furthermore, the online survey also includes questions regarding the participants' self-efficacy in detecting phishing attempts on MIM apps and their self-efficacy in using mobile devices. Examining users' self-efficacy aligns with the current literature, which identifies self-efficacy as one of the factors contributing to susceptibility to phishing in email, as highlighted in Section 2.2.1. The survey also contains questions on participants' awareness of phishing and their level of concern regarding phishing attacks in MIM Apps.

Ethics approval was granted by the researchers' departmental ethics Board, as highlighted in section 3.3. The participants' information and consent forms for this study are available in Appendix A.1.

4.3.1 Survey Instrument

The survey used in this study is designed to collect data on participants' self-reported behaviour of interacting with shared links in one-to-one and group-

based conversations in MIM apps. The reason for examining the participants' link-clicking behaviour in both one-to-one and group-based communication was to assess whether there were any differences in the participants' behaviour between these two conditions. The sections in the survey are:

Participant demographics: Four demographic questions related to age, gender, education, and country of residence, similar to most phishing studies in the literature, as highlighted in Section 2.2.1.

The sections in the survey are:

Participant demographics: Four demographic questions related to age, gender, education, and country of residence, similar to most phishing studies in the literature, as highlighted in Section 2.2.1. The researcher created the questions in this section, and they include:

- Please specify your age
- Please specify your gender
- Please specify your level of education
- In which country do you currently reside

Device and MIM app usage questions: The first part of this section contains five questions that enquired about the participants' present information and communication technology (ICT) devices and their frequency of usage. The participants were asked if they used any of the following devices: desktop, laptop, smartphone, tablet, and other ICT devices. The questions were based on a Likert-type scale and measured frequency using the following options: never, rarely, at least once a month, at least once a week, multiple times a week, and every day. These options are similar to those used in the literature on media usage (Shimoga, Erlyana and Rebello, 2019). This section allows the researcher to understand the type of device used by the participants and how frequently such devices are used.

The second part of this section presents a mobile computing self-efficacy construct adapted from Keith *et al.* (2015). The mobile computing self-efficacy construct was included to measure the participants' perceptions of their ability to use mobile phones to accomplish specific tasks. The scale has four items that measure participants' competence in smartphone use. It consists of questions relating to participants' abilities to: make purchases using mobile devices, correct common smartphone problems, add and remove features and apps and effectively use these apps. Prior literature has linked high mobile self-efficacy to risk-taking behaviours (Keith *et al.*, 2015). As a result, this study aims to find if there is a correlation between mobile computing self-efficacy and risk-taking behaviours during MIM.

The third part of this section contains two questions, one of which asks participants to specify the MIM apps they presently use from a curated list of applications, along with the frequency of their usage and the individuals they connect with through these applications. The advent of smartphones and mobile applications has led to the proliferation of diverse MIM apps, each possessing distinct functions and capabilities. However, the objective of this study was to examine the behaviours of users using specific applications, notably Signal, Slack, Telegram, Viber, Line, and WhatsApp. Thus, the curated list comprises these applications since they are currently the most popular (Stivala and Pellegrino, 2020). The frequency of usage was measured using a five-point scale with the following options: never, rarely, sometimes, very often, and always. The primary objective of this question was to eliminate people who do not utilise any of the apps mentioned in the list. The second objective was to ascertain the MIM apps the participants presently utilise.

The second question asked participants to indicate with whom they use these applications to communicate from a list of possible communication parties. The communicating parties include friends, family, work or business colleagues, other acquaintances, and unknown senders. These types of senders were selected based on findings from the literature that show users of MIM applications

primarily use these applications to communicate with friends, family, work/business colleagues (Quan-Haase and Young, 2010), as well as other acquaintances (Avrahami and Hudson, 2006), as highlighted in section 2.2.2. The sender of the type 'contact you don't know" was included, as the current setup of MIM apps allows strangers to send messages to users. Furthermore, users can be exposed to messages from strangers during group-based communication.

The actual questions for this section of the survey are presented below. The researcher created questions one, three, four, and five. Question two on the mobile computing self-efficacy construct was adapted from Keith et al. (2015).

- Please indicate how frequently you used any of the following devices
- Please indicate whether you agree or disagree with the following statements.
  - I believe I have the ability to make purchases using a mobile device
  - I believe I have the ability to identify and correct common problems with mobile devices
  - I believe I have the ability to install and remove features and applications from mobile devices
  - I believe I have the ability to use the productivity features offered by mobile devices (e.g., calendar, email, task scheduling, etc.).
- Which of the following mobile instant messaging applications do you currently have an active account with and use? (Check all that apply)
- How frequently do you use these applications?
- Who do you use these applications to communicate with? (Check all that apply)

Frequency of clicking links in MIM apps: This survey section assessed the participants' self-reported behaviour when clicking on links while communicating in MIM apps. The section contains a total of 17 questions, 15 of these questions were evaluated using a Likert-type scale with the following options: never, rarely, sometimes, very often, and always. The 15 questions include:

Two questions evaluated the frequency with which they clicked on links during one-on-one and group-based instant messaging. The researcher created all the questions. The questions are:

- How frequently do you click links when communicating in these Applications?
- For the groups you are part of, how frequently do you click on links sent by other members?

Five questions assessed the frequency with which participants clicked on links from the different categories of people they communicated with during one-to-one interactions. The researcher created all the questions. The questions are:

- How frequently do you click on links sent to you by your friends?
- How frequently do you click on links sent to you by your family members?
- How frequently do you click on links sent to you by your work/business colleagues?
- How frequently do you click on links sent to you by contact you know, but are not your friends, family or work/business colleagues?
- How frequently do you click on links sent to you by contact you do not know?

Five questions assessed the frequency with which participants clicked on links from the different categories of people they communicated with during group-based interactions. The researcher created all the questions. The questions are:

- For the groups you are part of, how frequently do you click on links that are sent in such groups by your friends?
- For the groups you are part of, how frequently do you click on links that are sent in such groups by your family members?
- For the groups you are part of, how frequently do you click on links that are sent in such groups by your work/business colleagues?
- For the groups you are part of, how frequently do you click on links that are sent in such groups by contacts you know, but aren't friends and family or work/business colleagues?
- For the groups you are part of, how frequently do you click on links that are sent in such groups by contacts you don't know?

One question about how frequently the participants checked the link preview before clicking a link, which was evaluated using a Likert-type scale with the following options: never, rarely, sometimes, very often, and always.

- How frequently do you check the link preview before clicking a link?

The last two questions of this survey section were created by the researcher and aimed to assess the phishing preventive behaviour of the participants while engaging with links. The questions are based on a Likert-type scale with the following options: never, rarely, sometimes, very often, and always. These actions are considered phishing preventive because security specialists often advise users to inspect the link before clicking (Reeder, Ion and Consolvo, 2017; Indiana University, 2022). The actual questions comprise the following:

- How often do you check the actual links sent during chats before clicking them?

- How frequently do you consider the sender of a message before clicking the link in that message?

One question seeks to know if the participants are part of any MIM app groups. The researcher created the question.

- Are you a member of any group?

One question asked if the participants are aware of the type of groups they are part of (i.e., whether private or public). An explanation of the two group types (public and private) was included in the question to help participants identify which category of groups they belong to. The researcher created the question.

- Do you know whether the groups you are a member of are private or public?

Frequency of forwarding links: This section assessed the participants' frequency of forwarding links to others. It contains six questions, evaluated using a Likert-type scale with the following options: never, rarely, sometimes, very often, and always. Two questions ask about the frequency with which the participants share links they received from others with their contacts. Two questions ask about the frequency with which the participants forward links they receive in public or private MIM groups to others. The final two questions inquire about the frequency with which participants assess the credibility of the sender before forwarding a link to others and whether they verify the links before sharing them with others. These questions were raised due to the message-forwarding feature of MIM apps, which, as highlighted in 2.2.2, enables users to instantly share received messages with a large audience, both through one-on-one and group-based communication. This tendency can facilitate the quick

dissemination of fraudulent messages throughout the network. The researcher created all the questions. The questions are:

- How frequently do you forward links to others?
- Do you follow links before forwarding them?
- How frequently do you consider who the sender of a link is before forwarding it to others?
- How frequently do you forward links that are shared with you by others?
- How frequently do you forward links that are shared in public groups
- How frequently do you forward links that are shared in private groups

Awareness and Concerns about Phishing in MIM Apps: This section contains five questions. The questions in this section assessed whether the participants were aware of phishing in MIM apps and if they were concerned about this threat. The participants were asked to express their familiarity with phishing using a 5-point Likert scale, ranging from "strongly disagree" to "strongly agree." They were also asked if they were aware of general phishing using yes/no options. Although the participants may have some knowledge of what phishing is, their familiarity is probably limited to email-based phishing, the oldest and most widely recognised type. To assess the participants' understanding of MIM-based phishing schemes, another question asked the participants to indicate whether they were aware of MIM phishing scams using yes/no options. The researcher created all the questions. The specific questions are:

- Please indicate whether you agree or disagree with the following.
- Are you aware of phishing scams?

- Are you aware of any mobile instant messaging scams?

Participants were instructed to express their level of concern regarding phishing on MIM Apps, using a scale ranging from "no concern" to "extreme concern." This question was because section 2.2.1 has highlighted that concern/anxiety about phishing is a predictor of phishing susceptibility. The researcher created all the questions. The specific question is:

- How concerned are you about phishing scams on mobile instant messaging applications?

In addition, the participants were asked to select how they respond to the problem of phishing scams on mobile instant messaging applications. They were provided with five statements to choose from. The statements include 1) I look for suspicious links, 2) I check the link preview, 3) I do not click links from unknown senders, 4) I search Google for further information on the website before giving out my details, and 5) I do nothing. These options align with the existing recommendations provided by cybersecurity experts to internet users regarding how to respond to phishing attempts (Orunsolu *et al.*, 2017). The specific question is:

- Which of these statements best reflect how you respond to the problem of phishing scams on mobile instant messaging applications? (Select all that apply)

Participants' phishing self-efficacy: Phishing self-efficacy is defined in this study as the participant's belief in their ability to effectively apply security measures that safeguard them against phishing threats (Liang and Xue, 2010). A scale was developed to measure an individual's self-confidence in identifying and avoiding phishing attempts in mobile instant messaging (MIM) apps. The scale consists of three questions, which were adopted from Li *et al.* (2014). The phishing self-efficacy questions assessed the participants' confidence in their

ability to safeguard themselves from phishing attacks. The questions were based on a five-point Likert scale, ranging from "strongly disagree" to "strongly agree."

- I believe I have the knowledge and skills to identify phishing scams when they are presented to me during communication on mobile instant messengers.
- I believe that I can protect my personal information from being stolen by phishers.
- I believe I have the skills to implement security measures to stop people from getting my confidential information.

Protection from others: Two questions asked participants whether they believed in the capabilities of their peers and MIM platforms to protect them from phishing. The questions were based on a five-point Likert scale, ranging from "strongly disagree" to "strongly agree." First, the question about the participants' belief in peers arose because MIM is primarily informal and occurs between individuals with some offline connections. Thus, participants' familiarity with their peers may influence their belief in their peers' ability to protect them. Second, the hype surrounding the privacy and security of MIM apps may influence their perception of the level of protection these apps offer. The researcher created all the questions, and they included:

- My contacts have the required skills and knowledge to detect phishing links and therefore can protect me.
- I believe mobile instant messengers have mechanisms in place to protect me from opening malicious links.

Phishing Protection Responsibility: This section comprises three questions that ask participants about their opinions on who is responsible for safeguarding

MIM app users from phishing scams. Existing literature suggests that both social media platforms and their users are responsible for addressing the problem of phishing (Parker and Flowerday, 2020). As a result, this part of the survey aims to investigate whether the users are aware of this responsibility. The researcher created all the questions, and they included:

- I believe it is the responsibility of mobile instant messaging platforms to protect their users from phishing scams.
- I believe mobile instant messengers' users should be responsible for protecting themselves from phishing scams.
- Protecting users from phishing scams should be a shared responsibility between mobile instant messaging platforms and their users.

4.3.2 Target participants and recruitment

Participation necessitated individuals to be at least 18 years old and possess the ability to complete an English survey successfully. In addition, the sample was restricted to individuals who utilise WhatsApp, Signal, Slack, Line, Telegram, and Viber due to their widespread usage and the range of functionalities they offer, including group communication, link previews, link sharing, forwarding messages/links, and the capability to join public groups through links shared by group administrators. This was done to ensure that the questions were meaningful and applicable to the participants.

The data collection strategy employed a snowball sampling methodology and utilised social media platforms to enlist volunteers. Snowball sampling is a non-random sample technique used to recruit study participants who are difficult to access or unfamiliar (Rashidi, Vaniea and Camp, 2016). Snowball sampling also allows researchers to recruit socially networked populations (Noy, 2008). These features are essential for the current research, as the target audience for this

study are MIM app users, who, as highlighted in 2.2.2, use these platforms for informal communication and are heavily reliant on offline connections.

The recruitment process begins when the researcher distributes the survey link to individuals in his existing MIM Apps network, including private and public groups. Furthermore, the survey link was disseminated through multiple social media platforms, including r/SampleSize on Reddit. Reddit is a large online discussion and content-sharing platform where users post links, text, images, and videos, and engage in threaded discussions. The link was also posted on samplesize on Facebook and SurveyCycle. Ultimately, the survey link was emailed to the researcher's colleagues, requesting their participation and urging them to forward the survey to further individuals. Throughout the recruitment process, the researchers asked participants to extend invitations to others to participate in the study by distributing the survey link.

4.3.3 Procedure

Before advertising the study and recruiting participants, the researcher sought and obtained ethics approval from their department. The advertisement described the purpose of the research, its procedures, the expected completion time, and the benefits. Participants were required to read and sign an informed consent document before participation. The consent form can be found in Appendix A.1 of this thesis. The participants were told that the study's goal is to determine whether the behaviours of users of mobile instant messaging applications expose them to security risks, such as phishing and that the survey hopes to provide insights about users' attitudes towards links that are shared on such platforms and the factors that influence their decision to perform specific actions such as clicking and forwarding of these links. Next, the participants were presented with a brief description of the task they would be required to do.

Before conducting the main study, a pilot study was conducted to help identify potential problems that might cause low participation or prevent participants

from responding to specific questions. The pilot study used six postgraduate students from departments other than the researcher's. The pilot study aimed to 1) identify the amount of time necessary to finish the survey, 2) solicit participants' feedback on the survey design, and 3) identify any instances of ambiguity in questions or statements.

The pilot study results show that the survey took an average of 15 minutes. The results also revealed that the survey was well-organised and clear. However, the participants suggested using "links" instead of "URLs", as many people may not know what "URL" stands for. Thus, some questions/statements were reworded due to these findings.

The main study was conducted from October 12, 2021, to November 5, 2021.

4.4 Results

This section presents the study's findings. First, data-cleaning strategies were discussed, including the number of participants excluded from the analysis and the justification for doing so. Data cleaning and analysis were conducted in SPSS. Please note that the term "frequency of clicking links" was used to indicate how often participants engage in the behaviour of clicking links.

4.4.1 Participants

A total of 129 participants accessed the survey. However, 18 participants who did not meet the screening criteria were excluded after the data cleaning. Exclusion criteria were: 1) whether the participants declined to participate in the study after reading the consent and participants' information note (n=2), 2) whether the participants did not use any of the MIM platforms mentioned in the study (n=6), 3) the participants failed to provide sufficient data (i.e. not answering questions on URL clicking and forwarding behaviour (n=8) and 4) whether the provided consent for the survey but did not respond to any of the survey

questions were removed (n=2). Therefore, the analysis was conducted using a sample of 111 participants who provided adequate data and fulfilled the study's screening criteria.

4.4.2 Participants demographics

Table 4-1 presents the demographic distribution of the respondents, indicating that most participants were Male (n=73) and aged 18-30 (n=54). Most participants also held a postgraduate education (n=60).

Table 4-1:Demographic Information of the Participants

		Frequency	Percentage
Gender	Male	73	65.8%
	Female	37	33.3%
	Prefer not to disclose	1	0.9%
Age Group	18-30	54	48.6%
	31-45	51	45.9%
	46-65	6	5.4%
	65 and above	0	0%
Level of Education	Secondary	5	4.5%
	Further education (i.e. colleges)	13	11.7%
	University undergraduate	33	29.7%

	Postgraduate (e.g. Masters, Doctorate)	60	54.1%

Responses were received from participants living in several countries. Figure 4-1 shows that the highest participation is from the United Kingdom (64, 58.7%), followed by Nigeria (18, 16.5%).

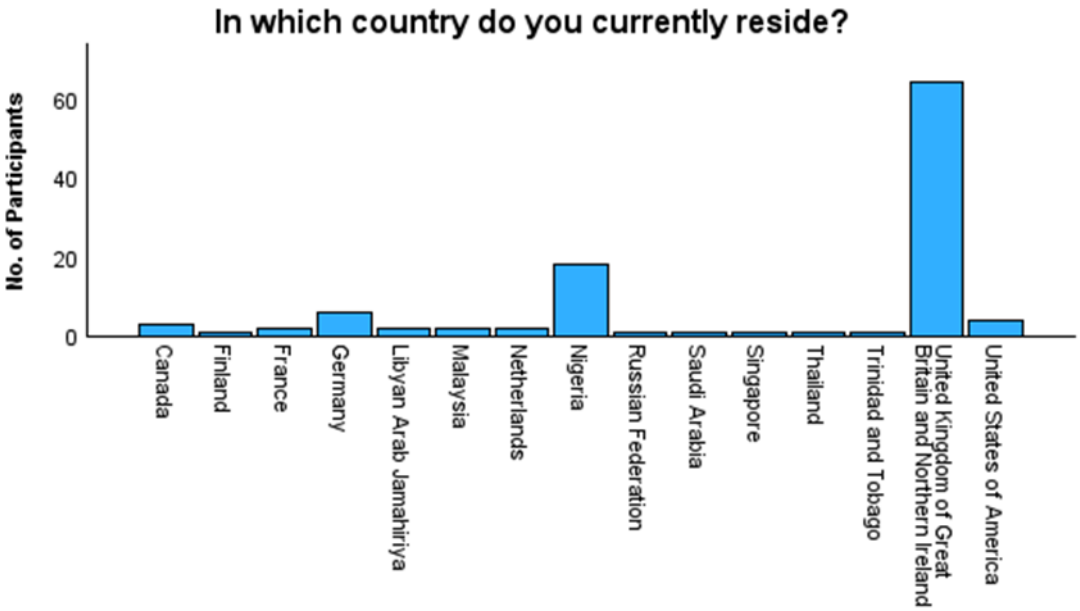


Figure 4-1:Number of respondents per country of residence

4.4.3 Participants' current ICT devices and MIM apps' usage

To gain a deeper understanding of the study participants, they were asked to specify the ICT devices they currently use and the frequency with which they use them. According to Figure 4-2, smartphones are the most frequently used devices, with 109 respondents (98.2%) reporting daily use. Laptops are the second most commonly used devices, with 69 participants (62.2%) currently using them, followed by desktops with a usage rate of 36 (32.4%).

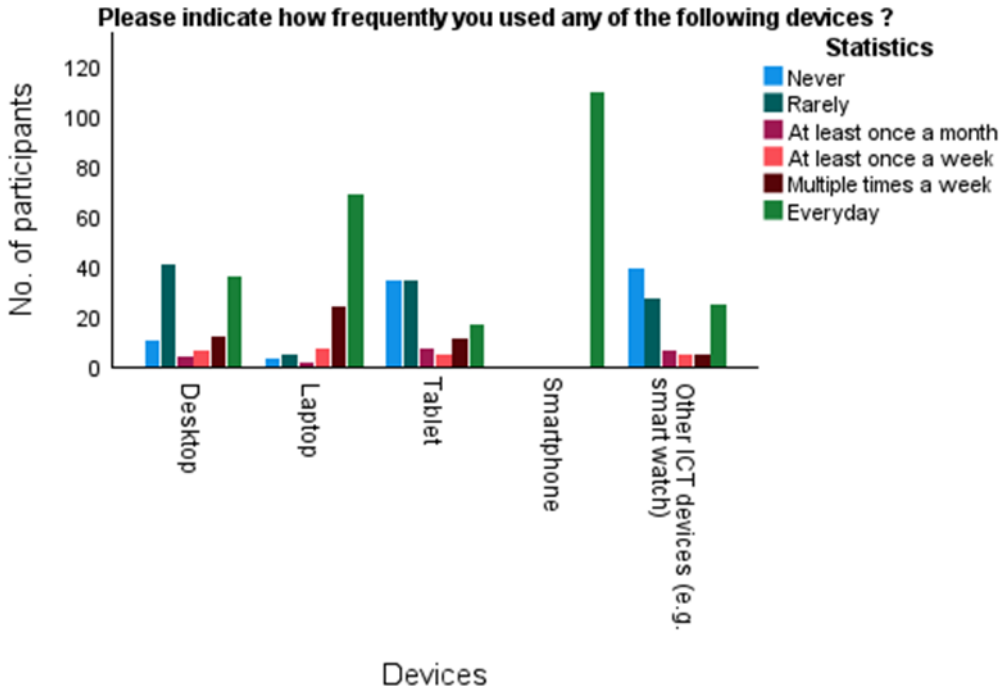


Figure 4-2: Participants' current devices and frequency of usage

According to Table 4-2, most participants (106, 95.5%) currently use WhatsApp, whereas Telegram is utilised by a smaller percentage (40, 36.0%). Additionally, participants were asked to indicate how frequently they used any of the apps mentioned in the survey. According to Figure 4-3, WhatsApp and Telegram are the platforms most frequently used by the participants.

Table 4-2: Participants current MIM apps

Answer Choices	**Responses**
WhatsApp	95.5% (n=106)
Line	3.6% (n=4)

Viber	5.4% (n=6)
Telegram	36.0% (n=40)
Slack	11.7% (n=13)
Signal	17.1% (n=19)

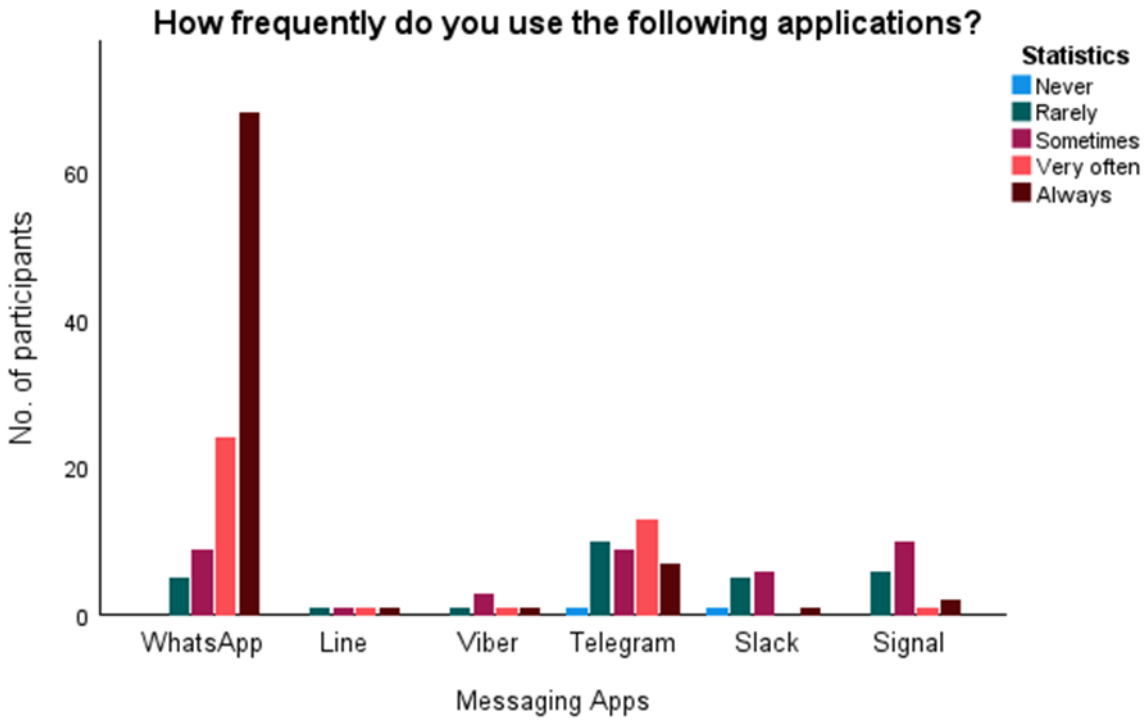


Figure 4-3:Participants' frequency of MIM Apps usage

Participants were also asked to specify the category of contacts they used MIM apps to communicate with. According to Figure 4-4, most participants (97, 87.4%) use MIM Apps for communication with friends, and (92, 82.9%) use them to communicate with family. However, as depicted in Figure 4.4, participants also

utilised MIM apps to interact with other categories of users, particularly work/business colleagues (80, 72.1%). But they also communicate with other acquaintances (47, 42.3%) and unknown contacts (14, 12.6%).

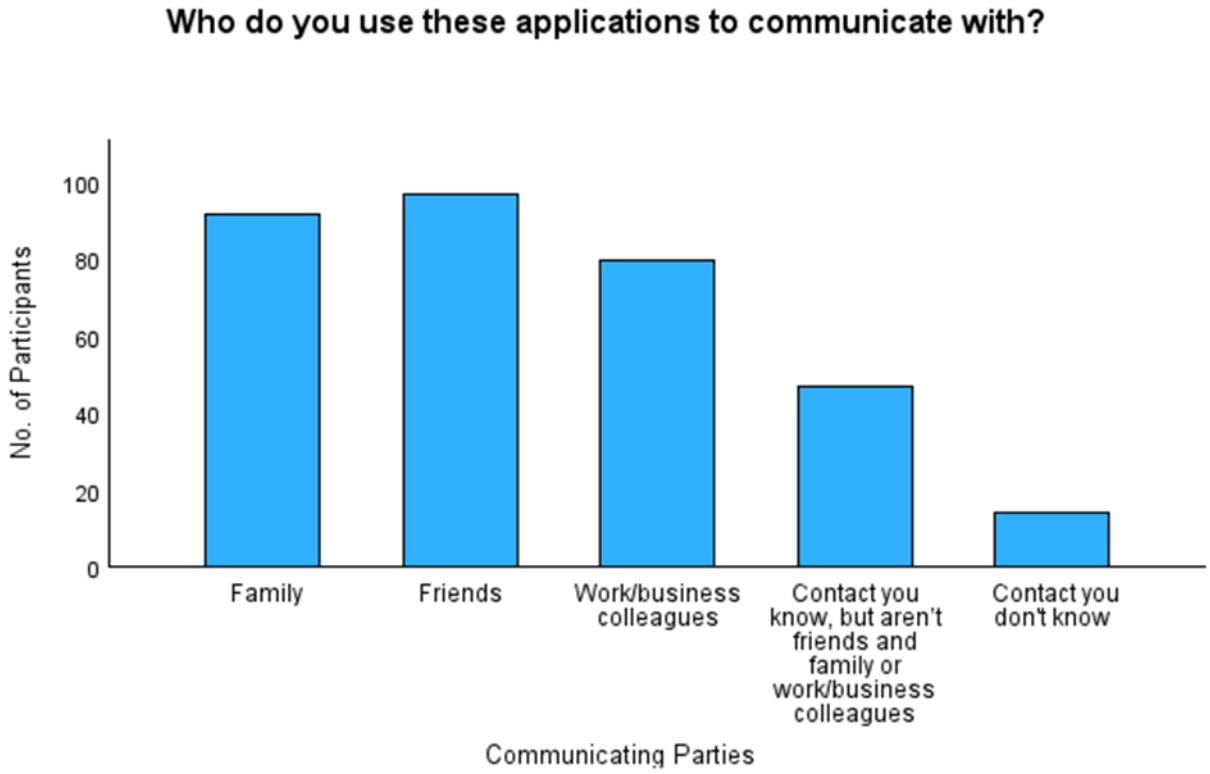


Figure 4-4: Participants' communicating parties (n=111)

Next, the participants were asked if they were members of any group in MIM Apps. As seen in Figure 4-5, (38, 35.5%) of the participants reported being members of groups for all the applications mentioned in this study, and (36, 33.6%) stated they belonged to groups in some apps but not all. In contrast, only (4, 3.6%) said they do not belong to any group.

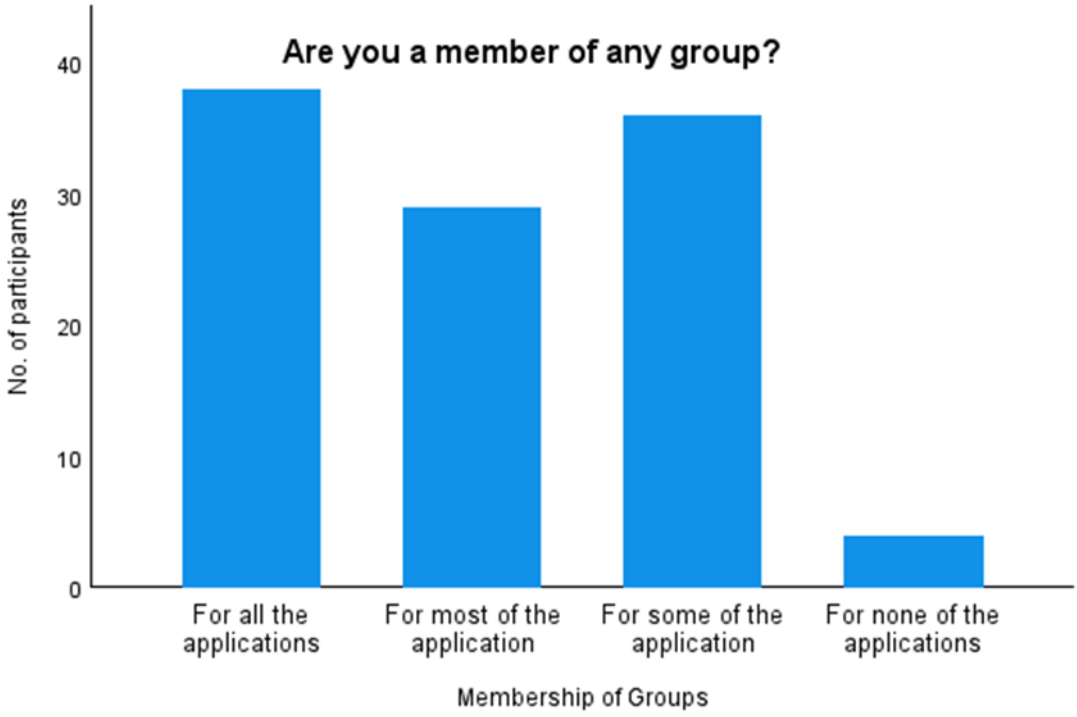


Figure 4-5: Group membership (n=107)

Subsequent examination revealed that all participants who stated they were not part of MIM Apps groups were users of WhatsApp (n = 4). One of the participants utilised Signal in conjunction with WhatsApp. Additionally, these individuals seem to utilise MIM Apps to communicate with friends, family, and work/business colleagues.

As seen in Figure 4-6, most participants reported knowing the type of groups they are part of (71, 68.9%).

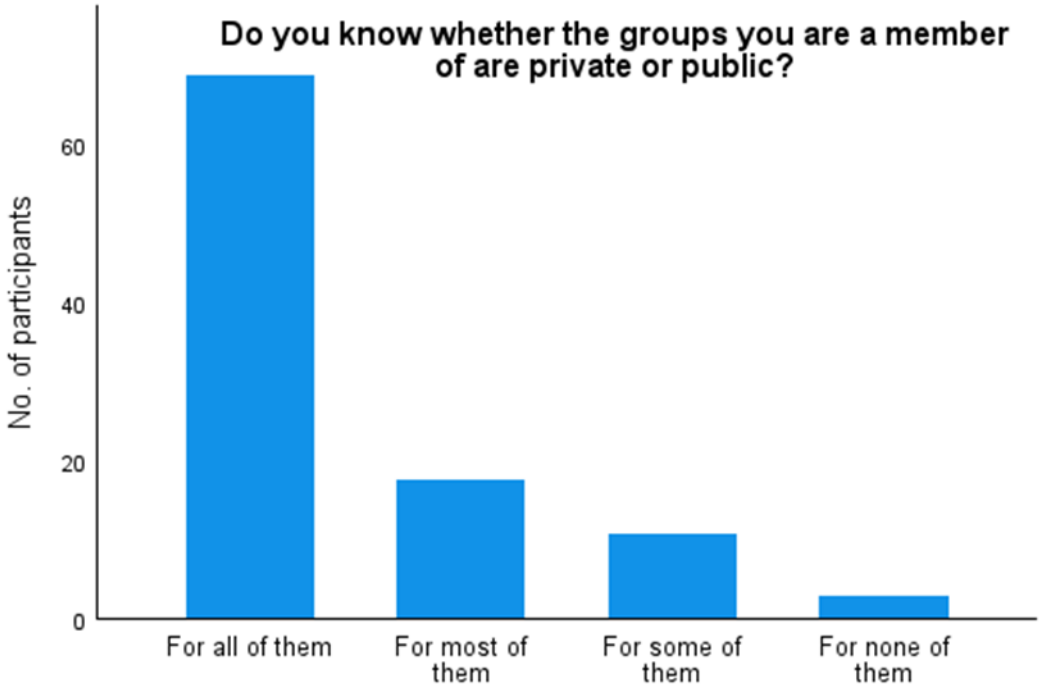


Figure 4-6: Participants' knowledge of types of groups (n=103)

4.4.4 Participants' awareness and concerns regarding phishing in MIM apps

Figure 4-7 shows that most participants agree or strongly agree when asked if they know what phishing is, indicating a high level of familiarity with the term.

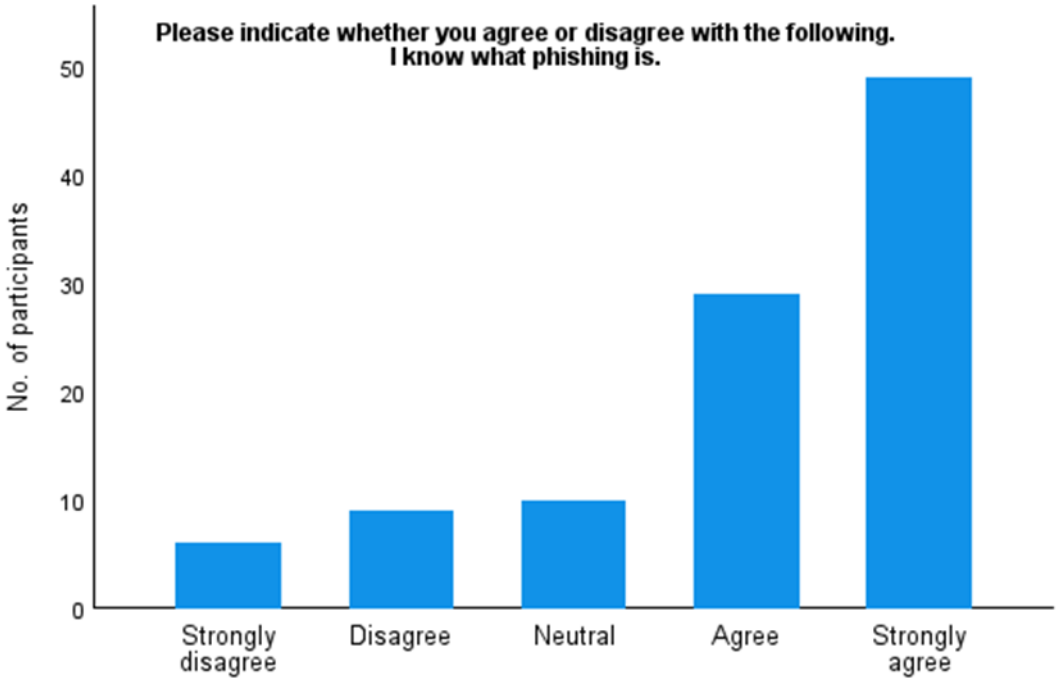


Figure 4-7: Participants' perceived understanding of what phishing (n=103)

Having expressed their knowledge of phishing, the participants were asked to indicate whether they were aware of general phishing scams. According to the data presented in Figure 4-8, most participants (n=86) were aware of phishing. Nevertheless, a considerable proportion (n = 18) indicated a lack of awareness.

Due to the novelty of phishing in MIM apps, it is anticipated that some participants may be unaware of these types of threats. Therefore, the participants were further asked to indicate if they were aware of MIM phishing scams. Figure 4-8 shows that most participants were aware of MIM phishing (n = 85). However, a considerable proportion (n = 19) indicated a lack of awareness. When asked whether they are concerned about phishing in MIM apps, (n=27) of the participants said they are quite concerned. At the same time, (n = 23) said they were extremely concerned (see Figure 4-9).

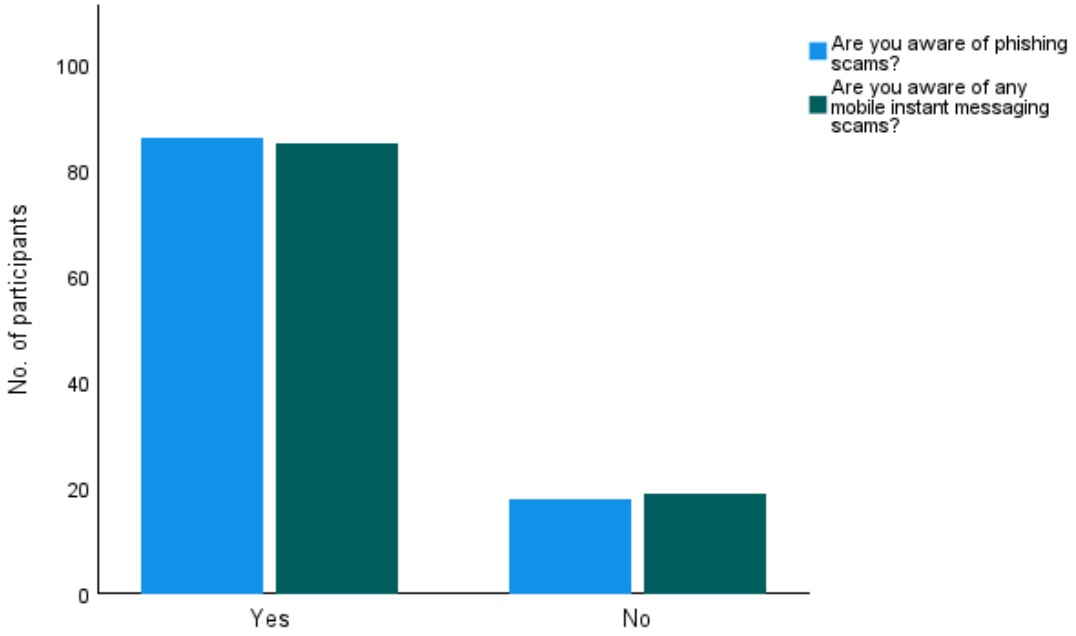


Figure 4-8: Participants' awareness of phishing (n=104)

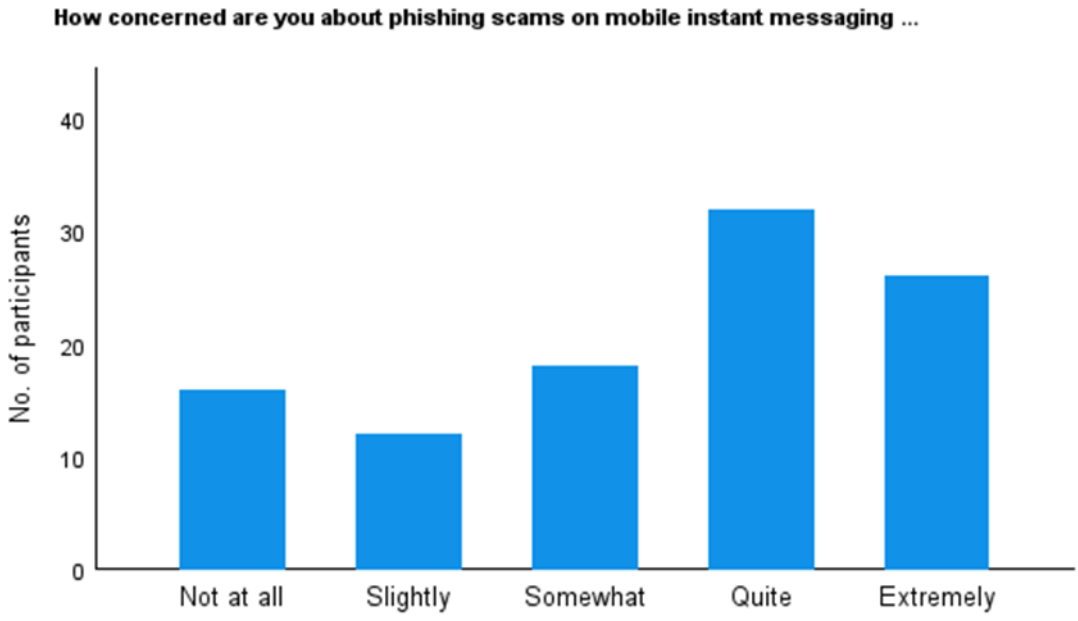


Figure 4-9: Participant concern about phishing in MIM apps (n=85)

4.4.5 How MIM app users respond to phishing attempts

In addition to evaluating participants' self-reported Knowledge, awareness of, and level of concern about phishing, it was also important to examine how they respond to suspicious links during communication in MIM apps. According to Figure 4-10, most participants (n=78) do not click on links from unfamiliar senders, accounting for 75% of the total. Additional choices that the participants favoured included keeping an eye for suspicious links (n=53, 51%), examining the link preview (n=48, 46.2%), and conducting a Google search to gather more information on the website before disclosing personal information (n=43, 41.3%).

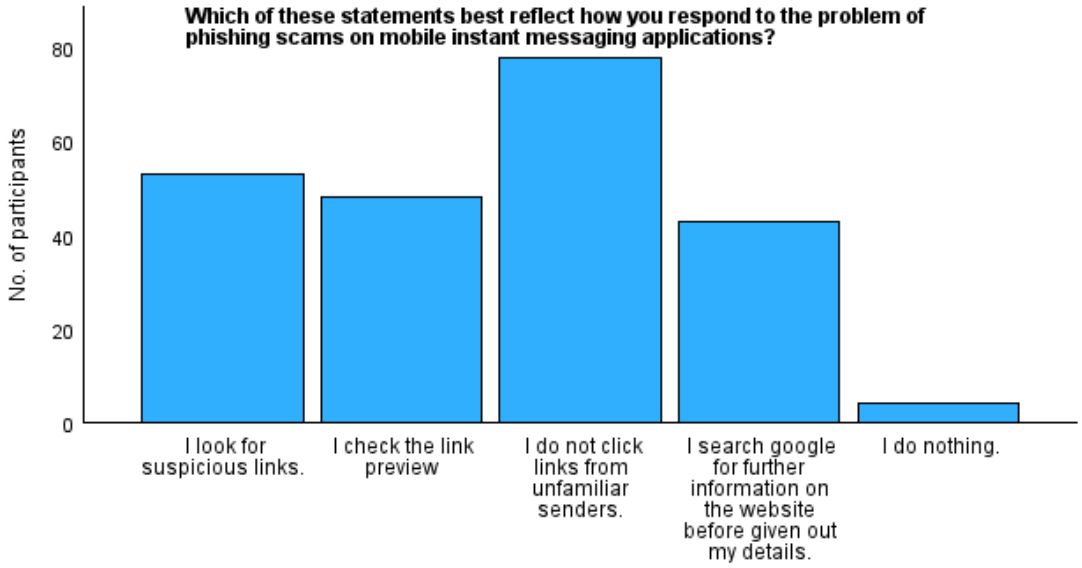


Figure 4-10:How participants respond to suspicious links during instant messaging

4.4.6 Participants' frequency of clicking links during MIM

The participants were asked about their behaviour when clicking links during instant messaging. Specifically, the participants were asked about the general frequency of clicking links and the frequency of clicking links shared in groups.

According to Figure 4-11, most participants occasionally click on links in both conditions. More than half of the participants (n = 52, 51%) reported that they sometimes click on links during communication in MIM apps. In comparison, 47 (46.1%) indicated that they sometimes click on links shared in groups. The data presented in Figure 4-11 suggest a slight disparity between the participants' general link-clicking behaviour and their behaviour for links shared during group communication. However, the Wilcoxon signed-rank test results in a Z-statistic of -0.394 and a p-value of 0.680, implying that this difference is not statistically significant.

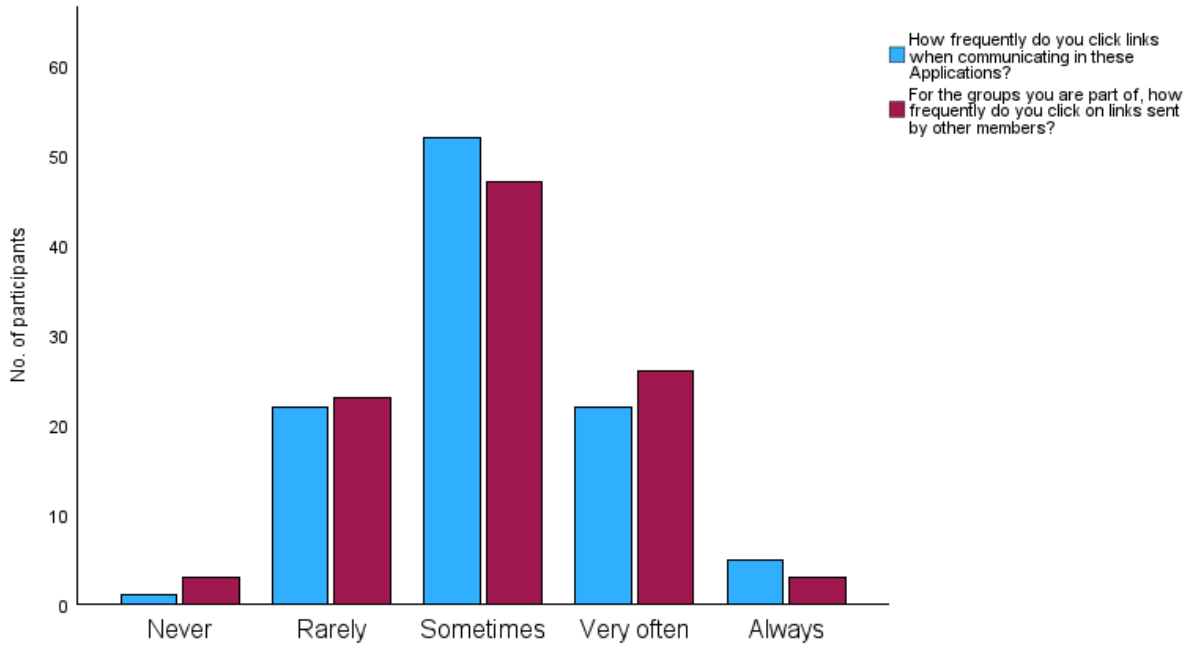


Figure 4-11: Frequency of clicking links by the participants (n=102)

4.4.7 Participants' link-clicking behaviour across different communicating parties

As reported in Figure 4-4, MIM Apps' users communicate with friends, family, work/business colleagues, acquaintances, and strangers. However, it is unknown whether the self-reported link-clicking behaviour of users across these communicating parties is different. Therefore, we asked the participants to indicate how frequently they clicked on links from these types of users within their social cycles. The survey employed display logic to ensure that questions related to communicating parties were only displayed to participants if they had indicated in the first part of the survey that they used MIM apps to communicate with them. Following this approach, the data revealed five groups of participants. The first group (n=10) stated that they utilised MIM apps to communicate with all categories of communicating parties. The second group (n=32) utilised MIM apps to communicate with four categories of communicating parties. The third group (n = 31) reported using MIM apps to communicate with three types of parties. The fourth group reported communicating with only two (n=21), and the fifth and final group reported communicating with only one communicating party (n=17).

Since the click frequency question focuses on finding the difference in user click behaviour based on two or more conditions, the analysis focused only on participants who communicate with more than one category of users.

Table 4-3 presents the click frequency for participants who converse with only two people. It shows that those who use MIM apps to chat with family and friends have the highest number of instances. Further analysis reveals only a slight difference in the frequency with which participants click on links received from friends or family (see Figure 4-12). Most participants indicated that they sometimes click on links from family (Mdn = 3) and friends (Mdn = 3). This similarity was confirmed by a Wilcoxon signed-rank test with a Z statistic of .00

and a p-value of 1.000, indicating no significant difference in the participants' self-reported behaviour in the two conditions.

Table 4-3:Different combinations of communicating parties for participants that selected two types of people

Selected communicating parties	Frequencies
Friends and Family	13
friends and work/business colleagues	3
Friends and known but not friends, family or business/work colleagues	1
Family and work business colleagues	2
Work/business colleagues and strangers	1
known but not friends, family or business colleagues	1

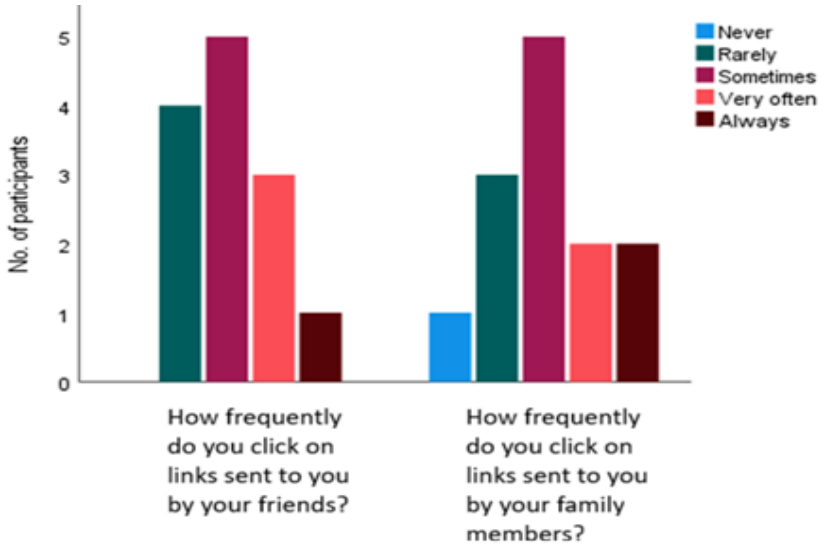


Figure 4-12: Frequency of clicking links by the participants (n=13)

Table 4-4 presents the combinations for participants who selected three communicating parties, along with their corresponding frequencies. The analysis focused on individuals who used MIM apps to chat with friends, family, and work/business colleagues, as this group has the highest frequency.

Table 4-4: Different combinations of communicating parties for participants who selected three types of people and their corresponding frequencies.

Selected communicating parties	Frequencies
Friends, family, and work/business colleagues	28
Friends, family and contacts that aren't their friends, family or business colleagues	2

Figure 4-13 shows that the participant frequently clicked links from the three communicating parties, as indicated by the median value (Mdn=4). Links from family and friends were also highly rated by many participants, as evidenced by the selection of the “always” option. The similarity in the participants' ratings across the three communicating parties was confirmed using the Friedman test, which showed no significant difference in the median ratings across the three categories, $\chi^2(2) = 2.655$, $p = .265$.

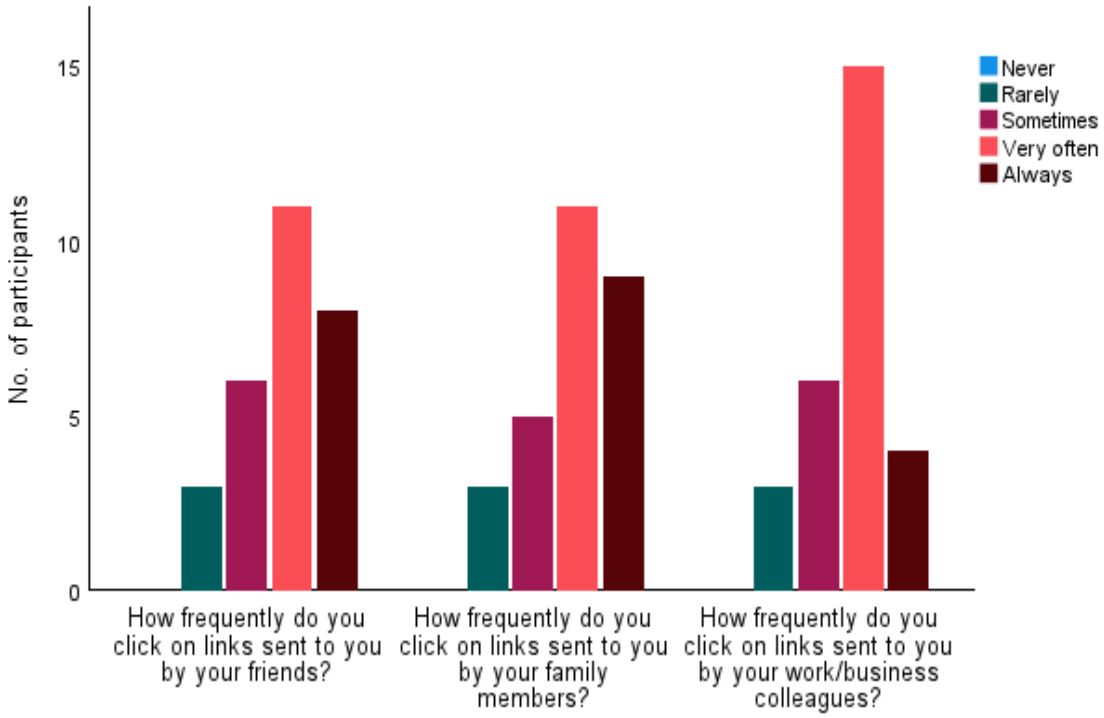


Figure 4-13: Frequency of clicking links from friends, family and work/business colleagues (n=28)

Table 4-5 displays the various combinations of participants who chose four categories of communicating parties. Most individuals in this group often select acquaintances, such as family members, friends, colleagues from work or the company, and others who are not part of their immediate circle of friends, family, or work or business colleagues. Additional examination indicated that most participants rarely (n=18) or never (n=6) engaged with links from contacts who are not part of their immediate circle of friends, family, or work or business colleagues, as shown in Figure 4-14. The highest frequency ratings were received by clicking links from family and work/business colleagues, with a median of (Mdn =4). This was followed by friends, with a median of (Mdn=3.50). The frequency of clicking links from contacts who are not family, friends, or work/business colleagues is rated the lowest, with a median score of 2. A Friedman test revealed a statistically significant disparity in the median ratings

among the four communicating parties, with a chi-square, χ^2 (3) =59.416, and p<.001. Subsequent pairwise comparisons using the Dunn-Bonferroni post hoc test revealed a statistically significant difference in the frequency of clicking links from family members, friends, colleagues from work or business and individuals who are not friends, family, or work/business colleagues (p<.001).

Table 4-5: Different combinations of communicating parties for participants that selected four types of people

Selected communicating parties	Frequencies
Family, friends, work/business colleagues and others they know but not friends, family or work/business colleagues	30
Friends, family, work/business colleagues and strangers	1
Friends, family, others they know but not friends, family or work, business colleagues and strangers	1

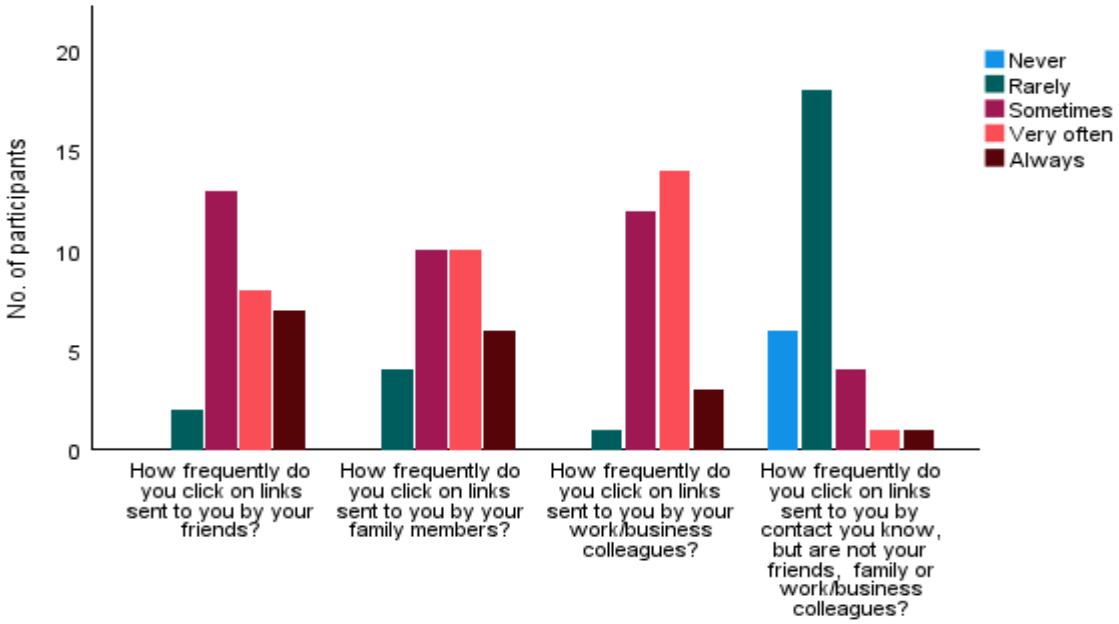


Figure 4-14: Frequency of clicking links from friends, family work/business colleagues, and others known (n=30)

Further analysis shows that 10 participants reported communicating with all five types of contacts. Figure 4-15 shows the frequency at which the participants click on links from these communicating parties. The figure suggests that the frequency of clicking links from friends, family, and work/business colleagues received the highest ratings, with a (Mdn=3.50). At the same time, the participants rated other communicating parties as lower. Specifically, the unknown contacts received the lowest ratings (Mdn = 2.0). A Friedman test revealed a statistically significant difference in the median ratings among the four communicating parties, with a chi-square $\chi^2(4) = 25.638$ and $p < .001$. A subsequent paired analysis using the Dunn-Bonferroni post hoc test reveals a statistically significant difference in the frequency of clicking links from work/business colleagues and those from unfamiliar contacts (p = .007).

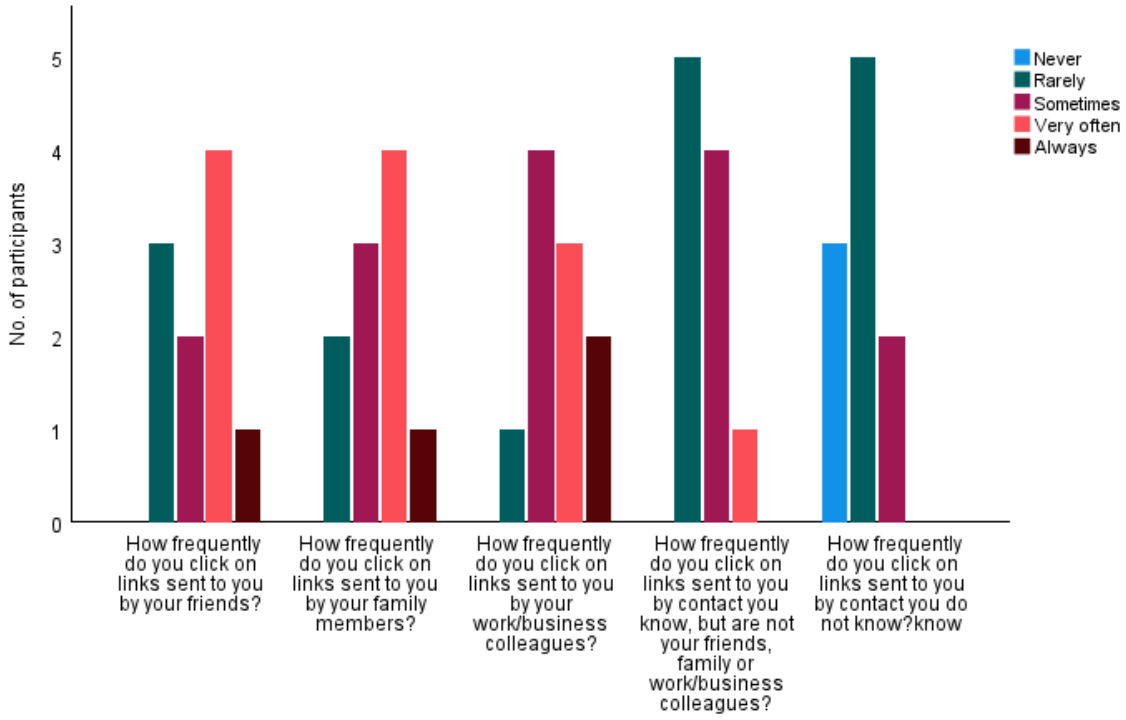


Figure 4-15: Participants' frequency of clicking links across five types of communicating parties (n=10)

4.4.8 Frequency of clicking of links in groups shared by various communicating parties

The study also analysed the participants' link-clicking behaviour across the different categories of people during group-based communication. The analysis revealed that participants who communicate with two types of contacts mostly do so with friends and family, as shown in Tables 4-6. Further analysis reveals that the rating levels for the participant's frequency of clicking links across the two conditions were above the midpoint of the scale (Mdn = 3), indicating that most participants sometimes, very often, or always click on links from their friends or family. Similar to the findings for general click frequency, the Wilcoxon signed-rank test outcome revealed no significant difference in the participants' ratings across these conditions, with a Z-statistic of 0.447 and a p-value of 0.655.

Table 4-6:Different combinations of communicating parties for participants that selected two types of contacts during group-based communication

Selected communicating parties	Frequencies
Friends and Family	11
friends and work/business colleagues	3
Friends and Unknown	3
Work/business colleagues and strangers	1
known but not friends, family or business colleagues and strangers	1

For the participants who indicated they communicate with three categories of people (n=31), only (n=28) responded to the questions on how frequently they click links from these categories during group-based communication. As shown in Table 4-7, the most commonly selected groups are friends, family, and

work/business (n = 27). Further analysis shows that the rating levels for the participant's frequency of clicking links across the three conditions were above the midpoint of the 5-point Likert scale. Specifically, clicking links from friends has a median of (Mdn=3), family (Mdn=4), and work/business colleagues (Mdn=4). Like the general click frequency, the Friedman test showed no significant difference in the median ratings across the three categories, χ^2 (2) =3.556, p=.169.

Table 4-7:Different combinations of communicating parties for participants that selected three types of people during group-based communication

Selected communicating parties	Frequencies
Friends, family, and work/business colleagues	27
Friends, family and contacts that aren't their friends, family or business colleagues	1

Out of (n=32) participants who communicate with four types of people during MIM communication, only (n=26) responded to questions on the frequency of clicking links across these categories during group-based communication. The most frequent combinations are those that involve communication with friends, family, work/business colleagues, and acquaintances who are neither friends, family, nor business colleagues (n=25). As a result, the analysis was focused on this combination. The participant's rating levels for the frequency of clicking links from family, friends, and work/business colleagues were above the midpoint of a 5-point Likert scale (Mdn = 3), while those for known contacts who are not family, friends, or work/business colleagues were below the midpoint (Mdn = 2). Further analysis using the Friedman test revealed a significant difference in the median ratings across the four communicating parties, $\chi^2(3) = 33.860$, $p < .001$. Follow-up pairwise using the Dunn-Bonferroni post hoc test indicates a significant difference between the frequency of clicking links from friends, family

and the contacts the participants know but not friends, family or business colleagues (p<.001) and from work/business colleagues and the contacts the participants know but not friends, family or business colleagues (p<.026). These findings are consistent with earlier observations regarding general click frequency.

The (n=10) participants who indicated that they communicate with all five types of people also responded to the questions on the frequency of clicking links during group-based communication. The participants sometimes clicked on links sent by these users, resulting in rating levels just above the midpoint of the 5-point Likert scale (Mdn = 3). The Friedman test showed a significant difference in the median ratings across the five communicating parties, $\chi^2(4) = 17.951$, $p < .001$. However, this significance was not noticed when the Dunn-Bonferroni correction adjusted the significance values for multiple tests.

4.4.9 Participants' link forwarding behaviour during MIM

The participants were asked to indicate how frequently they forward links to other users during MIM. Figure 4-16 shows that most of the participants (n=42) indicated that they sometimes forward links to others, as indicated by a median value (Mdn=3), and (n=18) indicated that they very often do so. In contrast, (n=31) said they do so, but infrequently.

Members of MIM app groups can receive messages from other members. However, as highlighted in section 4.3.1, groups can be either private or public. Therefore, the participants were asked how frequently they forward links they receive from public and private groups. The findings in Figure 4-17 suggest that the participants were more likely to forward links they received from private groups, as evidenced by (Mdn= 3). In contrast, those from public groups received a lower rating (Mdn=2). A similar sample test was conducted using the Wilcoxon Signed-Rank test. The results showed a statistically significant difference

between the medians of the two self-reported behaviours, with a z-score of 4.884 and a p-value of less than .001

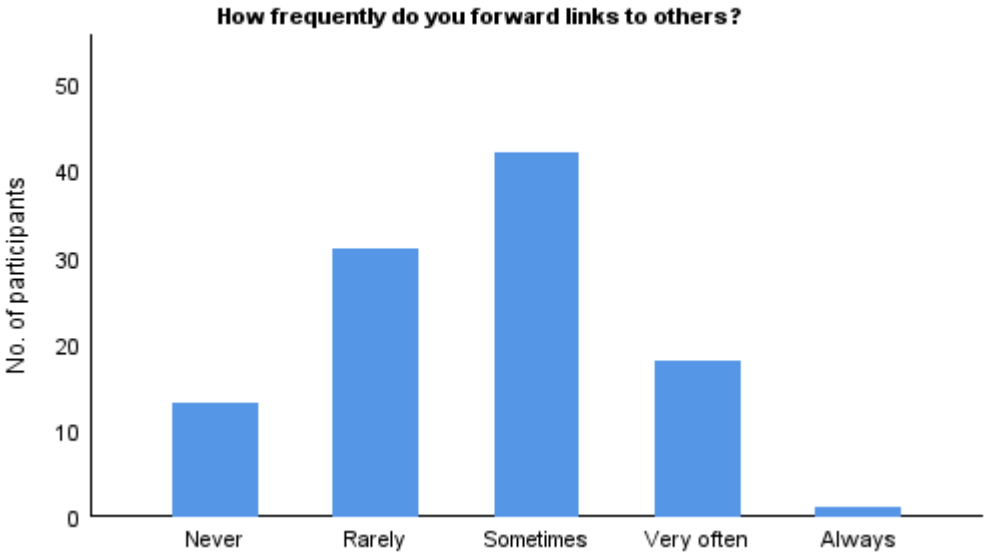


Figure 4-16: Frequency of forwarding links to other users (n=105)

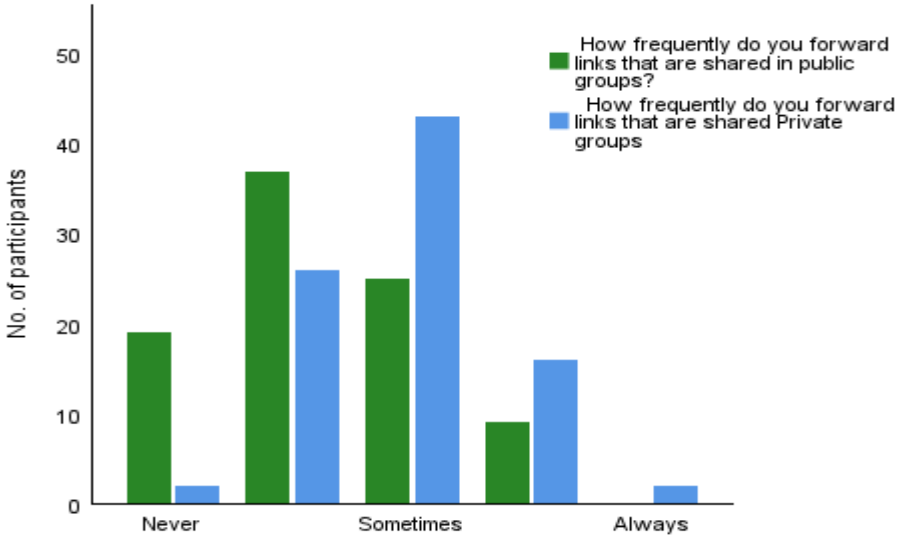


Figure 4-17: Participants' frequency of forwarding links received from groups (n=90)

4.4.10 Participants' phishing precautionary behaviour during instant messaging

The study investigated participants' cautious behaviours against phishing during instant messaging by using questions that assessed specific actions users might take while instant messaging. These practices include 1) verifying the link before clicking, 2) considering the sender of a message before clicking, 3) following the link before forwarding and 4) considering the sender before forwarding. These activities may indicate the extent to which users are aware of online security when instant messaging.

Figure 4-18 illustrates the participants' ratings of the frequency of checking links before clicking on them. The data suggest that most participants frequently or consistently verify links before clicking them. Further analysis revealed that the median of the participants' responses is higher than the midpoint of the scale (Mdn=4). Consequently, these results suggest that participants exercise caution when clicking links during instant messaging. Figure 4-19 shows that most participants (n=71) always consider the sender of a link before clicking it.

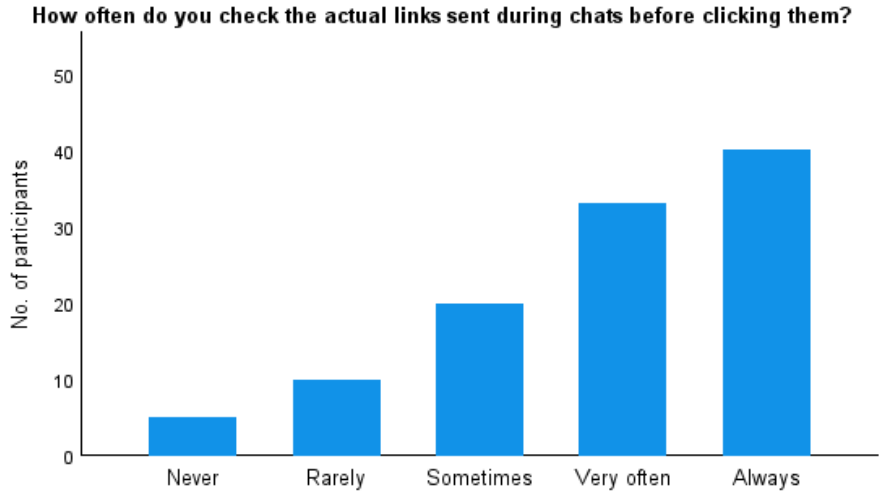


Figure 4-18: Participants' frequency of verifying the link before clicking (n=108)

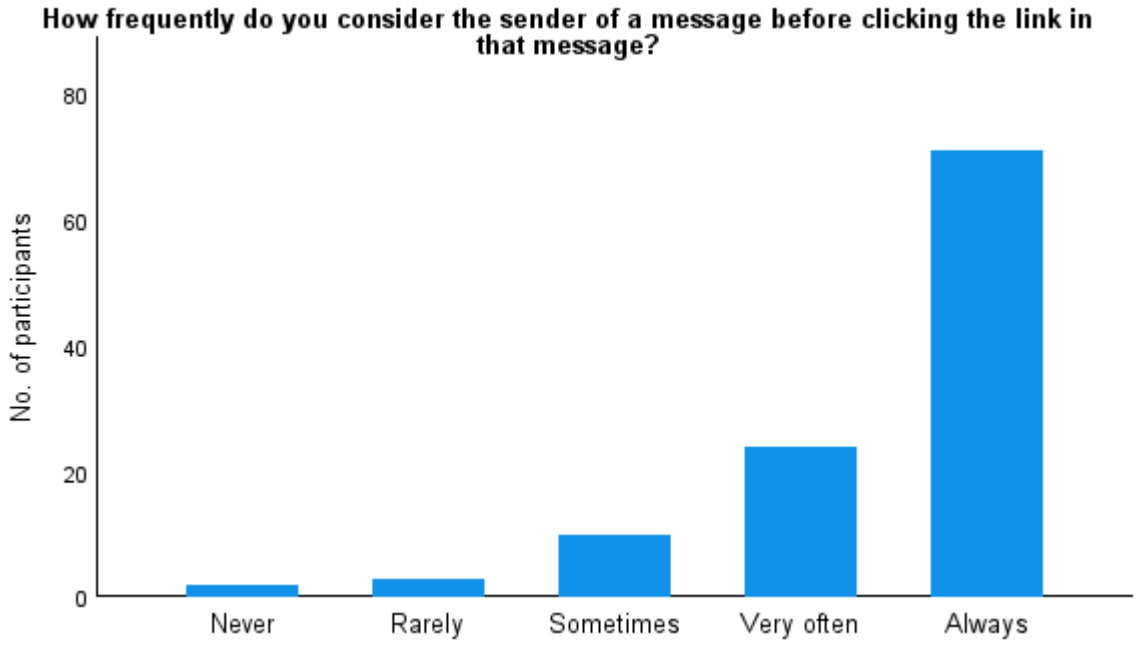


Figure 4-19: Frequency of considering the sender before clicking

In Figure 4-20, the data indicate that most participants (35, 31.5%) reported always following links before sharing them with others. A significant portion (26, 23.4%) stated that they do so very often. Some participants (19, 17.1%) indicated that they occasionally follow links, while others (9, 8.1%) reported seldom doing so.

A minority (3, 2.7%) said they never follow links before sharing. Figure 4-21 shows that most participants (43, 38.7%) always consider the sender of a link before forwarding it to others.

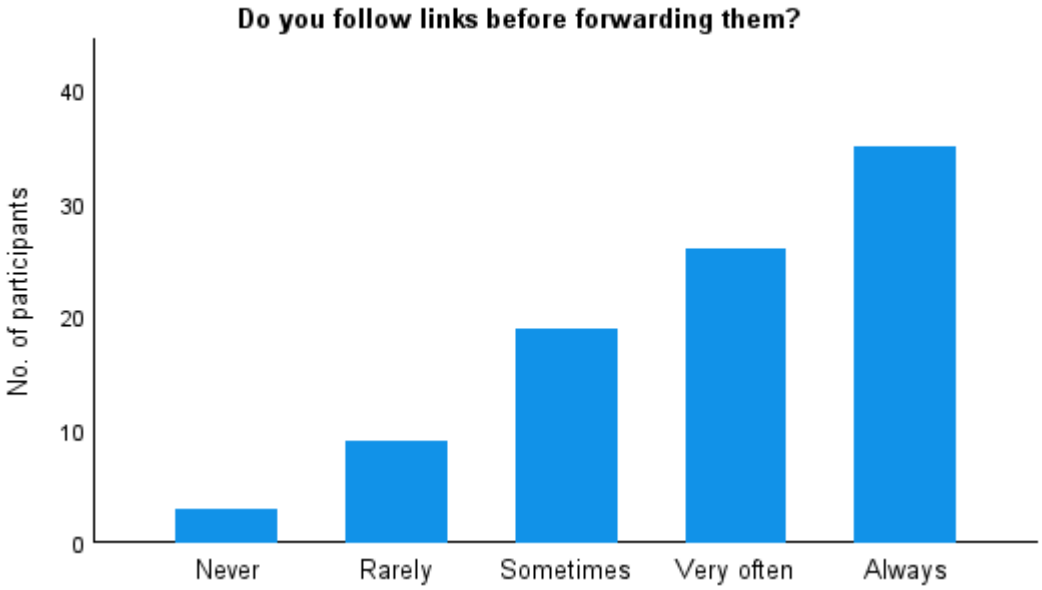


Figure 4-20: Participants' frequency of following links before forwarding them (n=92)

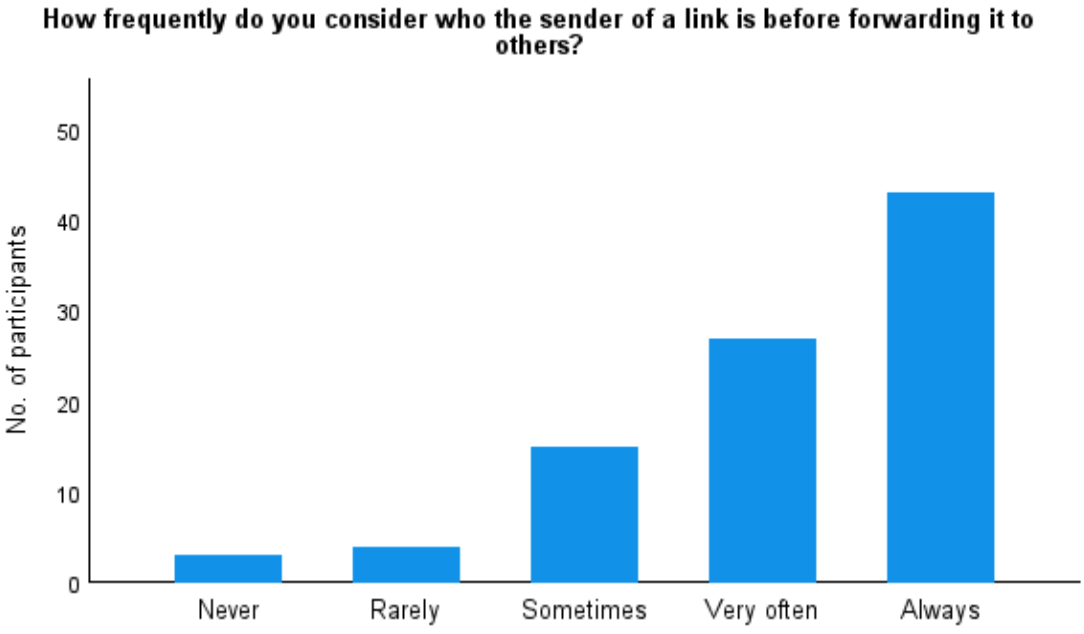


Figure 4-21: Participants' frequency of considering the sender a link before forwarding it (n=92)

4.4.11 Phishing protection

The phishing self-efficacy of the participants was measured using a three-item scale as highlighted in 4.3.1. The internal consistency of the construct was assessed using Cronbach's alpha reliability test. Table 4-8 shows that the Cronbach's alpha coefficient for phishing self-efficacy was 0.86. This value falls within the acceptable range, as the literature recommends a value between 0.6 and 0.7 is required for the internal consistency of a construct's components (Churchill Jr, 1979; Pallant, 2007; Rahimnia and Hassanzadeh, 2013).

Table 4-8: Reliability test for phishing self-efficacy scale

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.865	.866	3

Table 4-9 shows the questions and their corresponding rating scores. The scores from the table demonstrate that the participants rated themselves very highly in terms of their ability to protect themselves from MIM phishing scams.

Table 4-9: Participants' phishing self-efficacy scores

Items	Mean	SD
I believe I have the knowledge and skills to identify phishing URLs when they are presented to me during communication on mobile instant messengers.	4.07	0.908
I believe that I can protect my personal information from being stolen by phishers.	3.80	1.010

I believe I have the skills to implement security measures to stop people from getting my confidential information	3.64	1.064

When asked if they believe in the skills and knowledge of their contacts to protect them from phishing, (8,7.9%) strongly disagree and (24, 23.8%) disagree. The majority of participants (42, 41.6%) selected the neutral option. Some participants (23, 22.8%) agreed, while others (4, 4%) strongly agreed, as shown inFigure 4-22.

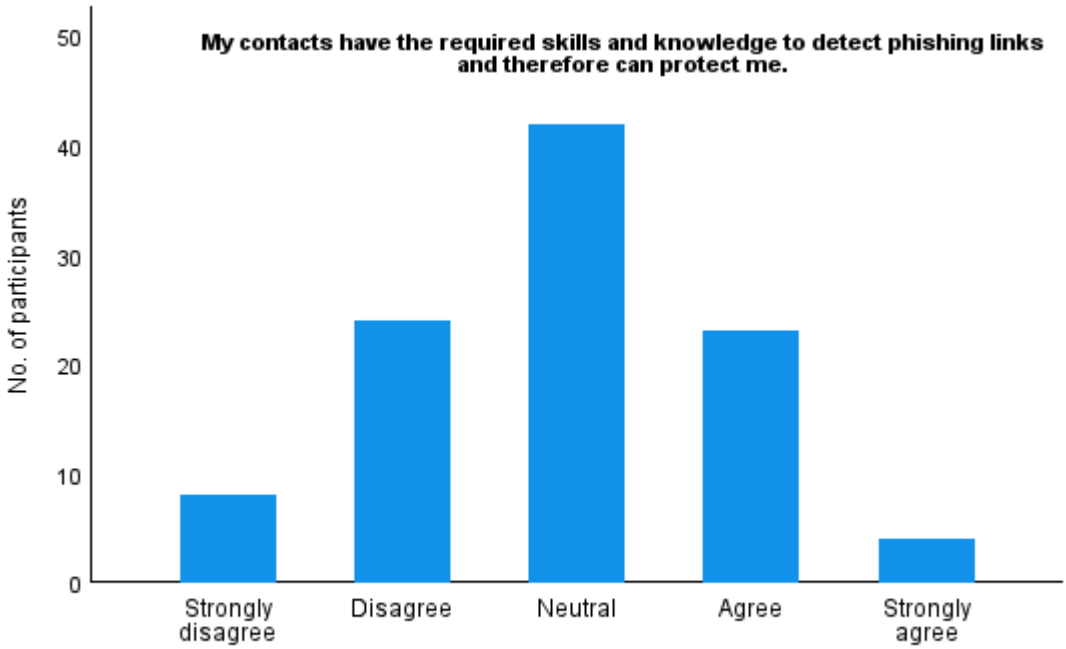


Figure 4-22:Participants believe in the capability of their friends to protect them (n=101)

Similarly, when asked if participants believe mobile instant messengers have mechanisms in place to protect them from opening malicious URLs, (10, 10%) strongly disagree, while (27, 27%) disagree. The majority of participants (36, 36%) selected the neutral option. Some participants (22, 22%) agreed, while others (5, 5%) strongly agreed, as shown in Figure 4-23.

The participants were also asked to indicate who they believe is responsible for protecting users from phishing in MIM apps. Table 4-10 shows that most participants believe the responsibility should be shared between platform providers and users.

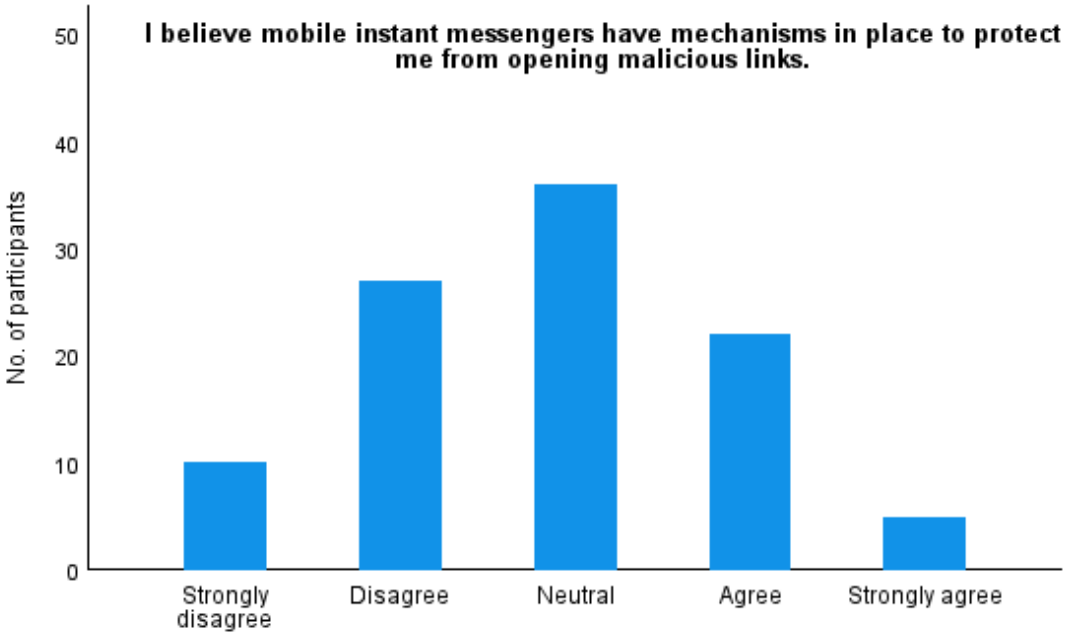


Figure 4-23: Participants believe in the capabilities of MIM apps to protect them (n=100)

Table 4-10: Phishing protection responsibility

Items	Frequency
I believe it is the responsibility of mobile instant messaging platforms to protect their users from phishing scams.	19
I believe mobile instant messengers' users should be responsible for protecting themselves from phishing scams.	10
Protecting users from phishing scams should be a shared responsibility between mobile instant messaging platforms and their users.	68

4.4.12 The correlation between adopting phishing preventive behaviours and phishing self-efficacy, perceived phishing knowledge, and concern over phishing in MIM apps

As highlighted in section 2.2.1, user knowledge, phishing self-efficacy and concern/anxiety about phishing decrease user susceptibility to phishing. In this section, an examination of the relationship between adopting phishing preventive behaviours and phishing self-efficacy, perceived phishing knowledge, and concern over phishing in MIM apps was conducted.

Spearman's rank-order correlation was used to examine whether there is a relationship between the participants' phishing self-efficacy and their self-reported phishing preventive behaviours, specifically considering the sender before clicking, checking the link before clicking, following the link before forwarding, and considering the sender before forwarding links to others. A

composite score was generated for each participant based on the five items of the phishing self-efficacy measure, as this was required for Spearman's rank-order correlation. Table 4-11 shows the relationship between phishing self-efficacy and the participants' self-reported phishing preventive behaviours. The table suggests a statistically significant weak positive correlation between the participants' phishing self-efficacy and the frequency of following a link before forwarding it to others. There was also a moderate correlation between the participants' phishing self-efficacy and the frequency with which they considered the sender before forwarding a link to others.

Table 4-11: Spearman's correlation coefficient (rs) between participants' phishing self-efficacy in MIM Apps and phishing preventive behaviours.

	Consider sender before clicking	Check the link before clicking	Follow the link before forwarding	Consider sender before forwarding
Phishing self-efficacy	($r_s(100)$ = .161, p = .105).	($r_s(100)$ = .185, p = .063).	($r_s(90)$ = .214*, p = .041).	($r_s(90)$= .406**, p <.001).

Figure 4-10 shows the participants' concerns about phishing in MIM Apps, and the findings indicate that most participants were quite or highly concerned about phishing in MIM apps. Further exploration was carried out using Spearman's rank-order correlation to understand if the level of concern someone has is associated with their self-reported phishing preventive behaviours in MIM apps. Table 4-12 shows a statistically significant, weak positive correlation between the participants' level of concern about phishing in MIM apps and the frequency of considering the sender before clicking and checking the link. Furthermore, the table shows a statistically significant moderate positive correlation between the participants' frequency of considering the sender before forwarding the link to others and their level of concern.

Table 4-12: Spearman's correlation coefficient (rs) between participants' concern about phishing in MIM Apps and phishing preventive behaviours.

	Consider the sender before clicking.	Check the link before clicking	Follow the link before forwarding.	Consider the sender before forwarding
Level of concern	($r_s(101)$ = .197*, p = .046).	($r_s(101)$ = .206*, p = .037).	($r_s(99)$ = .101, p = .340).	($r_s(90)$ = .407**, p<.001).

The concluding analysis of this section investigated the correlation between the participants' self-reported familiarity with phishing and their self-reported phishing preventive behaviours. Table 4-13 shows that the participants' perceived understanding of phishing has a statistically significant influence on their self-reported preventive behaviours during instant messaging. The results suggest a statistically significant weak positive correlation between the participants' perceived understanding of phishing and the frequency of considering the sender of a link before clicking, checking the link before clicking and following the link before forwarding. Furthermore, there was a statistically significant moderate positive correlation between the participants' frequency of considering the sender before forwarding the link to others and their reported perceived understanding of phishing.

Table 4-13: Spearman's correlation coefficient (rs) between participants' perceived understanding of phishing and phishing preventive behaviours.

	Consider sender before clicking	Check the link before clicking	Follow the link before forwarding	Consider sender before forwarding
Phishing knowledge	($r_s(100)$ =.376**, P<.001)	($r_s(100)$ =.257*, p=.009)	($r_s(89)$ .217*, p=.038	($r_s(89)$ .408**, p<.001

4.4.13 The correlation between mobile-computing self-efficacy and participants' risk-taking behaviours during MIM apps

Table 4-14 shows the mean scores and standard deviations for each mobile computing self-efficacy construct item. These scores indicate that most of the responses are at the high end of the scale. Furthermore, the internal consistency of the construct was assessed using Cronbach's alpha reliability test. Table 4-15 shows that the Cronbach's alpha coefficient for mobile computing self-efficacy was 0.74.

Table 4-14:Participants' mobile computing self-efficacy

Items	Mean	SD
I believe I have the ability to make purchases using a mobile device.	4.76	0.664
I believe I have the ability to identify and correct common problems with mobile devices.	4.22	0.889
I believe I have the ability to install and remove features and applications from mobile devices.	4.55	0.724
I believe I have the ability to use the productivity features offered by mobile devices (e.g. calendar, email, task scheduling, etc.).	4.66	0.625

Table 4-15: Reliability test for mobile-computing self-efficacy scale

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.738	.750	4

As highlighted in 4.3.1, the mobile-computing self-efficacy was included because prior research has linked high self-efficacy to risk-taking behaviours. Therefore, the relation between the participants' mobile-computing self-efficacy score and the risk-taking behaviour of clicking links from unknown senders was investigated using Spearman's rank-order correlation. Thus, a composite score was generated for each participant based on the four items of the mobile-computing self-efficacy measure, as this was required for Spearman's rank-order correlation.

As highlighted in 4.4.7, participants only saw questions relating to whether they clicked on links from a communicating party if they had indicated in the first part of the survey that they used MIM apps to communicate with that communicating party. As a result, only 14 participants indicated that they communicate with unknown users during one-to-one MIM. In comparison, 13 participants reported doing so during group-based communication. Therefore, the examination of the relation between the participants' mobile-computing self-efficacy score and the risk-taking behaviour of clicking links from unknown senders was based on this number of participants.

A Spearman's rank-order correlation was conducted to assess the relationship between mobile-computing self-efficacy and frequency of clicking links from unknown senders. There was a very weak, non-significant negative correlation between the two variables, $r_s(14) = –.072$, $p = .807$. In contrast, the analysis reveals a statistically significant moderate, negative correlation between mobile computing self-efficacy and the frequency of clicking links from unknown senders during group-based communication, $r_s(13) = –.667$, $p = .013$. This suggests that higher mobile efficacy is associated with a reduced likelihood of engaging in risky clicking behaviour.

4.5 Discussion

This section summarises the findings of this chapter whilst relating them to the prior literature on phishing susceptibility. The section also highlights the study's limitations and implications.

4.5.1 MIM usage, awareness of phishing, and preventive measures

Some general findings of this study are that most participants use WhatsApp for MIM, followed by Telegram. This is not unexpected, as the two applications are currently among the most widely used messaging applications globally (Dixon, 2024). The results also show that most participants used MIM apps to communicate with friends and family, suggesting that this media type is commonly used for informal communication. These results corroborate previous research indicating that instant messaging primarily occurs among friends or family (Sultan, 2014). Most participants were part of groups, which allowed them to participate in a dedicated sequence of conversations. Group communication can be beneficial, as it allows the exchange of useful information between members (Resende *et al.*, 2019). However, the shared information may deviate from the group's theme, sometimes including content that may put the members at risk of being phished (Resende *et al.*, 2019). Thus, members must be cautious of the type of links they click or forward to others. Fortunately, the data show that most participants were aware of MIM phishing (n = 85). However, a considerable number (n = 19) stated that they were unaware of such activities occurring on MIM apps, possibly because phishing in MIM apps is a relatively recent phenomenon. Nevertheless, those who indicated awareness of MIM phishing were found to be very concerned about the prevalence of such cyber threats on MIM apps.

The results also indicated that most participants were cognisant of phishing aimed at them and tried to decrease their vulnerability by adhering to

recommended anti-phishing advice. For example, the findings suggest that most participants (n = 78, 75.0%) refrain from clicking on links sent by unknown senders as a strategy to mitigate their vulnerability to phishing attacks in MIM apps. In contrast, Cain, Edwards and Still (2018) reported in their exploratory study of cyber hygiene behaviours that 96% of participants admitted to clicking on links from unknown senders during email communication. Similarly, the research Shah and Agarwal (2020) revealed that smartphone users in India had inadequate online security practices, such as clicking on links from unfamiliar sources. Nevertheless, the participants of the current study also reported being alert and looking for suspicious links to protect themselves from phishing attempts during MIM. Additionally, the results indicate that some participants conducted further research on the link before handling their personal details. This strategy can be effective since security researchers often write articles and blogs about ongoing phishing campaigns. These findings support and reinforce the results in 4.4.4, providing further evidence that the participants were aware of phishing and actively implementing the suggested precautions to safeguard themselves. The results also revealed that most participants believed protecting users from phishing scams should be a shared responsibility between mobile instant messaging platforms and their users, aligning with the view of many security professionals who emphasise that both social media platforms and their users are responsible for addressing the problem of phishing (Parker and Flowerday, 2020).

Another strategy used by the participants is examining the link preview (n=48, 46.2%). However, Stivala and Pellegrino (2020) have shown that cybercriminals can manipulate the elements of the link preview to direct users to fraudulent pages. Therefore, the effectiveness of this strategy depends on whether the participants are aware of this and can detect this manipulation.

4.5.2 Frequency of clicking and forwarding links

In terms of the frequency with which individuals click on and forward links shared by others during communication in MIM apps. The findings indicate that a significant number of participants engage in the action of clicking on and then sharing links within MIM apps. Some individuals engage in this behaviour not only in one-on-one conversations but also in group communication, as the findings revealed no statistically significant difference in the frequency of clicking on shared links in both scenarios. Although it may be okay to engage in such conduct, it becomes problematic when it becomes a persistent pattern, as this can impact how users process information (Frauenstein and Flowerday, 2016). Moreover, given the prevalence of phishing in MIM apps, individuals who frequently click on links and forward messages are at risk of exposing both themselves and others to phishing attacks. Nevertheless, this study uncovers that participants' self-reported behaviour is more intricate, as they tend to click on links shared by individuals in their social network, such as friends, relatives, or coworkers. This finding is similar to what was observed in a data privacy study conducted by Alqhatani (2021). The author found that users are generally willing to share various forms of data with family, friends, and significant others while demonstrating lower comfort in sharing with other parties. Although Alqhatani's study is not explicitly phishing-related, security and privacy are closely intertwined.

Additionally, this study's findings indicate that individuals are more likely to share links they receive in private rather than in public groups. This is likely because private groups consist of familiar contacts, such as friends, family, and coworkers. Unfortunately, it is concerning that certain users are unaware of the specific types of groups they are utilising, since they may share phishing links they receive from public groups with other users. Although the reasons for engaging in these behaviours remain unknown, trust may be a significant component since the participants unconsciously or openly depend on those

within their social circle to safeguard them against phishing attacks. Nevertheless, it remains uncertain if this dependence is justified. However, an interesting results was noticed when the participants were asked how confident they are in the skills of their contacts to protect them from phishing, as the majority of participants disagreed (37, 37%) whilst (36, 36%) selected the neural option, with only (27, 27%) agreeing that their friends have the skills to protect them. This data suggests that the participants had low confidence in their contacts' ability to protect them. Similar results were observed when participants were asked about their confidence in the platforms' ability to protect them. Perhaps the participants' low rating of their contacts and the platform was due to their having seen these fraudulent activities shared on MIM apps, indicating a deficiency in platform security mechanisms and the proficiency of individuals disseminating them. However, going back to the previous finding where the participants were seen as more likely to click links shared by those in their social cycles, the results indicating the participants didn't trust their contacts may appear as contradicting the assumption made earlier that the participants' high frequency of clicking links from those in their social cycle may suggest a level of trust and confidence in their contacts' ability to protect them. However, it is essential to note that the participants were asked to rate the ability of their contacts. Therefore, the ratings were not specific to those within the participants' social circle.

4.5.3 Factors that Influence Phishing Preventive Behaviours in MIM apps

Prior literature has reported that phishing self-efficacy or individuals' self-confidence in recognising phishing leads to higher motivation for performing anti-phishing behaviour (Sun *et al.*, 2016). Similarly, the more concerned individuals are about becoming victims of phishing, the more likely they are to take precautionary measures against it (Lei, Hu and Hsu, 2023). Therefore, this study examined the correlation between participants' self-efficacy in identifying

phishing attempts and their engagement in preventive actions against phishing. In addition, the study examined the impact of other factors, such as participants' perceived understanding of phishing and their concerns, on the adoption of phishing preventive measures during MIM.

In terms of the participants' phishing self-efficacy, it was found that the participants assessed themselves as having a high level of confidence in their ability to protect themselves from phishing in MIM apps. Further analysis showed that the participants' belief in their ability to identify and protect themselves from phishing attacks significantly influences their self-reporting behaviour of evaluating the sender before sharing a link with others, as a moderate correlation was found between the ratings of individuals' confidence in identifying phishing attempts and how often they thought about the sender before sharing a link with others. There was also a statistically significant, albeit weak, positive correlation between the participants' phishing self-efficacy and the frequency of following a link before forwarding it to others.

While the phishing self-efficacy score did not predict all preventive behaviours, the results are consistent with the findings in (Ge and Love, 2014; Safa *et al.*, 2015; Torten, Reaiche and Boyle, 2018), where self-efficacy was found to have a significant impact on the performance of security-conscious behaviour.

This study examines the correlation between participants' self-reported understanding of phishing and their self-reported phishing preventive behaviour when engaging in mobile instant messaging. Findings from this study revealed that the participants' self-reported familiarity with phishing significantly influenced their phishing preventive behaviours. This finding suggests that individuals who strongly believe they understand the concept of phishing are more likely to engage in the recommended proactive measures to prevent phishing attacks. However, these findings contradict those found in an email phishing susceptibility study by Diaz, Sherman and Joshi (2020). The authors found that individuals who lacked awareness or comprehension of phishing were

less likely to click on phishing links than those who were knowledgeable about phishing. Diaz, Sherman and Joshi (2020) found their conclusion surprising, as they had anticipated that individuals who lack awareness or comprehension of phishing would be more vulnerable to phishing attempts. Nevertheless, the current study's findings demonstrate a correlation between user awareness of cyber threats and the performance of security-conscious behaviour.

Regarding the participants' concern about phishing, the analysis ultimately found that the individual's level of concern has a statistically significant, weak positive correlation with their frequency of considering the sender before clicking and checking the link before doing so. Furthermore, a statistically significant moderate positive correlation was observed between the participants' frequency of considering the sender before forwarding the link to others and their concern about phishing in MIM apps. Thus, the findings suggest that the participants' concern about phishing in MIM Apps is associated with adopting preventive behaviours. This finding indicates that individuals concerned about phishing in MIM apps are more inclined to adopt preventive measures when communicating through these applications. This finding is similar to the results reported in an email study (Wang, Li and Rao, 2017), where the authors found that anxiety and concerns regarding phishing had a significant impact on individuals' intention to adopt phishing prevention measures.

4.6 Limitations

The primary limitation of this study is its relatively small sample size. Furthermore, snowball sampling, although effective, often leads to the inclusion of individuals with greater interconnectedness compared to what would be observed in the real world. This may have resulted in the introduction of certain biases in the sample being studied. Moreover, the survey respondents possess a high level of education, as indicated by the fact that most have completed either an undergraduate or postgraduate degree. This educational background could potentially influence the observed behaviours. Additionally, most participants

are male. Yet, the impact of gender on user security behaviour has shown varying results in the literature (Zhuo *et al.*, 2023).

Additionally, the study sample appears to be young, with only a few individuals above 46 years old. However, this demographic distribution may be attributed to the greater prevalence of MIM app usage among younger individuals. Most participants are UK residents, with many more residing in Nigeria, which potentially could have influenced the reported behaviours. Hence, extrapolating the findings of this study to the extensive and varied user base of the MIM app carries certain risks. Ultimately, this study is based on self-reported behaviours, meaning the data may not accurately represent how users interact with links in real-world situations.

4.7 Chapter summary and conclusion

This chapter presented the findings of an exploratory study that used a survey to investigate 1) MIM users' awareness of phishing scams and the preventive measures they use to protect themselves from such scams, 2) MIM app users' frequency of clicking and forwarding links during MIM and 3) factors that may impact MIM app users' behaviours towards links shared in MIM apps. The findings of this chapter revealed that the participants are aware of phishing in MIM apps and expressed concerns regarding the widespread occurrence of such fraudulent schemes. The findings highlight that the study participants try to decrease their vulnerability to phishing by adhering to recommended anti-phishing advice. However, the participants appeared to rely on the link preview to check the legitimacy of links before clicking. However, this self-reported behaviour increases the participants' risk of being phished since empirical evidence from the literature already shows that phishers can manipulate the elements of the link preview to direct users to fraudulent pages.

The findings of this study also show that the participants frequently click on links shared by others and tend to forward such links further, exposing themselves and

others to the risk of phishing. Furthermore, the participants reported having engaged in this behaviour both during one-on-one and group-based communication. This behaviour can potentially increase the exposure to phishing attacks. Nevertheless, the self-reported data suggest that participants try to reduce their susceptibility to phishing attacks by clicking on links shared by individuals in their social circles, such as friends, family or co-workers. Additionally, they have a lower probability of sharing links in public groups than in private ones. Although the underlying causes of these behaviours remain uncertain, trust is likely to play a significant role. Participants may implicitly or explicitly rely on members of their social circle to safeguard them against phishing attempts. However, Phishers can exploit interpersonal trust within social networks, and empirical evidence suggests that, in the context of phishing emails, trust can increase users' susceptibility to phishing attacks.

Moreover, the results indicate that individuals' confidence in identifying phishing attempts, understanding of phishing, and concern about phishing in MIM apps significantly influenced their engagement in precautionary actions against phishing.

The findings of this chapter support the argument presented in Section 2.2.2, which states that MIM apps have features that can potentially increase users' susceptibility to phishing attacks. This study shows that the way these features are used exacerbates the problem. Furthermore, despite the participants' belief in their ability to detect phishing, some of their behaviours and protective measures are not very effective. Thus, the study in this chapter has identified aspects of concern because they have the potential to increase susceptibility to phishing in MIM apps.

Chapter 5

Content analysis of mobile instant messaging phishing attacks

Participants' self-reported data from Chapter 4 show that the behaviours of MIM app users may expose them to security risks such as phishing. To further our understanding of phishing in MIM apps, this chapter presents a study that examines the strategies employed by cybercriminals to phish users of MIM apps, with a particular focus on the persuasion principles used and the URL construction techniques applied. To this end, this study conducted a deductive content analysis of real-world MIM phishing attacks. First, the study is briefly introduced, and the research questions it aims to answer are outlined. Second, the method used to collect and analyse the data was discussed. Third, the chapter presents and discusses the study findings, outlining the limitations. Finally, the chapter concludes the study.

5.1 Introduction

Addressing the problem of phishing in any media requires an understanding of the characteristics of phishing messages in such media. As highlighted in Section 2.2.3, the literature currently suggests that the perpetrators of phishing employ different persuasion principles depending on the attack channel, with differences found across various communication media. Yet there have not been any studies in the literature on which persuasion principles phishers employ during MIM phishing campaigns. Phishers may employ different persuasion techniques when carrying out MIM phishing, as opposed to other attack channels, due to the

medium's unique features, which are highlighted in Section 2.2.2. Furthermore, section 2.2.5 has highlighted that phishers use URL obfuscation techniques during phishing attacks. However, the limited screen size of mobile devices and the use of URL previews may influence how phishers construct deceptive URLs during MIM phishing. In terms of the preview, the key concern, as highlighted in section 2.2.5, is that some of the key components of the preview can be manipulated to deceive users into thinking they are interacting with legitimate pages. This approach can impact how phishers construct URLs during MIM phishing. For example, the phisher can include the details of the impersonated organisation in the title, site description, Image/logo, and domain name of the link preview, but use URL obfuscation techniques, such as TinyURL, to hide the actual URL from the users. This means users relying on the preview as a proof of legitimacy can easily be deceived into thinking they are interacting with a legitimate URL. Thus, we need to understand how the existence of a preview affects the URL manipulation techniques that phishers use during MIM phishing.

The study in this chapter aims to investigate the tactics and techniques employed by cybercriminals during MIM phishing, specifically focusing on the persuasion and URL construction techniques used. This will enable researchers and practitioners to gain a clear understanding of the problem, facilitating the development of technical and user awareness approaches to protect users.

Following this, the study in this chapter aims to answer the following question.

RQ2: What are the tactics and techniques used in the content of phishing messages targeting the users of MIM apps?

RQ2 was further refined into the following questions:

RQ 2.1. What persuasion principles are employed by phishers in MIM phishing, and what was the frequency of their usage?

RQ 2.2 Do phishers combine multiple persuasion principles in MIM phishing messages?

RQ 2.3. How are the different principles of persuasion implemented?

RQ 2.4. What URL construction methods do phishers employ to create deceptive URLs during MIM phishing?

RQ 2.5. Does the prevalence of URL construction methods vary according to the type of organisation being impersonated?

The rationale for these questions was due to the related work from Chapter 2. Specifically, sub-questions one, two and three were formulated based on the findings from section 2.3.3, which highlighted that the perpetrators of phishing employ different persuasion principles depending on the attack channel. These questions were also motivated by similar research from the literature. Specifically, related work on vishing attacks (Jones et al., 2020) and email phishing (Akdemir and Yenal, 2021). This approach will allow for comparison of the findings of the current study with existing literature. Sub-questions 4-5 were formulated due to the findings from the existing literature, as highlighted in section 2.3.5, which shows that phishers use different URL obfuscation techniques during phishing attacks to either enhance source credibility by deceiving users into believing they are interacting with a legitimate entity or make URL parsing difficult for users. Sub-questions three and five were therefore introduced to examine which URL construction techniques were used during MIM phishing

5.2 Materials and Procedures

This study utilised qualitative content analysis (QCA) to investigate the strategies employed by cybercriminals to defraud users of MIM apps through phishing campaigns. As Chapter 3, section 3.2.3 points out, QCA is appropriate for

situations where little is known about the phenomenon under investigation or where it is investigated in its natural state, primarily from a naturalistic viewpoint. Adopting the guidelines for conducting QCA (McKibben *et al.*, 2020), the following steps were utilised to conduct the study.

5.2.1 Data collection

Similar to other qualitative research methods, QCA requires the sample to be systematically selected (Krippendorff, 2018). Sampling methods that can be used to choose units systematically for analysis include convenience, purposive, quota, stratified and multistage (McKibben *et al.*, 2020). Purposive sampling involves the deliberate selection of units based on their specific qualities (Etikan, Musa and Alkassim, 2016). Purposive sampling can mainly be applied when the researcher wishes to target a particular communication medium (McKibben *et al.*, 2020). In this research, units were selected based on the purposive sampling method. This choice was made because the current study targets MIM as a communication medium and phishing messages targeting users of such a medium. The sample used for this study consists of images of phishing messages that were sent to users of MIM apps via the platforms.

The initial stage involved gathering images of real-world MIM phishing attacks, adopting a similar methodology used by Jones *et al.* (2020) and Akdemir and Yenal (2021). Similar to Akdemir and Yenal (2021), this study utilised the Google search engine to collect images of MIM phishing messages posted on different websites by individuals and organisations. The target images were those of MIM phishing targeting the users of Telegram, WhatsApp, and Viber. These platforms were targeted due to their high susceptibility to being targeted by malicious actors (Kaspersky, 2021).

Data collection took place between August 8 and August 15, 2022, using a Google search that covered the period between January 1, 2018, and August 15, 2022. This period was chosen to obtain up-to-date data on the growth of MIM phishing

since the incidence began to appear in the literature in 2018 (Murashka, 2018). The data collection started when the researcher navigated to the Google website and selected the image search section. The researcher then used the phrases "phishing messages", "phishing frauds", and "phishing scams" in combination with either "WhatsApp", "Telegram", or "Viber" in the search bar to search for images of MIM phishing attacks. The researcher utilised the advanced search operators "Before" and "After" in Google's image search tool to search for phishing images posted online within the preceding year. As an illustration, the search commenced by inputting the query "phishing scams WhatsApp After: 2018-01-01 before 2019-12-01" to retrieve all images that were posted online between the dates of "2018-01-01" and "2019-12-01". This method was employed for successive years until August 15, 2022, when the data collection process commenced. Each search result resulted in a large number of images, including legitimate images. As a result, the researcher had to manually scan each search result looking for phishing messages targeting the users of Telegram, WhatsApp, and Viber. The researcher conducted an exhaustive review of the Google Images search results, such that every image returned by the query was examined for inclusion. The image selection criteria required that the phishing messages be written in English and exhibit clear indicators of intended fraudulent activity. The researcher applied domain-specific knowledge of phishing to evaluate both the associated URLs and message content to determine whether the materials met the criteria for classification as phishing attacks. The selection process focused on the content of the images and did not consider the context of the page on which the images were posted. In addition, only messages with links were chosen since this study focuses on phishing rather than other types of social engineering fraud. Moreover, renowned cybersecurity agencies such as the NCSC agree that links or attachments are crucial elements of phishing (NCSC, 2018). This methodology yielded 67 images depicting MIM phishing attacks. This number is similar to what was used in (Jones *et al.*, 2020).

Furthermore, five messages were excluded from the analysis of URL construction techniques due to incomplete URLs or URLs redirecting to legitimate file hosting platforms. Specifically, one message was excluded because it contained links to an APK file, and another was excluded because it contained a link to Google Docs. Two messages were excluded because they contained links to WhatsApp groups. Another message was excluded because the URL was masked and therefore was not viewable (n=1). Therefore, 62 messages were used to answer RQ2.4 and RQ2.5. Appendix B.3 provides access to the dataset used in this study, which is now publicly available through the University of Strathclyde's KnowledgeBase portal.

5.2.2 Development of coding scheme and data coding

This stage aimed to identify the tactics and techniques used by cybercriminals to make MIM app users vulnerable to phishing attacks. The objective was to examine the collected messages to determine the existence of Cialdini's six persuasion principles and the utilisation of URL obfuscation techniques. Therefore, following the data collection, the subsequent stage analyses and categorises the data based on the categories of interest. As highlighted in 3.2.3, the study in this chapter used QCA to answer the research questions. Also, it was highlighted in section 3.2 that QCA can be either inductive or deductive.

A deductive approach was adopted because prior research existed on the techniques and tactics used by cybercriminals during phishing on other media. In line with this approach, a codebook was developed using categories derived from prior research (Akdemir and Yenal, 2021). Specifically, the study used the six categories from Cialdini's framework of social influence (Cialdini, 2006) and the URL obfuscation techniques identified by (Lin *et al.*, 2011; Fernando and Arachchilage, 2020). The codebook was created based on guidelines from (Lin *et al.*, 2011; Ferreira and Jakobsson, 2016; Jones *et al.*, 2020; Akdemir and Yenal, 2021). The codebook includes category names, definitions/explanations, and text

examples, as recommended by (McKibben *et al.*, 2020). This approach helps researchers achieve internal validity of their coding schemes and ensures interrater reliability. The codebook had 27 statements, each explicitly formulated to encapsulate a distinct principle of persuasion or URL obfuscation technique. The complete codebook can be found in Appendix B.1**.** However, some of the codes used are presented in Table 5-1. A coding sheet was also created. The coders used the sheet to label the images according to the categories in the codebook.

Table 5-1: Codebook

Category Name	Description	Example
Authority	Words expressing authority	”your account will be blocked.” (Ferreira and Lenzini, 2015)
Commitment and Consistency	Leverage an individual’s previous commitment or consistency with a previous situation	“As you have done before, please update your password using the following link ”(Wright *et al.*, 2014)
Liking	Content seems to come from those the recipient likes (i.e., friends or those they share similarities with, such as group members, same gender, etc.)	“a scam email from a supposed friend of the recipient asking him/her to visit a website that is really interesting.”(Ferreira and Jakobsson, 2016)
Reciprocity	The content shows the recipient has to perform an action in response to a favour.	“If you respond, you will have a chance to upgrade to our new version.”(Akdemir and Yenal, 2021)
Scarcity	The content of the message shows that the recipient can miss out on offers by not acting immediately (Ferreira and Lenzini, 2015)	“You can save 20% on your next electric bill by filling out our online survey within the next three days.”(Oliveira *et al.*, 2017)

Social proof	The content shows that other users have acted in this manner in the past (Jones *et al.*, 2020)	"I would like to invite you to join the thousands of other clients who have experienced our vacation excursions" (Oliveira *et al.*, 2017)
Similar Domains	The URL contains a domain that is similar to known legitimate addresses	"amazon.checkingoutbooksonline.ca" (Lin *et al.*, 2011)
IP-address	The URL displays an IP address rather than the domain name	http//192.168.111.112/login (Lin et al., 2011)
Letter substitution	One or more letters/characters in the domain are substituted with one or more characters that look similar to those in actual domains	www[.]uca1gary[.]ca rather than www[.]ucalgary[.]ca (Lin et al., 2011)
Complex URLs	The length of the URL is extended with illogical characters making it difficult to read	http://www.amazon.ca.checkingoutbookonline.ca/Golden-Mean-Annabel-Lyon/ (Lin et al., 2011)
Tiny URL	The URL is presented as a tiny URL. Adapted from (Sameen, Han and Hwang, 2020)	

5.2.3 Pilot Testing

A pilot test was conducted to clarify both the codebook and the coding sheet. Two coders participated in the study. The researcher and a second coder, a native English speaker employed in the Cybersecurity sector at the time of the analysis, took part in the pilot process. The pilot aimed to ensure that all coders consistently interpret and categorise the images (McKibben *et al.*, 2020). The researcher randomly selected 15% of the 67 images for the piloting. The author

and the external coder code the data together while noting inconsistencies, discussing, and reaching a consensus. This approach ensured that both coders had the same interpretations of the codebook.

5.2.4 Main Coding

After the pilot test was completed, the main coding phase was initiated. During this phase, the coders categorised the images one by one. However, unlike the pilot stage, coding was done independently. However, the coders regularly came together to assess interrater reliability and agree on inconsistencies. The coders analysed each statement in the codebook to evaluate its relevance for every case of phishing. Suppose a statement assigned to a principle or a URL obfuscation technique is considered true. In that case, the related principle or technique is noted as present in the phishing example. The value of "1" was used to denote the existence of a principle or a URL obfuscation technique, while "0" denoted its nonexistence.

Cohen's kappa was used to test inter-coder reliability. The reliability among coders was significant, with a coefficient of 0.76, which is considered moderate (McHugh, 2012). According to the literature, any Kappa value above 0.60 indicates adequate agreement among the raters. As a result, a code set was chosen at random to be used as the final code for analysis.

Furthermore, each phishing message was examined to identify the companies impersonated and classify them based on their respective industries. The companies were classified according to the categorisations outlined by Zielinska *et al.* (2016) and the Global Industry Classification Standard, developed by MSCI and S&P Dow Jones in 1999 (O'Connell and Curry, 2022). The data was analysed using Python. Due to the utilisation of publicly accessible data, ethical approval was deemed unnecessary for this study.

5.3 Results

It is recommended that the results emanating from the deductive coding process with numerically defined values be reported as frequency counts and percentages (McKibben *et al.*, 2020). This approach was adopted in this thesis and can be seen in related studies (Jones *et al.*, 2020; Akdemir and Yenal, 2021). The descriptive data tells the reader how frequently a particular category appears in the sample.

5.3.1 Initial Analysis

Out of the 67 phishing messages examined, it was discovered that 29 of them, accounting for 43.3% of the total, contained a logo as part of a link preview. Approximately 32 messages (47.8%) included a modified link preview, which accounts for almost half of the total. These modified link previews contained names and logos of the impersonated organisations. Figure 5-1 shows an example of a manipulated link preview that was seen. Upon initial examination, it was found that (n=52, 77.6%) of the phishing attacks propagated using WhatsApp, followed by Telegram users (n=14, 20.9%), followed by Viber (n=1, 1.5%). Further analysis reveals that most phishing messages were related to giveaways and rewards (n=44, 65.7%). These messages were mainly characterised by free vouchers and rewards from very prominent retailers. Other notable topics found in messages were customer support (n=4, 5.97%), security, and mobile application updates (n=4, 5.97%), job offer (n=2, 2.99%) and others (14.93%). The customer support and security updates were targeting cryptocurrency customers. An analysis of the prominent words in the sample reveals that the top five words were "free," "get," "click," "given," and "anniversary," as depicted in Table 5-2.

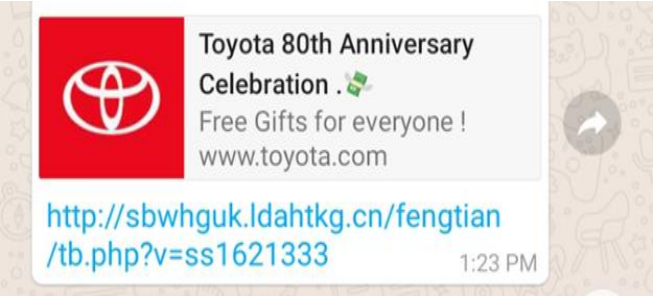


Figure 5-1: An example of a manipulated link preview

Table 5-2: Most prominent used words MIM phishing

Word	Frequency
Free	35
Get	26
Click	17
Giving	16
Anniversary	16
Link	16
Away	15
Please	15
Wallet	15
Account	13
Send	13
Ticket	10

5.3.2 Main findings

The first part of the main analysis examined the categories of companies impersonated during the phishing attacks. The analysis revealed that companies in the retail sector were the most impersonated, with a total of 15 instances, followed by fintech with 13 cases. Table 5-3 displays the classifications of companies most impersonated and their respective frequencies.

Table 5-3: Categories of companies impersonated and their frequency

Company Category	Frequency	Percentage
Retail	15	24.2%
Fintech	13	21.0%
Food and beverages	5	8.1%
Government	5	8.1%
Software	5	8.1%
Transportation and logistics	5	8.1%
Hotels, restaurants and leisure	3	4.8%
Household products	2	3.2%
Interactive media and Internet entertainment	2	3.2%
Automobiles	1	1.6%
Footwear and apparel	1	1.6%

RQ 2.1. What persuasion principles were employed by phishers in MIM phishing, and what was the frequency of their usage?

Figure 5-2 shows that phishers utilised the social proof principle very frequently. The social proof principle was seen in (n = 51, 76.1%) of phishing messages. It was followed by liking, which was employed in 50 cases (74.6%). The authority principle was used in (n = 37, 55.2%) of the messages, while scarcity was employed in (n = 23, 34.3%). Reciprocity was utilised in (n = 11, 16.4%), while commitment and consistency were only employed in (n=1, 1.4%) of the messages (See figure 5.2). Inspection of Figure 5-2 suggests that cybercriminals used certain persuasion principles more than others.

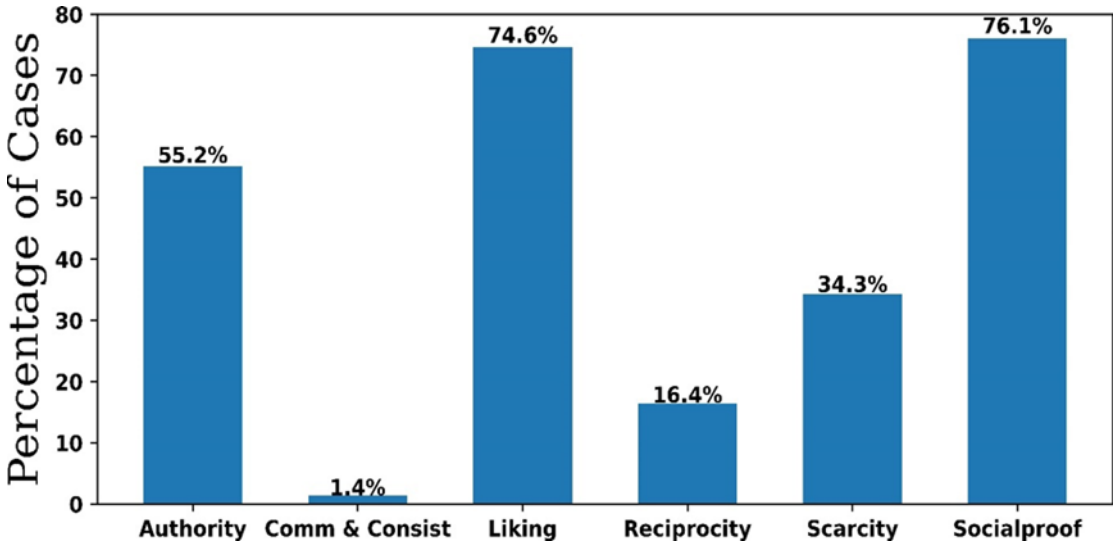


Figure 5-2: Overall persuasion principles prevalence across MIM phishing examples

Upon examining each category of impersonated organisations, the findings indicate that phishers apply the principles of authority, liking, scarcity, and social proof in most organisation types (refer to Figure 5-3). Nevertheless, the results show that phishers extensively use the liking and social proof principles in messages targeting the retail industry, as illustrated in Figure 5-3, which presents 12 cases of the use of the liking principle and 12 cases of the use of the social proof principle. The extensive application of the liking and social proof principles was also evident in messages targeting the Fintech sector, with 10 cases utilising social proof and 12 cases applying the liking principle. Figure 5-3 indicates that phishers employ the principle of reciprocity when carrying out attacks that specifically target fintech, hotels and restaurants, retail, and software industries. Overall, the social proof principles appear to be used across all the industry categories except the Automobile industry.

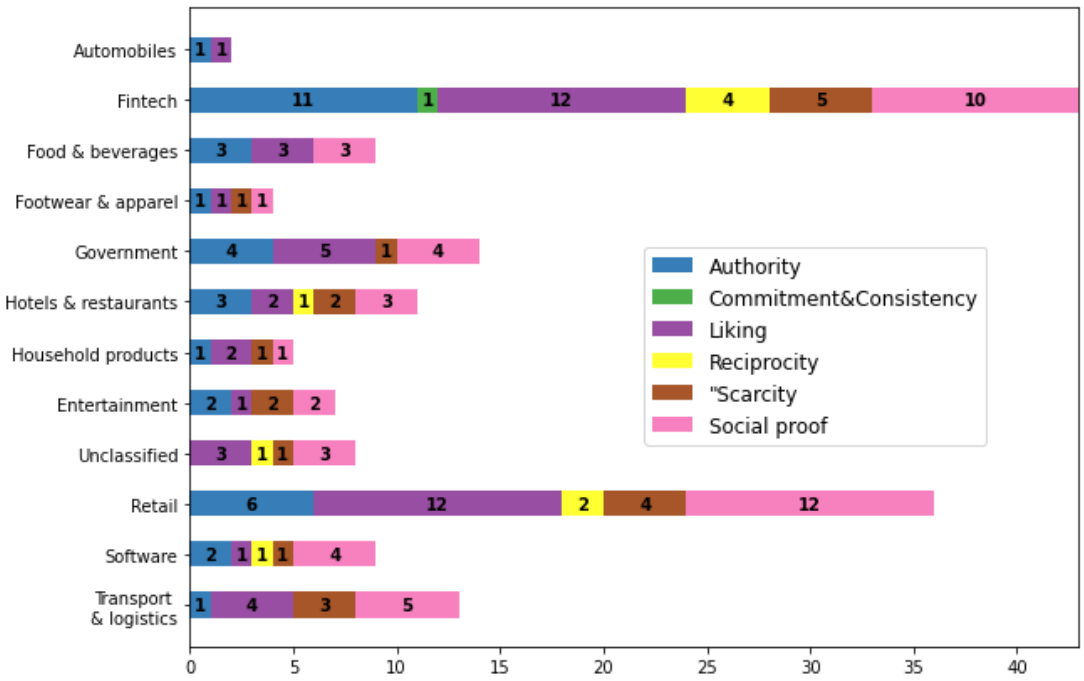


Figure 5-3:Persuasion principles across industry categories (interactive media and internet entertainment were shortened to "Entertainment").

RQ 2.2 What is the typical number of persuasive principles found in most phishing messages?

The dataset was transformed into a Pandas DataFrame to answer this question, with each row representing a phishing message. A Python script was created to count the occurrences of the value "1" in each row of the Pandas DataFrame and save this count in a list. This method enabled quantifying the number of persuasion principles present in each communication. The most common score in the list was determined using the Scipy Python module.

The analysis reveals that most phishing messages (n = 21, 31.3%) had three persuasive principles, while (n = 16, 23.8%) messages contained two principles. Others (n=10, 14.9%) contained four persuasion principles. Phishers were

observed employing five persuasive principles across five messages. Figure 5-4 presents the findings.

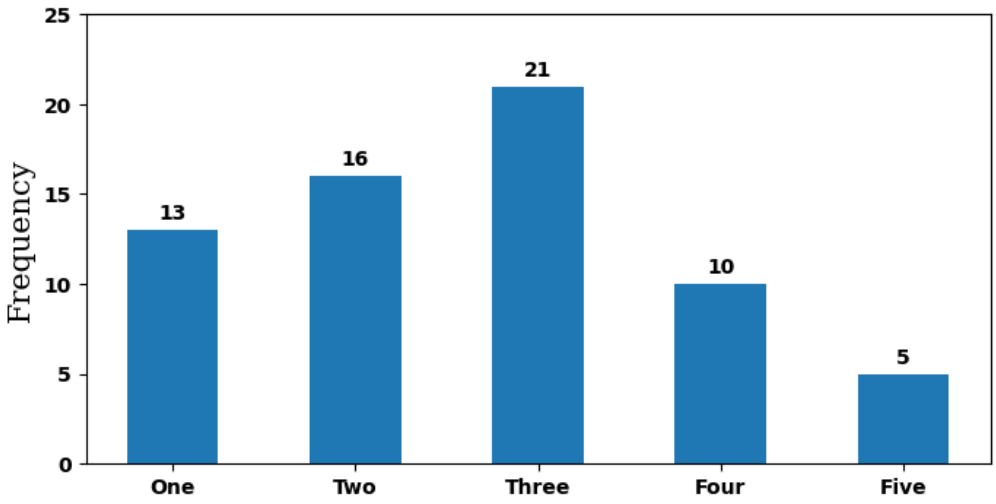


Figure 5-4: Number of persuasion principles in messages

Upon further examination of each combination, it was found that out of the n=16 phishing attacks utilising two principles, (n=8, 50%) were based on the liking and social proof principles. Authority and liking were used together in (n = 6, 37.5%), while authority and social proof were used in (n = 2, 12.5%) phishing attacks. Out of the n=21 messages utilising three principles, the most commonly used combination is: authority, social proof, and liking (n = 8, 38.1%), followed by social proof, scarcity, and liking (n = 7, 33.3%). Additional combinations include social proof, reciprocity, liking (n = 4, 19%), and finally authority, social proof, and liking (n = 2, 9.5%). When phishers employ four persuasion principles (n = 8), they often use liking, authority, scarcity, and social proof (n = 8, 80%). However, phishers used liking, authority, reciprocity, and social proof in two messages. Phishers used five persuasion principles in messages involving cryptocurrencies delivered to users using the Telegram platform. They utilised the principles of liking, authority, reciprocity, social proof, and scarcity in four

specific messages. In one phishing case, phishers utilised the principles of authority, commitment and consistency, liking, social proof, and scarcity.

A subset of the messages (n = 13, 19.4%) employed a single persuasive principle to attract their intended recipients, with the majority (n = 5, 38%) applying the social proof principle followed by authority (n = 4, 30.7%), liking (n = 2, 15.3%), scarcity (n = 1, 7.6%), and reciprocity (n = 1, 7.6%).

RQ 2.3. **How were the different principles of persuasion implemented?**

The analysis reveals that phishers establish authority by employing the logos and names of reputable organisations. The social proof principle was implemented by referencing well-known individuals and emphasising the advantages of visiting the deceptive webpage. For instance, in a specific case of Bitcoin fraud, phishers impersonated the CEO of Coinbase. Furthermore, social proof was demonstrated by including phrases that confirm to the message recipient that the sender had already complied and benefited from such an act. Consequently, the recipient is encouraged to comply by visiting the deceptive webpage. As an illustration, most giveaways and voucher scams use the phrase "I have received mine". This method deceives the recipient into assuming the sender benefited from the scam. Phishers exploit the liking principle by creating the illusion of assistance towards the recipient. The concept of scarcity was effectively illustrated by emphasising the need for recipients of phishing messages to take urgent action, or risk missing out on attractive offers. Phishers implement the scarcity principle by employing phrases like "free air tickets for 200 customers".

Messages employing the reciprocity principle prompt users to act in return for a favour. The content of these emails primarily revolved around fraudulent activities involving cryptocurrencies. As an illustration, phishers deceived customers by informing them that they had won a specified quantity of Bitcoin and could only retrieve it by transferring a designated amount to a

cryptocurrency wallet. Table 5-4 provides the various strategies used to implement these principles.

Table 5-4: How the different persuasion techniques were implemented

Principle	Implementation method	Example
Authority	-logos and names of popular organisations	The logos of Adidas are used as part of a link preview
Social Proof	-Referencing well-known people -Emphasising the advantage of visiting the fraudulent page	-Referencing the CEO of Coinbase -Most messages contained the sentence: "I have received mine"
Liking	-The message appeared to be helping the recipient -Content seems to come from those the recipient likes (i.e. friends or those they share similarities with, such as group members, same gender etc.).	-"Click here to get free vouchers" -Message is forwarded by friends or shared in MIM app groups
Scarcity	-Comply or risk of losing out on offers	"free air tickets to 200 customers"
Reciprocity	-Act in response to a favour	"We are holding a giveaway of 50 bitcoins. To be entered into the giveaway, send 0.02 btc to the address"
Commitment & Consistency	-Leveraged on previous support of group members	"We have been utterly fortunate to have the support and backing of many followers"

RQ 2.4. What URL construction methods do phishers employ to create deceptive URLs during MIM phishing?

As mentioned in section 5.2.1 only 62 emails were used when answering RQ4. Out of these emails, the analysis reveals that the most frequently employed URL construction approach was crafting URLs with a domain similar to the impersonated one (n = 31, 50%), using HTTPS to deceive users (n = 30, 50%) and creating complex URLs with random characters to make them long and difficult to interpret (n = 24, 38.7%). The presence of homoglyphs, which complicates the

identification of forged domains, was also seen in (n = 9, 14.5%) messages. The technique of using TinyURL was seen in (n = 6, 9.6%) and replacing a letter from the original domain name (n = 7, 11.3%). There was no attempt to replace the hostname with an IP address in any of the URLs analysed.

Phishers frequently employ a variety of URL construction strategies during phishing attacks. In the current study, phishers used multiple URL construction strategies in the same campaign. Specifically, the findings indicate eight distinct sets of URL creation techniques. Similar domains and complex URLs are often applied together (n = 8, 12.9%). Complex URLs and HTTPS were found to be used together (n=7, 11.3%). For example, the URL:

http[://]play[.]google[.]store[.]apps[.]details[.]settings[.]pw/play?=1,

which emulated Google Play by utilising complex URLs and similar domains. Table 5-5 displays additional URL creation approaches that were seen being utilised in conjunction

Table 5-5: Overall prevalence of observed URL construction techniques sets in the sample

URL obfuscation sets	Freq	Percentage
Similar domains and complex URLs	8	12.9%
Complex URLs and HTTPS	7	11.2%
Similar Domains, Letter substitution and homoglyph	6	9.6%
Similar Domains and HTTPS	5	8.1%
Tiny URLs and HTTPS	5	8.1%
Similar Domains Complex URLs and HTTPS	4	6.4%
Similar Domains and homoglyph	3	4.8%
Similar Domains Letter substitution	1	1%

Section 5.1 argued that phishers' ability to manipulate certain key components of the preview can influence the URL construction techniques they use during MIM phishing. To examine this claim, a detailed analysis was undertaken to determine the specific URL construction techniques employed by phishers in instances where a phishing message contained a manipulated link preview. Table 5-6 presents the various URL construction techniques employed by phishers in conjunction with manipulated link previews to deceive users of MIM applications into clicking on fraudulent links. The percentages are out of the 32 messages identified as having a link preview. Table 5-6 suggests that phishers often combine Complex URLs with manipulated link previews in many instances. At the same time, the Tiny URL technique was used in only two instances.

Table 5-6: Use of URL construction techniques together with manipulated previews

URL obfuscation sets with preview	Freq	Percentage
Complex URLs, Preview	4	12.5%
Similar Domains, HTTPS, Preview	3	9.3%
HTTPS, Preview	4	12.5%
Tiny URL, HTTPS, Preview	2	6.3%
Similar Domains, homoglyph, Preview	1	3.1%
Similar Domains, Preview	1	3.1%
Similar Domains, Letter substitution, homoglyph, Preview	3	9.3%
Tiny URLPreview	1	3.1%
Complex URLs, HTTPS, Preview	4	12.5%
Similar Domains, Letter substitution, Preview	1	3.1%
Similar Domains, Complex URLs, Preview	4	12.5%

RQ 2.5. Does the prevalence of URL construction methods vary according to the type of company being impersonated?

An examination of the frequency of various URL construction methods among different company categories reveals that phishers often employ similar domains while pretending to be organisations in the retail sector. They commonly employ HTTPS to trick users during phishing attacks that target Fintech companies. The retail industry exhibited the greatest utilisation of homoglyphs. Figure 5-5 displays the frequency of various methods among different types of organisations.

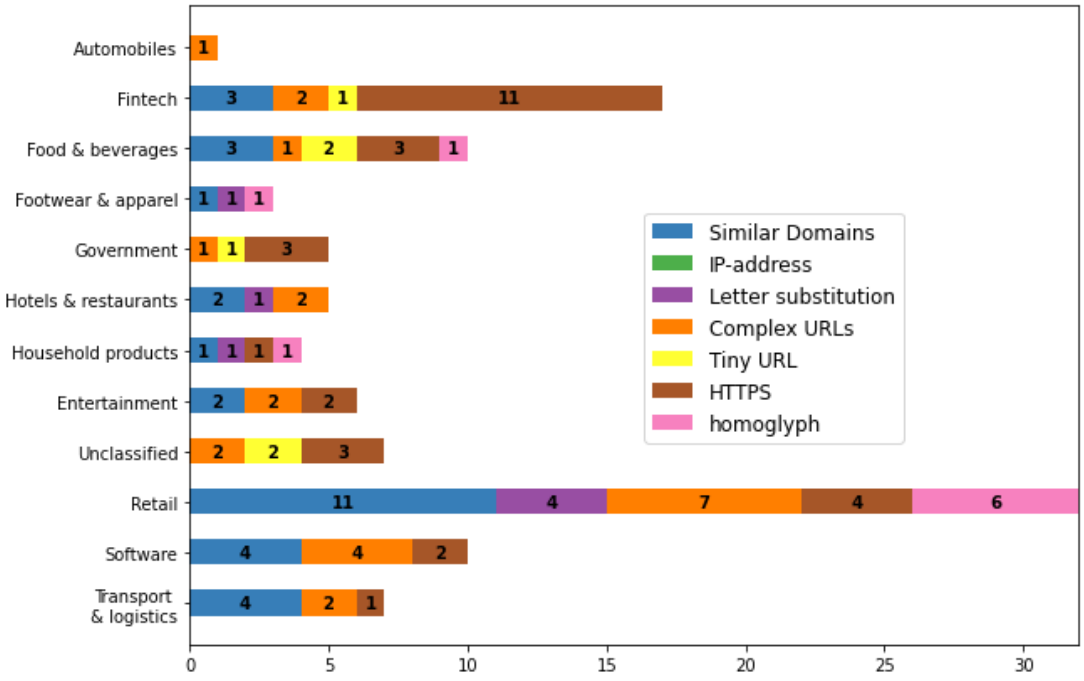


Figure 5-5: Techniques used to construct URL across company categories (interactive media and internet entertainment was shortened to "Entertainment").

5.4 Discussion

This study demonstrates phishers' utilisation of MIM apps to disseminate phishing campaigns. Most phishing communications were carried out through WhatsApp, which aligns with the data from (Kaspersky, 2021). The frequent targeting of the WhatsApp platform is likely due to its large user base, with recent statistics suggesting that WhatsApp Messenger has a user base of approximately two billion users every month (Dixon, 2024). The findings of this study indicate that phishers use giveaways as a lure during MIM phishing. They used well-known phrases such as "free vouchers" or “free giveaway” to lure users into divulging sensitive details. This is similar to what was seen in email phishing (Stojnic, Vatsalan and Arachchilage, 2021). This illustrates the prevalence of financial gain and reward, a common tactic employed in phishing emails. In addition to customer support targeting cryptocurrency users, security updates and phishing attacks also targeted the same users.

The words "get" and "click" were frequently used to persuade users to click on fraudulent links, similar to what was observed in Stojnic, Vatsalan and Arachchilage (2021). The findings indicate that, regarding URL construction strategies, phishing in MIM apps is not significantly different from email phishing, as phishers employ common URL manipulation techniques, including using similar domains, HTTPS, and creating complex URLs to mislead users. However, the results show that phishers consistently combine complex URLs with modified link previews during MIM phishing. By using this approach, phishers can deceive users and instil a sense of legitimacy by including the details of the impersonated organisation in the title, site description, Image/logo, and domain name of the link preview, whilst using the URL construction technique of complex URLs to obfuscate the true destination of the URL. This finding also suggests that phishers are aware of the hardware constraints of smartphones, such as their small screen size, and exploit this limitation to make it challenging for users to parse URLs accurately.

Another interesting observation is the widespread use of HTTPS in MIM phishing, which is similar to what is currently observed in email phishing (Stojnic, Vatsalan and Arachchilage, 2021). However, this is an intentionally deceptive approach, considering that online anti-phishing tips typically advise users to look for HTTPS when transmitting critical information over the internet (Roberts *et al.*, 2019). This advice aligns with the functionality of HTTPS, which is to encrypt data as it passes from one communicating party to another. However, in most cases, users misconstrue the actual functionality of HTTPS by erroneously believing that it signifies the reliability and credibility of a website (von Zezschwitz, Chen and Stark, 2022), whilst, in reality, HTTPS is a cryptographic protocol employed to ensure that a third party does not intercept communication between two parties on the internet. It was noted that the URLs of most phishers that pretended to be fintech organisations employed HTTPS. A recent survey of fintech users indicates that a significant number of them possess advanced educational qualifications. The use of HTTPS in fintech scams may suggest that phishers are aware that fintech consumers, who are likely to possess advanced technological skills, may exhibit scepticism towards URLs that do not include HTTPS.

Regarding the persuasion principles used, the findings from this study revealed that phishers frequently employ the social proof principle and the liking and authority principle during MIM phishing. Other principles most used are scarcity, reciprocity, commitment and consistency. Moreover, in most cases, phishers employ three persuasive principles, typically a blend of social proof, authority, and liking. Phishers commonly establish their credibility by using the logos and names of reputable organisations. Utilising the logos and names of reputable businesses enhances the credibility of the source and persuades the message recipient to take action (Kim and Kim, 2013), aligning with the source credibility theory, which posits that individuals are more inclined to accept and respond to communication when the source portrays itself as trustworthy (Hovland, C.I., Janis, I.L. and Kelley, 1953). However, with MIM apps, the user forwarding the message or the organisation being impersonated can be considered the source,

as users could use any of these elements to establish the credibility of the communication. Perhaps this is the rationale behind phishers frequently requesting that users share the message with additional recipients, since this approach allows them to leverage the existing trusting relationship between the user and their contact (Cuddeford, 2018).

Another notable finding is that most giveaway and free voucher campaigns include the phrase "I have received mine", purposely inserted to mislead users into believing that the senders of the message have benefited from the scam. This strategy can be highly efficacious when the recipient has confidence in the sender. Evidence has shown that phishers have successfully convinced numerous people to respond to MIM phishing messages via this method, as reported in ActionFraud (2018) and iRadio (2018). Moreover, this strategy aligns with the social proof principle, which relies on shared risk. In this case, alerting the recipient that previous users have taken similar actions to reassure them that complying will lead to some benefit.

Although social proof is commonly employed in MIM phishing, it is not widely used in email phishing (Ferreira and Lenzini, 2015). This shows a potential disparity between MIM and email phishing. Nevertheless, this study's findings align with previous research on vishing scams, highlighting the wide usage of the social proof principle to emphasise the advantages of compliance, indicating their reliance on it (Jones *et al.*, 2020).

The findings of Chapter 4 showed that people used MIM apps informally and often between people with some offline connection. It is also among those with a common interest, such as friends or members of the same MIM app group. This facilitates the implementation of the liking principle by phishers, as the messages will likely originate from individuals the recipient likes. The idea is that after successfully deceiving a user or compromising their account, phishers can access the compromised user's contacts and move to additional users. The prevalence of the liking principle in MIM phishing scams suggests that, in most cases,

phishers rely on others to propagate their MIM phishing campaigns. This is the same approach phishers use in other media, such as email. For example, in a recent analysis of multistage adversary-in-the-middle (AiTM) phishing and business email compromise (BEC) attacks affecting various organisations in the financial sector, the Microsoft threat intelligence team observed that phishers began by compromising a trusted vendor and then used the compromised account to engage in a series of AiTM attacks and follow-on BEC activity across many organisations (Microsoft Threat Intelligence, 2023).

The findings also indicate that the reciprocity, commitment, and consistency principles were used with less frequency, aligning with the existing body of work on vishing, where phishers employed these principles to a lesser extent (Jones *et al.*, 2020). On the other hand, Ferreira *et al*. (2015) discovered that reciprocity, commitment, and consistency are frequently employed in emails that promise substantial amounts of money, and tend to be more personalised and often informal. They also tend to ask for favours from the target. Implementing the concepts of reciprocity, commitment, and consistency in MIM phishing can be a difficult task. For instance, it may be difficult for phishers to demonstrate an individual's previous commitment or consistency, except in situations where the attack is targeted and a level of confidence has been established between the phisher and the victim in the past. This may be challenging to implement for the type of attacks we see in this study. Likewise, reciprocating may be more feasible in a targeted phishing attack that necessitates the phisher to develop a certain level of trust with the intended victim. Yet, there was an attempt to get users to reciprocate. For example, Table 5-4 shows that users were told to send 0.02 BTC to a particular Bitcoin address before they could be entered into a giveaway of 50 bitcoins.

Regarding the organisations that were impersonated the most, the findings indicate that retail companies were the most targeted, followed by fintech platforms. This contradicts previous email phishing studies, where phishers impersonated education and health institutions the most (Zielinska *et al.*, 2016;

Akdemir and Yenal, 2021). However, this is probably because retail and finance firms use MIM applications to communicate with customers (Vasiliu *et al.*, 2023).

The results of this study indicate that most phishing instances contain modified link previews. A modified link preview can potentially mislead users into believing they are accessing the correct webpage. In their 2020 paper, Stivala and Pellegrino expressed concerns about the possibility of exploiting link previews to mislead visitors into accessing deceptive web pages (Stivala and Pellegrino, 2020). The current study validates the utilisation of link previews by phishers as a means to redirect consumers to deceptive web pages. Phisher's ability to modify the link preview is aiding them in implementing the authority principle by allowing them to include logos of impersonated organisations in their campaigns, deceiving users into thinking they are interacting with legitimate content. This finding highlights how certain software design and development practices can hinder cybersecurity professionals' efforts in combating phishing threats.

5.5 Limitations

The initial step of this study involved extracting the example phishing messages from publicly accessible websites using the Google search engine. The dataset does not accurately represent all MIM phishing attacks that were circulated within the given period. MIM phishing attacks that were not reported or shared online were not recorded. Nevertheless, due to the novelty of phishing in MIM apps, there is a lack of authenticated databases containing information about such schemes. Consequently, the only recourse is to rely on the existing technique. Hopefully, in the future, researchers will create a specialised MIM phishing corpus to aid in tackling the problem of phishing in MIM apps. However, despite the relatively small sample size, the sample is within the range previously reported in the literature in similar studies, for example (Jones *et al.*, 2020).

Furthermore, despite its potential, deductive content analysis is thought to occasionally hinder researchers' capacity to identify additional contextual

characteristics that may be relevant to the phenomenon under investigation (Hsieh and Shannon, 2005). It is possible that other elements of MIM phishing relating to platform-specific affordances, or new attacker–victim interaction patterns, have not been identified due to reliance on existing categories from the literature.

5.6 Chapter summary and conclusion

The study in this chapter provides valuable insights into the characteristics of phishing in MIM apps by examining the techniques used by phishers to create misleading URLs and employ Cialdini's persuasion principles in MIM phishing. The findings indicate that these concepts are commonly employed in MIM phishing, along with the utilisation of popular URL construction techniques that combine domain spoofing, HTTPS, and complex URLs to generate misleading URLs. The first conclusion that can be drawn from this study is that phishers exploit the vulnerability in the link preview generation mechanism in MIM to create deceptive links. The findings show how users' susceptibility to MIM phishing is increased by phishers' practice of combining complex URLs with HTTPS to create deceptive URLs. Furthermore, the modification of link previews by phishers to create deceptive URLs can increase user susceptibility. Specifically, phishers manipulate the link previews by including the logos of impersonated organisations to increase the likelihood of users responding to the phishing campaigns. Moreover, the findings from the exploratory study conducted in this thesis have shown that participants currently check link previews when verifying the legitimacy of links.

The second conclusion that could be derived from this study is that phishers leverage the existing social trust between users of MIM apps to propagate phishing campaigns and increase user susceptibility. The findings show that phishers leveraged relationships to enhance phishing credibility. They frequently apply the social proof and liking principle, constructing phishing communications using phrases that convey the idea of shared risk to convince recipients that the

senders of the message have already benefited from the scam. This approach can potentially increase users' likelihood of responding to the phishing messages.

The above points highlight how phishing in MIM apps incorporates a unique approach by utilising two aspects: the organisation being impersonated and the sender, to ensure source legitimacy, potentially enhancing the chance of successfully defrauding unsuspecting victims.

Furthermore, it is interesting to see that phishers rely heavily on the social proof principle during MIM phishing, often demonstrating that other users have acted in this manner in the past. This raises the question of whether this strategy is effective, especially considering the informal nature of instant messaging, where users are likely to receive messages from those they know. Furthermore, most of the messages analysed relate to financial gain and reward. Given the current user awareness of phishing and the fact that most of these attacks appear too good to be true, it is unlikely that anyone will fall for them. However, it is unclear if this is the case. Due to these findings, the next chapter will investigate how susceptible users are to these forms of attacks and also examine the effect of some key features of MIM apps on phishing susceptibility.

Chapter 6

The effect of specific characteristics of MIM apps on user susceptibility to phishing

The study presented in this chapter examines how specific characteristics of MIM apps affect susceptibility to phishing in these apps, using WhatsApp as a case study. First, the chapter introduces the study and outlines the study's aim and hypotheses. Then, it describes the method used to design the experiment, the survey instrument, and the data collection. Next, the chapter outlines the procedure used to recruit participants. It then discusses the analysis approach, followed by a detailed discussion of the study's findings and limitations.

6.1 Introduction

This chapter investigates the effect of specific characteristics of MIM apps on user susceptibility to phishing. It takes into account the findings of Chapters 4 & 5, together with the findings of section 2.2.2, to identify potential characteristics of MIM apps that may impact susceptibility to phishing.

As highlighted in 2.2.2, the literature currently suggests that communication within mobile instant messaging (MIM) applications tends to be predominantly informal and typically occurs between individuals who already have an existing offline relationship. The results of the exploratory study in Chapter 4 have also confirmed this, as many participants reported using MIM apps to communicate with friends, family, and colleagues at work or in business. Unfortunately, the findings of the qualitative content analysis of MIM phishing messages in Chapter

5 suggest that cybercriminals try to abuse the trust users have in their contacts by asking them to forward phishing messages to them. Yet, it has been shown that internet users assess the credibility of online resources using a social confirmation or social proof approach (Metzger, Flanagin and Medders, 2010). This approach assumes that credibility can be established through the actions and beliefs of others and relies on endorsement heuristics, which refer to individuals' inclination to perceive information as credible because others perceive it as such. Social confirmation will likely play a key role in phishing within MIM apps, primarily because cybercriminals rely on victims to disseminate their campaigns through such media. This approach can help cybercriminals reinforce credibility, as the recipient of the message is likely to believe that the sender has acted similarly. Thus, there is a need to understand if the relationship the user has with the sender of a message impacts the trust assessment of phishing messages in MIM apps. Furthermore, there is also a need to understand how this relationship affects users' susceptibility to phishing in MIM apps.

In Section 2.2.4, this thesis highlights existing studies that demonstrate how design cues or features related to the aesthetic presentation of emails or websites influence individuals' trust perceptions of these platforms. Section 2.2.4 also highlights the literature on phishing, which reveals that participants rated phishing emails with logos of known brands and copyright statements as more trustworthy than those without these features. In section 2.2.2, it was explained that MIM apps generate a link preview as part of a strategy to give users an idea of the website they are visiting. These previews often contain images, including the logos of well-known brands. However, including logos in link previews raises the question of whether such logos impact users' susceptibility to MIM phishing.

Given the findings in Chapter 5, which show that cybercriminals manipulate the content of link previews to include the logos of impersonated organisations, it is essential to determine the extent to which the link preview affects phishing susceptibility. However, there does not seem to be any thorough study or analysis

demonstrating the extent to which link previews impact users' susceptibility to phishing in MIM apps. Seemingly, there is evidence that previews increase link click-through rates during online marketing campaigns (McLachlan, 2021). However, the extent to which people use link previews to assess the legitimacy of messages during communication in MIM apps remains unclear.

Section 2.2.1 highlights findings from the existing literature that reveal individuals' habits during email communication and social media interaction affect their susceptibility to phishing. With instant messaging becoming a significant part of many people's lives, the use of these apps may become habitual and potentially influence how users respond to phishing.

6.2 Study aim and hypotheses

The primary aim of this study is to examine the impact of both the user's relationship with the sender of a message and the use of a link preview on susceptibility to phishing in MIM apps, using WhatsApp as a case study. The focus on WhatsApp is due to its being the most popular MIM app and also the most widely used by cybercriminals for circulating phishing campaigns. The study aims to investigate whether the types of relationships in WhatsApp (i.e. friends, family members, work/business colleagues, someone the user knows but not family, friends or work/business colleagues, and unknown) and link preview influence participants' likelihood of clicking suspicious links embedded in WhatsApp phishing messages and how trustworthy they consider such messages. Furthermore, the study will investigate whether habitual WhatsApp usage affects users' susceptibility to phishing on WhatsApp. Susceptibility in this study refers to the self-reported likelihood of users clicking on phishing links during communication in MIM apps.

Whilst this chapter seeks to answer **RQ3,** which is: Do specific characteristics of MIM apps influence user susceptibility to phishing within these platforms?

The following sub-questions were formulated to help in successfully answering the research questions. First, sub-question one was introduced due to the discussion on the characteristics of mobile instant messaging apps in section 2.3.2 which highlighted how communication in MIM apps tends to be between known contacts but also raised concern regarding evidence from the literature that indicates that users assess the credibility of online resources using a social confirmation or social proof approach, perceiving information as credible because others have deemed it credible. Sub-question two was introduced due to section 2.3.4, highlighting that design cues or features related to the aesthetic presentation of emails or websites influence individuals' trust perceptions of phishing emails. Finally, sub-question three was introduced due to section 2.3.1 highlighting that individuals' email and social media habits affect their susceptibility to phishing. The sub-questions are:

RQ3.1 Does susceptibility to phishing in WhatsApp differ by the user's relationship with the sender of a message?

RQ3.2 Are users of WhatsApp more susceptible to phishing with a link preview than those without a link preview?

RQ3.3 Does habituation to WhatsApp influence susceptibility to phishing within it?

The following hypotheses have been proposed to achieve the study's aim and answer its research questions.

H1: The relationship with the sender of a message influences users' likelihood of responding to phishing in WhatsApp.

- **H1a** Users will be more likely to report the intention to click on links in phishing messages sent by contacts they know than by unknown contacts.
- **H1b** Users will be more likely to report the intention to click on links in phishing messages sent by family, friends, or work/business colleagues

than from contacts they know who are not family, friends or work/business colleagues.

H2: The relationship with the sender of a message influences users' perceived trust ratings of phishing messages in WhatsApp.

- **H2a** Users will be more likely to report higher trust ratings for phishing messages sent by contacts they know than for those sent by unknown contacts.

- **H2b** Users will be more likely to report higher trust ratings for phishing messages sent by family, friends, and work/business colleagues than from contacts they know who are not family, friends or work/business colleagues.

H3: Users will be more likely to report the intention to click on links in phishing messages with a link preview.

H4: Users will be more likely to report higher trust ratings for phishing messages with a preview.

H5: The higher the individual's self-reported habituation to WhatsApp, the more likely they are to report an intention to click on phishing links

H6: Higher self-reported habituation to WhatsApp will lead to higher trust ratings of phishing messages

6.3 Method

6.3.1 Design

The study employed a within-subject design and utilised a factorial survey (FS) methodology, also known as the Experimental Vignette study (EVM), to investigate participants' likelihood of clicking on phishing links in WhatsApp and

their trust ratings of such messages in different scenarios. Details of the factorial survey (FS) methodology and the justification for its adoption in this research are presented in 3.2.4. The study employed the FS methodology to experimentally manipulate the "participant's relationship with the message sender" and "presence or absence of link preview". A total of 10 hypothetical vignettes were created to ensure these factors occur in every combination. Table 6-1 shows the levels of each factor and the associated phrases used.

Table 6-1: Vignette variables and their levels

Factor /variables	Levels
Participant's relationship with the message sender	Family Friends Work/business colleague Someone you know but not family, work or business colleagues Uknown contact
Link Preview	Link with a preview Link without a preview

6.3.2 Vignettes development

This study examines participants' likelihood of clicking on phishing links and their trust ratings of phishing messages, rather than their ability to distinguish between phishing and legitimate messages. Therefore, only phishing vignettes were created. The phishing messages impersonated 10 organisations. The organisations were selected based on two criteria. First, the organisation falls under the type of organisation that phishers currently impersonate during MIM phishing campaigns. Second, the organisation appeared in Cloudflare Radar Domain Rankings of the top 20,000 websites visited. This ensured that the organisations were familiar to the participants and represented the type of

organisations seen in the wild. Phishing-related studies often use websites from Alexa's top 500 to ensure familiarity. For example, the study by Alsharnouby, Alaca and Chiasson, (2015) selected websites from Alexa's top 500 list of most visited domains in Canada to ensure that their participants were familiar with the sites being tested. Alexa's top list was also used in a recent study of Dark Patterns (Krisam *et al.*, 2021). However, Amazon shut down (retired) Alexa on May 1st, 2022 (Duò, 2022). Therefore, Cloudflare Radar Domain Ranking was used instead of Alexa. Cloudflare Radar Domain Rankings offers more comprehensive domain rankings than Alexa. The Cloudflare website says their ranking "*identifies the top most popular domains that reflect how people use the Internet globally*" (Martinho and Zejnilovic, no date). Consequently, websites from this ranking will likely be familiar to the study participants. Table 6-2 lists the selected organisations. All organisations except Coca-Cola have been used by phishers in MIM phishing, as evidenced in the qualitative content analysis of phishing messages in Chapter 5. However, Coca-Cola was selected for impersonation in the study rather than Heineken, despite evidence that Heineken had been the actual target in real-world phishing attempts. This decision was made due to Coca-Cola's higher global brand recognition, which was deemed more suitable for eliciting participant responses.

Table 6-2:List of companies selected

Organisation	**Cloudflare Radar Domain Rankings**	**Global presence**
Booking.com	Yes	Online presence
Qatar Airways	Yes	Online presence
Netflix	Yes	Online presence
Nestle	Yes	188 countries
Amazon	Yes	Online presence
Ikea	Yes	Presence in 61 countries
Binance	Yes	Online presence
WhatsApp	Yes	Online presence
Google	Yes	Online presence
Coca-Cola	Yes	Presence in over 190 countries

6.3.3 Pretexting

The findings in Chapter 5 revealed the tactics and techniques used in MIM phishing. These findings served as a guide in creating the pretext of the phishing messages used in this study. These messages were crafted using approaches similar to those used by cybercriminals to phish users of MIM apps (See APPENDIX C.1 for the full messages used). While the messages differed in content, examining the impact of content variation falls outside the scope of this study. Nonetheless, efforts were made to ensure that the messages were comparable in terms of the attacker's intent and employed tactics.

This was achieved by ensuring that:

1) Each message contained the two frequently used persuasion principles in MIM phishing, namely authority and social proof. For example, the social proof principle was integrated by ensuring that the content implies that complying with the phisher's request (i.e clicking the link) will benefit the user. For authority, each message contains either the name of the organisation being impersonated as part of the URL or a logo, as in the case of the link preview.

2) A pair of phishing messages (i.e., one with a link preview and the other without) conveys the same meaning and context. For example, the phishing messages on Google Play and WhatsApp, as shown in Appendix C, aimed to trick users into downloading malware on their mobile phones.

Furthermore, phishing messages based on giveaways reflect the type of campaigns conducted by phishers, as evidenced in the findings of Chapter 5. These campaigns are also likely to be seen as legitimately conducted by these companies. For example, Netflix conducted a similar legitimate discount campaign (Mirror, 2023).

Two of the vignettes were designed to come from unknown contacts. These vignettes were crafted to reflect similar campaigns seen in the wild. These messages impersonated the organisations Airbnb and Binance. The Binance message was a modified version of a phishing campaign targeting users of the Binance cryptocurrency platform (Binance, 2021), while the Airbnb message was a modified version of a recent scam targeting Booking.com users (Sowerby, 2023). Airbnb was used instead of Booking.com to reduce suspicion, as some participants may have been aware of or had seen the Booking.com phishing messages.

Unlike the remaining vignettes, messages from unknown contacts may immediately trigger user suspicion. For example, a participant who is not a Binance user may likely flag an account security update message from Binance, coming from an unknown number, as suspicious. Therefore, the participants were instructed to assume the role of users of these platforms. For example, the Vignette for Binance reads: "*You are a user of Binance, a Cryptocurrency exchange platform. You received the following message from an unknown number via WhatsApp*". This thesis adopted this approach because it reflects a real-world scenario where users are contacted from unknown numbers, as seen in the Airbnb Community Centre, where users highlighted being contacted via WhatsApp by someone claiming to be their hosts with numbers unknown to them (AirBNB Community Center, 2018). Nonetheless, the vignettes integrate text and images. The text captures the participant's relationship with the message sender, and the image captures the link preview. The text was personalised to reflect the names provided by the participants. For example, Figure 6-1 below shows a vignette for a participant who provided the name "Marc" for their friend.

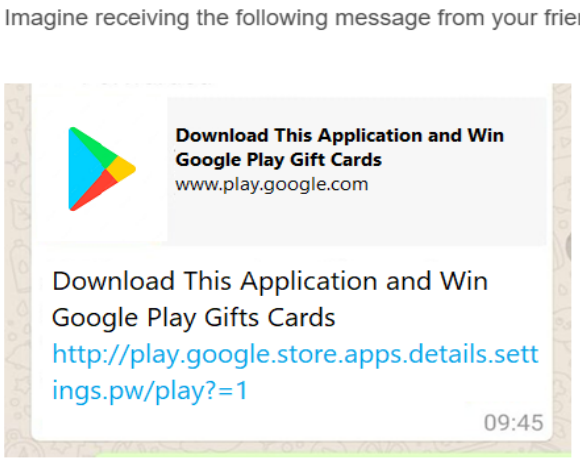


Figure 6-1: Figure 6 1: Example Vignette

6.3.4 Survey instrument and implementation

The experiment was implemented as a survey using the online survey creation software Qualtrics. The survey contained demographic questions (age, gender, educational level, and country of residence). It also included four items that provide a composite score of a participant's WhatsApp habituation. Three items were adapted from Vishwanath (2015), where the author measured Facebook habituation. The items were modified to reflect the current research. They included "I feel my WhatsApp use has gotten out of control," "I would miss checking my WhatsApp if I could no longer do it," and "I find myself checking my WhatsApp (posting or checking other people's status updates) around the same time each day". The fourth item, " Checking WhatsApp is part of my daily routine", was adapted from LaRose and Eastin, (2004) study on internet use and gratifications. The items of the WhatsApp habituation scale were based on a five-point Likert response, ranging from "strongly disagree" to "strongly agree."

Furthermore, the survey contained questions that asked participants to provide a unique pseudonym each for their friend, family member, work/business colleague, and someone they know who is neither a family member, friend, nor

work/business colleague. These pseudonyms represent the relationship between the message sender and the participant, followed by the ten crafted hypothetical vignettes from section 6.3.2, each presenting a "scenario" within one of ten combinations. Table 6-3 summarises the scenarios and combinations. Each hypothetical vignette/scenario was followed by two questions: one asking participants how likely they were to click the link in the vignette and the other about their trust ratings of such messages. The question on likelihood was based on a 5-point Likert scale from 1) extremely unlikely, 2) somewhat unlikely, 3) neither likely nor unlikely, 4) somewhat likely, and 5) extremely likely. This approach is used extensively in the literature to measure users' susceptibility to phishing (Vishwanath *et al.*, 2011; Lin *et al.*, 2019; Parsons *et al.*, 2019). Similarly, the trust rating question was based on a 5-point Likert scale from 1) not at all trustworthy, 2) slightly trustworthy, 3) somewhat trustworthy, 4) very trustworthy, and 5) extremely trustworthy. This is similar to what is used in a related study by Williams and Polage, (2019).

The survey also contained two questions that asked participants to state the reasons why they were likely or unlikely to follow the links in the vignettes. The questions contained multiple-choice (select all that apply) options. However, the multiple options the participants saw depended on whether they selected a likelihood of clicking or not. When a participant selects either *somewhat likely* or *extremely likely*, the options are: 1) because I know the sender, 2) the message referenced a well-known organisation, 3) because of the logo, and 4) others. However, the option “because of the logo” was only presented when the vignettes featured a link preview. Similarly, the option “Because the message came from an organisation that I know” was only offered in the case of vignettes depicting unknown relationships, ensuring the options were contextually relevant.

Similarly, when participants selected either *"extremely unlikely" or "somewhat unlikely,"* options were provided to understand the reason behind their choices. The options depended on whether the message was from a known relation or not. For messages from a known relationship, the options are: 1) because I know the

sender, and 2) Other. However, for messages coming from an unknown sender, the options are 1) because I don't know the sender, and 2) other. Furthermore, for vignettes without a link preview, the option "because of the absence of a logo" was also included to understand if the absence of a logo may be a reason for the unlikelihood of clicking. Regardless of their chosen option (i.e. likelihood to click or unlikelihood), the option "other" was provided to elicit qualitative responses. Figure 6-2 shows the complete flow of the options provided to the participants.

Table 6-3: Summary of the vignette scenarios/combinations

Vignette ID	Type of phishing message used	Apparent identity and relationship of the sender	Link preview
1	Software download in exchange for enhanced functionalities	Friend	No link preview
2	Free return ticket giveaway to two lucky winners	Family member	No link preview
3	Discount Promo	Work/business colleague	No link preview
4	Giveaway	Someone you know but not friends, family or work/business colleague	No link preview
5	Account confirmation	Unknown contact	No link preview
6	Software download in exchange for a free Google Play gift card	Friend	Has a link preview
7	Free Voucher	Family	Has a link preview
8	Discount Promo	Work/business colleague	Has a link preview
9	Giveaway	Some you know but not friend, family or work/business colleague	Has a link preview
10	Payment verification	Uknown contact	Has a link preview

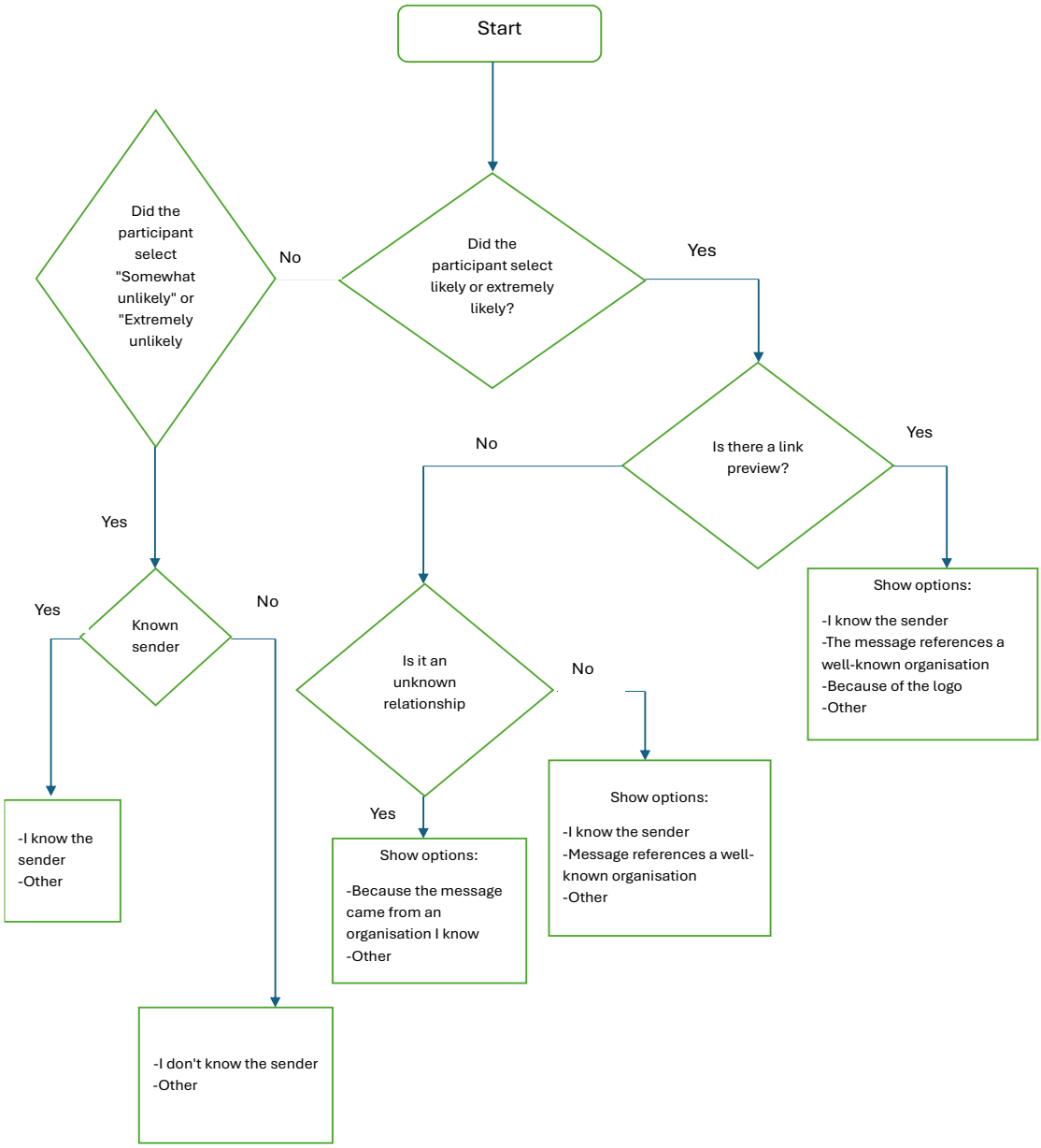


Figure 6-2: A flow chart that shows options available to participants when they choose likely to click or unlikely to click

An attention check question was used to assess data quality. Using the recommended practice of the **semantic synonym technique** (DeSimone, Harms and DeSimone, 2015), two similar items were included, one at the beginning and

the other at the end of the survey. In the first part, participants were asked if they currently used WhatsApp. In the second part, participants were provided with a list of MIM apps, which included both WhatsApp and other MIM apps. Participants were asked to indicate which of the apps they currently use. The idea is that those who stated they used WhatsApp in the first question should select WhatsApp in the second question.

As the study employs a within-subjects design, there is a potential for an order and carry-over effect, where participants' responses in one condition may influence their responses in later conditions. As a result, randomisation was implemented in Qualtrics using the advanced functionality of branching and randomisation to minimise this effect. This approach randomly displays the vignettes to the participants. Randomisation is the recommended approach for minimising the risk of carry-over effects (Dhamija, Tygar and Hearst, 2006; Petelka, Zou and Schaub, 2019).

The efficacy of the survey questions was tested through a pilot study conducted with the research supervisor and three other research students. The participants in the pilot study recommended improving the clarity of the introduction and consent note, including the “other” option, to better understand the reasons behind the participants' decisions to click. The pilot participants also recommended asking the participants to indicate why they decided not to click on the link. Feedback from the pilot study was integrated into the final survey. APPENDIX C.1 provides the full vignettes and questions used in the study.

6.3.5 Target participants and recruitment

The target participants were WhatsApp users aged 18 years and above. As mentioned in section 6.2, WhatsApp was chosen because it is currently the most popular instant messaging application (Dixon, 2024) and is the most frequently targeted by phishers (Kaspersky, 2021). The participants were recruited using a variety of methods. The survey link was posted on multiple social media groups,

including r/SampleSize on Reddit, Sample Size on Facebook, and SurveyCycle, as well as the "Dissertation Survey Exchange – Share Your Research Study, Find Participants" group on Facebook, which has 35,700 members. The researcher also shared the link to the survey with contacts on his current MIM Apps, including those in his private and public WhatsApp groups. Participation was encouraged by offering participants a chance to win one of five £20 shopping vouchers.

6.3.6 Procedure

The final study was conducted between November 6th, 2023, and February 1st, 2024. Before the study went live, the researcher applied for and received ethical approval from his department. The advertisement described the purpose of the research, its procedures, the expected completion time, and the benefits. Participants were asked to read and sign an informed consent document before participating. Full details of the information and consent form provided to the participants can be found in APPENDIX C.2.

The participants were not informed about the actual purpose of the study. Instead, they were told that the study aimed to investigate how WhatsApp users react to information shared by different types of people. This approach ensured that the participants were not primed. This approach was adopted because prior literature has revealed that priming participants can influence their responses to questions in phishing studies (Parsons *et al.*, 2013, 2015; Naiakshina *et al.*, 2017). Thus, adopting this approach ensured that the survey measured the natural behaviour of the participants.

Next, the participants were presented with a brief description of the task they would be required to do. Specifically, they were instructed to provide a pseudonym for each of the following types of contact: friend, family member, work/business colleague, and someone they know who is neither a family member nor a work/business colleague. An explanation of what a pseudonym is

was provided in the text. Specifically, they were instructed to provide four pseudonyms: one related to a friend, a family member, a work or business colleague, and someone they know but not a friend, family member, or work or business colleague. They were told that a pseudonym is a fake name that a person or group uses for a special purpose and that the purpose of the pseudonym is to mimic real-world situations based on the assumption that they are likely to receive messages from those whose pseudonyms they have provided. They were told to use a unique pseudonym for each of the communicating parties.

Each participant was presented with ten hypothetical vignettes that reflected all possible conditions of the two contextual variables used in this study. As mentioned earlier, eight of the vignettes were personalised to reflect the pseudonyms provided by the participants.

6.4 Results

This section presents the study's findings. First, data-cleaning strategies were discussed, including the number of participants excluded from the analysis and the justification for doing so. Data cleaning and descriptive statistics were conducted in SPSS, while inferential statistics were performed in STATA.

6.4.1 Participants

A total of 814 participants responded to the survey. However, after cleaning, only 127 participants qualified for analysis. The excluded participants were removed for several reasons, including data quality. Figure 6-3 shows the stages of the data cleaning process. Further explanation of each phase is provided in the paragraphs that follow.

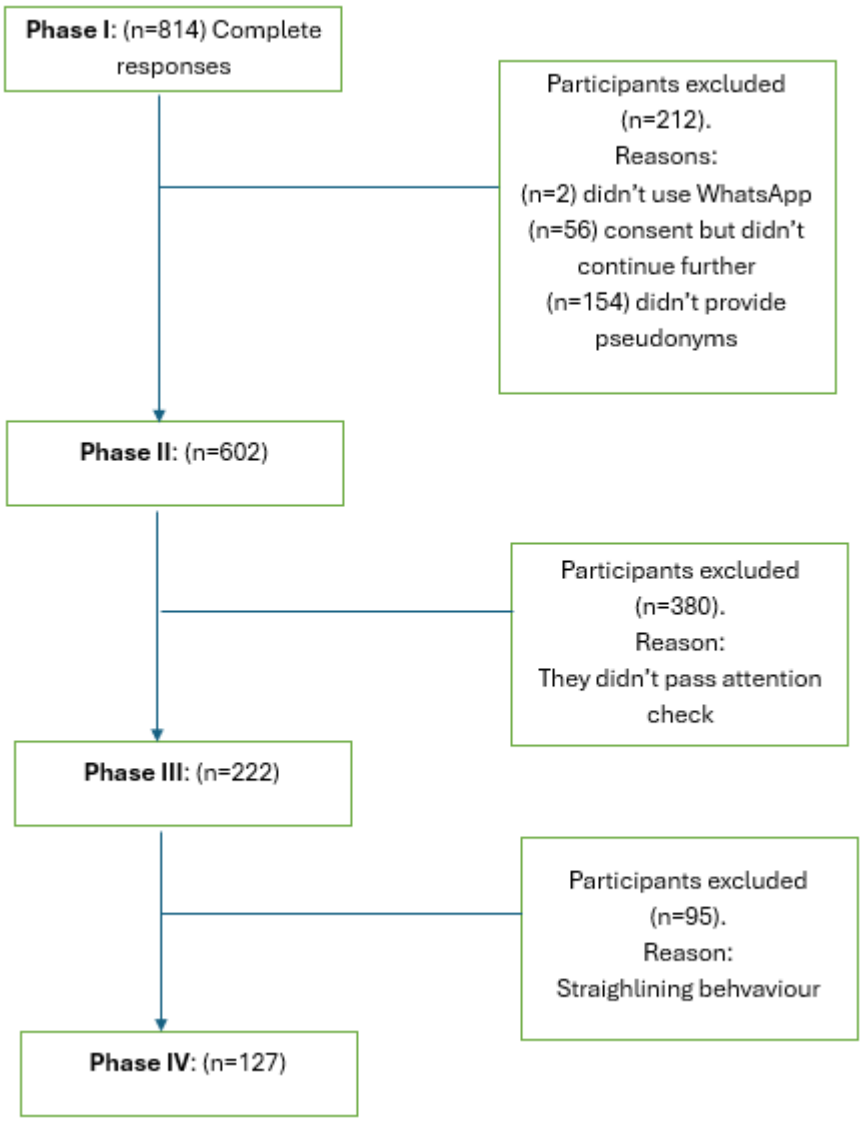


Figure 6-3: Data Cleaning Strategy

Phase I: A total of 814 responses were exported from Qualtrics upon completion of the survey.

Phase II: Since the study targeted WhatsApp users, participants' first question after consenting to participate was whether they currently used WhatsApp. Participants who reported not using WhatsApp were excluded (n=2). Participants who agreed to participate in the survey but did not continue further (i.e. did not provide any further information) were excluded (n = 56).

The survey required participants to give pseudonyms for the different relationship types. Answers to the pseudonyms were mandatory as they were required to customise the vignettes. Therefore, participants who didn't provide these answers were excluded (n=154). This process resulted in 602 participants being carried to the next stage.

Phase III: At this stage, (n=380) participants who didn't pass the attention checks, which were discussed in 6.3.4 were excluded, leaving (n=222). Further analysis of the (n=380) participants who failed attention checks shows that most of them selected the midpoint answer for both likelihood-to-click and trust rating questions, as can be seen in Table 6-4 and Table 6-5. This behaviour indicates evidence of straight-lining (Holbrook, Green and Krosnick, 2003), a known indicator of participant inattention during online surveys. Researchers have also found that straight-lining relates to other attention indicators, such as instructional manipulation checks (Paas and Morren, 2018).

Table 6-4: Cross-tabulation between each of the hypothetical vignettes and click likelihood responses for participants that didn't pass the attention checks (n=380)

		Click					
		1	2	3	4	5	Total
Vignette	1	2	2	366	5	3	378
	2	1	4	361	8	2	376
	3	0	7	361	8	2	378
	4	3	4	360	7	4	378
	5	0	4	362	9	3	378
	6	3	2	364	7	3	379
	7	1	5	361	7	2	376
	8	0	2	365	9	3	379
	9	1	3	363	7	3	377
	10	1	6	359	9	2	377
Total		12	39	3622	76	27	3776

Table 6-5:Cross-tabulation between each of the hypothetical vignettes and trust rating responses for participants who didn't pass the attention checks (n=380)

		Trust					
		1	2	3	4	5	Total
Vignette	1	0	6	363	6	5	380
	2	0	8	361	3	7	379
	3	1	8	361	2	7	379
	4	0	7	363	5	5	380
	5	0	6	362	4	6	378
	6	0	7	365	3	5	380
	7	0	6	362	3	7	378
	8	0	5	366	5	4	380
	9	0	11	360	2	6	379
	10	0	7	360	4	7	378
Total		1	71	3623	37	59	3791

Phase IV: The number of participants left from the previous stage was (n=222). However, evidence of straight-lining behaviour was also observed. Specifically, many participants selected neither "likely" nor "unlikely" for the likelihood question and "somewhat trustworthy" for perceived trustworthiness, as shown in Table 6-6 and Table 6-7. This behaviour is what researchers refer to as satisficing, a behaviour on the part of the participants that is characterised by the use of minimal attention or little effort to answer questions accurately, which has been shown to compromise the quality of survey data (Reuning and Plutzer, 2020). Satisficing is another form of straight-lining (Zhang and Conrad, 2014). Researchers recommend excluding these participants from the analysis (Paas and Morren, 2018). However, Reuning and Plutzer (2020) contend that, in certain circumstances, providing identical responses to a series of questions may constitute an appropriate and justified course of action. These instances

demonstrate what Schonlau and Toepoel (2015) refer to as plausible straight-lining. For example, consider a respondent selecting "Extremely unlikely" for the question that asks how likely they are to respond to the phishing messages. This is the expected behaviour and does not portray a straight-lining behaviour. However, Schonlau and Toepoel (2015) cautioned researchers to be cautious when determining what constitutes plausible or non-plausible straight-lining behaviour, as it is challenging to distinguish between satisficing behaviour and best-effort responses. However, there appears to be an agreement within the research community that selecting the mid-point/neutral answer across the entire survey is evidence of a satisficing behaviour (Paas and Morren, 2018; Reuning and Plutzer, 2020). The recommendation is to remove these types of respondents (Paas and Morren, 2018). This approach was followed in this study, and as a result, 84 participants were excluded, reducing the number of participants to (n=138).

Further examination also revealed that some participants (n=22) selected the same response from either the negative or positive end of the scale. Specifically, (n=9) cases selected "extremely unlikely" and “not at all trustworthy” as their response to the measured variables. Another (n=2) selected “extremely likely” and “extremely trustworthy”. However, these participants justified their choices by responding to the follow-up questions. Thus, they were considered plausible straight-lining cases and were not removed. Some responses (n=11) were considered non-plausible and were excluded because these participants didn’t respond to any follow-up questions and, therefore, didn’t justify their choices. Specifically, (n=5) selected “somewhat likely” and “somewhat trustworthy”, (n=3) selected "extremely likely” and “extremely trustworthy”, and another (n=3) selected "extremely unlikely" and “not at all trustworthy” across the entire survey questions.

Table 6-6: Cross-tabulation between each of the hypothetical vignettes and click likelihood responses for participants who passed the attention check (n=222)

Count

		Click					
		1	2	3	4	5	Total
Vignette	1	14	30	127	40	11	222
	2	15	30	127	34	16	222
	3	19	32	126	30	14	221
	4	14	29	127	39	12	221
	5	27	35	123	26	11	222
	6	13	18	133	43	13	220
	7	14	7	125	61	15	222
	8	11	17	118	62	13	221
	9	17	23	123	44	14	221
	10	20	20	129	34	17	220
Total		164	241	1258	413	136	2212

Table 6-7: Cross-tabulation between each of the hypothetical vignettes and trust rating responses for participants who passed the attention checks(n=222)

Count

		Trust					Total
		1	2	3	4	5	
Vignette	1	34	40	111	26	11	222
	2	33	45	104	26	14	222
	3	39	50	102	17	14	222
	4	34	45	109	24	10	222
	5	51	44	106	11	10	222
	6	25	43	116	26	12	222
	7	18	31	131	29	13	222
	8	18	39	114	38	12	221
	9	39	31	117	24	11	222
	10	36	41	110	21	14	222
Total		327	409	1120	242	121	2219

6.4.2 Demographics of the selected participants

As mentioned in the previous section, the outcome of the data cleaning resulted in 127 participants whose responses were deemed qualified for analysis. Table

6-8 shows the demographics of these participants. The genders appear to be relatively balanced. However, most participants tend to fall within the age groups of 18-29 and 30-39 years old. Most participants held an undergraduate degree (n = 78) and a postgraduate degree (n = 25). Many resided in the United States (n = 61), followed by the United Kingdom (n = 51). The other category comprises participants from Afghanistan, Algeria, India, and Switzerland, with (n=1) each. Canada and France, with (n=2) each. Albania has (n=3) participants and Nigeria (n=4).

Table 6-8: Demographic Information of the Participants

		No. of users	Percentage
Gender	Male	59	46.5
	Female	67	52.8
	Missing	1	0.8
Age Group	18-29	66	52.0
	30-39	45	35.4
	40-49	8	6.3
	50–59	5	3.9
	60 and above	3	2.4
Level of Education	Secondary	5	3.9
	Further education (i.e. colleges)	17	13.4
	University undergraduate	78	61.4
	Postgraduate (e.g. Masters, Doctorate)	25	19.7
	Missing	2	1.6
Country of Residence	United Kingdom	51	40.2
	United States	61	48.0
	Others	15	11.8

6.4.3 Descriptive Statistics

Figure 6-4 shows the participants' reported likelihood of clicking the links across each hypothetical vignette. It shows that for most of the vignettes; the number of participants who intend to click the links in the phishing messages exceeds those exhibiting the desired behaviour of reporting an unlikelihood of following the links. For example, the most frequent selection (mode) was four for vignettes where the sender is a friend, family member, work/business colleague, or someone known but not a family member, friend, or business colleague. This suggests that the participants frequently selected somewhat likely as their response. On the other hand, the most frequent selection (mode) for the vignette where the sender was unknown and had no preview was two, suggesting that participants mostly selected somewhat unlikely options. Interestingly, the vignette where the sender was unknown but a link preview was available had a mode of three, which suggests that most participants were neither likely nor unlikely to click on the link.

Furthermore, Figure 6-4 shows that vignettes containing previews had more participants reporting an intention to click (i.e., selecting "somewhat likely" or "extremely likely") than those without previews. Specifically, when the vignette was from a friend and had no preview, 40.2% of the participants expressed the intention of clicking, compared to 44.1% when the vignette had a preview. For a vignette from a family without a preview, 39.4% indicated that they were likely to follow the link, compared to 59.8% when the vignette had a preview. For a vignette from a work or business colleague, 34.6% of the participants indicated a likelihood of clicking when the vignette had no preview, compared to 59% when the vignette had a preview. Similarly, for vignettes from known contacts who aren't family, friends or business colleagues, 40.1% of the participants indicated a likelihood of clicking when the vignette had no preview compared to 45.6% when the vignette had a preview. Finally, for the vignettes from unknown contacts, 29.2% indicated the likelihood of clicking when the vignette had no

preview compared to 40.2% when the vignette had a preview. These results suggest a potential impact of the link preview on susceptibility to phishing in WhatsApp. Furthermore, the analysis shows that the vignettes with the highest likelihood of clicks are those where the sender is a family member or a work or business colleague, and has a link preview.

The family-preview vignette impersonated IKEA and used homoglyphs as its URL construction technique. The work/business colleague Vignette impersonated Amazon and promised users an 80% discount on products, but the URL itself was fraudulent.

Although the focus of this study is on examining the influence of the type of relationship a user has with the sender of the message and the link preview on phishing susceptibility, further efforts were made to check if the participants' click ratings also differ by the companies impersonated and the message content. Table 6-9 shows that despite the variation in message content and companies impersonated, Vignettes six, seven, eight and nine, which contained previews, had the highest click ratings. These vignettes impersonated Google, IKEA, Amazon and Coca-Cola. More details about these vignettes can be found in APPENDIX C.1.

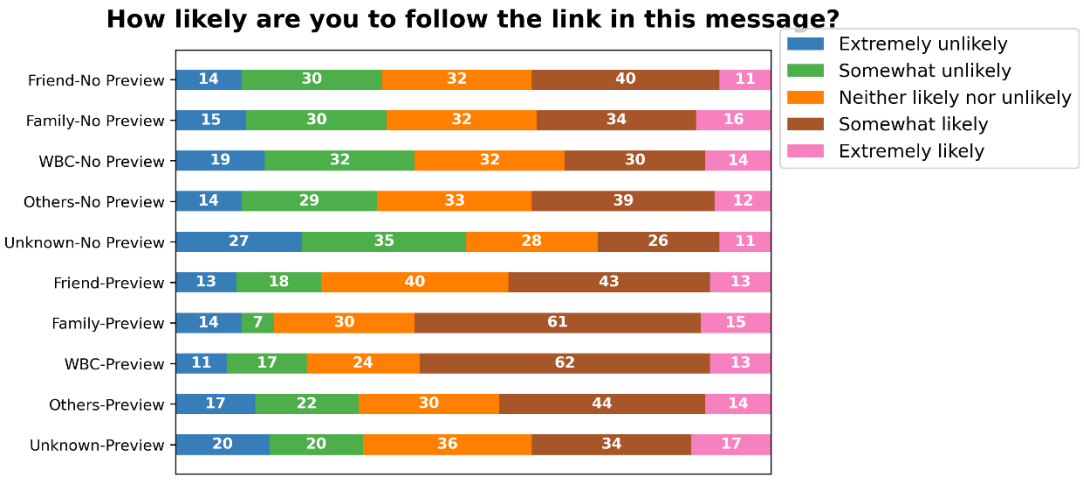


Figure 6-4: Participants click ratings across each hypothetical vignette

Table 6-9: Click ratings based on message content and companies

Vignette ID	Type of phishing message used	Company Impersonated	No. of users likely to click (sum of somewhat likely and extremely likely)	% of users likely to click (sum of somewhat likely and extremely likely)
1	Software download in exchange for enhanced functionalities	WhatsApp	51	40.2%
2	Free return ticket giveaway to two lucky winners	QatarAirways	50	39.4%
3	Discount Promo	Netflix	44	34.6%
4	Giveaway	Nestle	51	40.1%
5	Account confirmation	Binance	37	29.2%
6	Software download in exchange for a free Google Play gift card	GooglePlay	56	44.1%
7	Free Voucher	IKEA	76	59.8%
8	Discount Promo	Amazon	75	59%
9	Giveaway	CocaCola	58	45.6%
10	Payment verification	AirBnB	51	40.2%

The data was checked further using the Friedman test. The results revealed a statistically significant difference in the participants' click likelihood ratings across the different vignettes with a chi-square $\chi 2$ (9) =77.139 and p<.001. A subsequent paired analysis using the Dunn-Bonferroni post hoc test reveals that the observed difference where between 1) a vignette where the sender is unknown and the message has no preview, and vignette where the sender is a friend and the message has a preview, 2) a vignette where the sender is unknown and without a preview and a vignette where the message is from work/business

colleague the message has a preview, 3) a vignette where the sender is unknown and the message has no preview and a vignette where the message is from a family with a preview, 4) where the sender is work/business colleague with a preview and where the sender is work/business colleague but without a preview, and 5) vignette where the sender is a family and the message has a preview and vignette where the message is from work/business colleague and without a preview. This information is presented in Table 6-10.

Table 6-10: Dunn-Bonferroni Pairwise Comparisons for participants click ratings

Pairwise Comparisons

Comparison	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Unknown No Preview-friend_Preview	-1.319	.380	-3.471	.001	.023
Unknown No Preview -WBC preview	-1.854	.380	-4.881	.000	.000
Unknown No Preview-Family Preview	-1.969	.380	-5.181	.000	.000
WBC_No Preview-WBC_Preview	-1.319	.380	-3.471	.001	.023
WBC_No Preview-family Preview	-1.433	.380	-3.772	.000	.007

Asymptotic significances (2-sided tests) are displayed. The significance level is .050.

Figure 6-5 shows the participants' trust ratings of the hypothetical vignettes. Most participants trust the messages, as evidenced by many selecting either slightly trustworthy, somewhat trustworthy, very trustworthy, or extremely trustworthy. However, only a few appear to trust the messages extremely. This is similar to what was seen in Figure 6-4 where only a few were extremely likely to click. Figure 6-5 shows that some participants specifically rated the messages as either very trustworthy or extremely trustworthy. In particular, 28.4% rated the vignette from a friend, but without a preview, as trustworthy, compared to 29.9% when the vignette had a preview. For the vignette from a family member, 31.5% rated the message as trustworthy when it had no preview compared to 33% when the vignette had a preview. For the vignette from a work/business colleague, 23.6% rated the message as trustworthy when it had no preview, compared to 39.3% when the message had a preview. For vignettes from others known to the participants but not their friends, family, or work/business colleagues, 26.8% rated the message as trustworthy when it had no preview, compared to 27.6% when the message had a preview. Finally, for vignettes from unknown contacts, 16.6% rated the message as trustworthy when it had no preview, compared to 27.5% when the message had a preview. Similar to the responses on participants' intentions to click, these results suggest a potential impact of the link preview on trust ratings of phishing messages.

Further checks using the Friedman test revealed a statistically significant difference in the participants' trust ratings across the different vignettes with a chi-square χ^2 (9) = 86.202 and p<.001. Similar to the results for participants' click likelihood, subsequent paired analysis using the Dunn-Bonferroni post hoc test reveals that the differences were observed between 1) a vignette where the sender is unknown and the message has no preview, and vignette where the sender is a friend and the message has a preview, 2) a vignette where the sender is unknown and the message has no preview and a vignette where the message is from family with a preview,3) a vignette where the sender is unknown and the message has no preview and a vignette where the message is from work/business

colleague with a preview, 4) vignette where the sender is a family and the message has a preview and vignette where the sender is work/business colleague without a preview, 5) where the message is from work/business colleague the message has a preview and where the sender is work/business colleague without preview, 6) where the sender is known but is not a friend, family or work/business colleague and the message has a no preview and a vignette where the sender is family and the message has a preview 7) where the sender is known but is not a friend, family or work/business colleague and the message has a no preview and a vignette where the sender is work/business colleague and the message has a preview, 8) where the sender is a friend and the message has a no preview and a vignette where the sender is family and the message has a preview and 9) where the sender is a friend and the message has a no preview and a vignette where the sender is work/business colleague and the message has a preview and the message has a preview. This information is presented in Table 6-11.

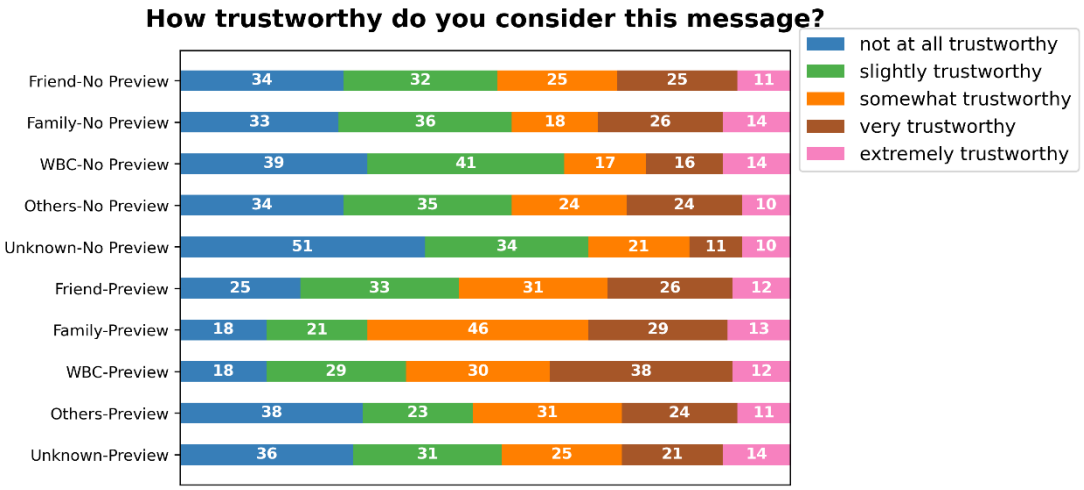


Figure 6-5: Participants' trust ratings across each hypothetical vignette

Table 6-11: Dunn-Bonferroni Pairwise Comparisons for the participants trust ratings

Pairwise Comparisons

Comparison	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Unknown No Preview-Friend Preview	-1.413	.380	-3.720	.000	.009
Unknown No Preview-Family Preview	-2.004	.380	-5.274	.000	.000
Unknown No Preview -WBC Preview	-2.094	.380	-5.513	.000	.000
WBC No Preview-Family Preview	-1.591	.380	-4.186	.000	.001
WBC_no Preview-WBC Preview	-1.681	.380	-4.425	.000	.000
Known No Preview-Family Preview	-1.413	.380	-3.720	.000	.009
Known No Preview-WBC Preview	-1.504	.380	-3.958	.000	.003
Friend No Preview-Family Preview	-1.256	.380	-3.306	.001	.043
Friend No Preview-WBC Preview	-1.346	.380	-3.544	.000	.018

Asymptotic significances (2-sided tests) are displayed. The significance level is .050.

The habitual WhatsApp usage construct was tested for internal consistency using Cronbach's alpha. The Cronbach's alpha value for the four-item construct is 0.637. This value falls within the acceptable range, as the literature recommends a value between 0.6 and 0.7 is required for the internal consistency of a construct's components (Churchill Jr, 1979; Pallant, 2007; Rahimnia and Hassanzadeh, 2013). The item total statistics value is presented in Table 6-12. Upon reviewing Table 6.12, it is evident that removing the first item, "Checking WhatsApp is part of my daily routine", would enhance Cronbach's alpha value to 0.65, approximating 0.7. This item also has a Corrected Item-Total Correlation value (the correlation between one item and the sum score of the other items) of 0.278. It is also important to note that this item was the only one adapted from (LaRose and Eastin, 2004).

Nevertheless, for the items to measure the same underlying construct, a Corrected Item-Total Correlation value of 0.3 or greater is recommended (Devriendt *et al.*, 2012). As a result, the first item was deleted because it failed the recommended checks. Therefore, from this point, each participant's habitual WhatsApp usage score is a composite score (sum or mean) of their ratings of items 2-4 of the habitual WhatsApp usage scale. Interestingly, the habitual WhatsApp usage construct had Cronbach's alpha value of 0.370 before data cleaning. This further supports the data-cleaning strategy adopted in this thesis.

Table 6-12: Item Total Statistics Habitual WhatsApp usage

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Checking WhatsApp is part of my daily routine	10.10	5.517	.278	.104	.651
I feel my WhatsApp use has gotten out of control	11.01	4.184	.352	.156	.633
I would miss checking my WhatsApp if I could no longer do it	10.42	3.990	.557	.406	.460
I find myself checking my WhatsApp (posting or checking other people's status updates) around the same time each day	10.37	4.428	.518	.381	.501

6.4.4 Hypothesis Testing

The descriptive statistics and the Friedman test in the previous sections suggest a likely impact of the user's relationship with the message sender and the link preview on participants' self-reported likelihood of clicking on phishing links during WhatsApp communication. These factors also appear to impact trust ratings of phishing messages. Further analysis using inferential statistics was conducted to test the study's hypothesis. Specifically, an ordinal logistic regression with Cluster-robust standard errors was fitted.

The choice of ordinal logistic regression was because the dependent variables were measured on ordinal scales (Jamieson, 2004). Furthermore, the fact that each participant responded to 10 different vignettes means that the observations were clustered among respondents (i.e. any 10 ratings from one participant are more likely to be similar than ratings from another participant). This violates the

assumption of independent observations required by standard regression techniques (Auspurg and Hinz, 2014). In addition, the variance of errors within clusters can vary (heteroscedasticity), introducing additional nuisance. Combining ordinal logistic regression with Cluster-robust standard errors will account for this and allow the model to adjust for the unequal variances caused by the clustering (Auspurg and Hinz, 2014). Using ordinal logistic regression enables the researcher to regress the participants' decision outcomes on the values of the vignette dimensions. The regression coefficients indicate which levels of the vignette dimensions were the most influential for participants' ratings.

Initially, an exploratory modelling step was conducted by fitting a one-predictor model using the primary variable, sender type. This approach was used to test the proportional odds assumption (e.g., using the Brant test) to determine if ordinal logistic regression is the appropriate model for the data. Table 6-13 shows the results of the regression analysis. On the top right of the logistic regression model, the number of observations in the dataset, the log likelihood ratio test statistic, the associated p-value, and the pseudo R^2 are shown. The dataset has 1,270 cases for the analysis. The associated p-value with the log likelihood ratio chi-square test is Prob > chi2 = 0.0014, which indicates that the one-predictor model provides a better fit than the null model with no independent variables in predicting the cumulative odds of being at or below a category of link click likelihood. The Pseudo R^2 = 0.0036, which is the likelihood ratio R^2 L, or McFadden's R^2 , suggests that the relationship between the response variable, click likelihood, and the predictor, sender type, is small.

The ordinal logistic regression table displays the Odds ratios, robust standard errors of the log-odds estimate, the 95% confidence interval for the odds ratio, the Wald z-statistics, and the associated p-values. The p-values across the various sender type categories are less than .05, implying that the participants were more likely to click messages from any of these categories than from an unknown sender. The 95% confidence interval of the regression coefficients across the

different categories does not contain zero, which indicates the coefficient is significantly different from zero. Next, the brant command is used to assess the proportional odds assumption in ordinal logistic regression. Introduced by Brant, (1990), the test evaluates this assumption by examining the individual fits of the binary logistic models that correspond to each threshold of the ordinal outcome. A non-significant omnibus test indicates that the proportional odds assumption holds. The test also provides assessments for each predictor. When the model includes only one independent variable, the results of the omnibus and individual tests are identical. The Brant test of the parallel regression assumption yields $X^2(3) = 2.97, p = 0.396$. This indicates that the proportional odds assumption for the model is satisfied, implying that the effect of the explanatory variable, sender type, remains consistent across the binary models fitted at each cumulative threshold. This analysis confirms that the ordinal logistic model is right for the data.

Table 6-13: Summary of one predictor model regression for participants' likelihood to click on links across the vignettes using unknown sender as reference category

				Number of obs = 1,270 Wald chi2(4) = 17.69 Prob > chi2 = 0.0014		
Log pseudolikelihood = -1936.3471				Pseudo R^2 = 0.0036		
Variable	Value	Odd Ratio	Robust std. err.	[95% conf. interval]	Z value	P- Value
Sender type	Friend	1.429	.1751398	[1.123 -1.817]	2.91	0.004
	Family	1.779	.2552041	[1.343 - 2.357]	4.02	0.000
	Work/Business Colleagues	1.533	.1800251	[1.218 - 1.930]	3.64	0.000
	Someone known but not friend, family or work/Business Colleague	1.388	.1604845	[1.107-1.741]	2.84	0.005

The significance of the model fit was reported at p<.05

After the initial analysis, four multiple ordinal logistic regression models were fitted to test the study's hypothesis. The independent variables were the participant's relationship with the message's sender, the link preview and the habitual WhatsApp usage score. The following sections present the results from the regression analysis.

Table 6-14 presents the estimated results for the first regression model, which seeks to test the hypotheses **H1a**, **H3**, and **H5**. For the independent variable, the participant's relationship with the message's sender, the reference category was the "unknown sender"; for the link preview, the reference category was "link without a preview".

To recap, hypothesis **H1a** posits that users will be more likely to report the intention to click on links in phishing messages sent by contacts they know than by unknown contacts. The estimated odds in Table 6-14 show that participants were more likely to click phishing links from known contacts than from unknown contacts. Specifically, participants were more likely to click on phishing links when a message was from a friend, family member, work colleague, or someone they knew who wasn't a family member, friend, or work colleague than when the message came from an unknown contact. For example, the odds of clicking a phishing link when a message is from a friend are 1.438 times higher compared to phishing links in messages received from unknown contacts. In other words, the odds indicate a 43.8% higher likelihood of opening phishing links in messages from a friend compared to an unknown sender. These estimated results support (**H1a**) of this study.

The third hypothesis (**H3**) posits that users would be more likely to click phishing links in messages with a link preview than those without a preview. Table 6-14 shows that the study participants were more likely to click phishing links in messages with a link preview than in messages without one. The results show that the odds of clicking a phishing link in messages with a link preview are 1.596 times higher compared to messages without a link preview. In other words, the

odds indicate a 59.6% higher likelihood of opening phishing links in messages with a link preview compared to messages without a link preview. Therefore, **H3** is supported.

The fifth (**H5)** hypothesis posits that individuals with higher self-reported habitual WhatsApp usage will be more likely to click on phishing links in WhatsApp. Please note that the habitual WhatsApp usage was measured as a continuous variable. Nevertheless, Table 6-14 shows that the odds of clicking phishing links in WhatsApp increase by 1.979 for every one-unit increase in habitual WhatsApp usage score. In other words, for a one-unit increase in habitual WhatsApp usage, the odds of clicking phishing links in WhatsApp increase by 97.9%. Therefore, **H5** is supported.

Overall, the estimated model's log-likelihood ratio chi-square test LR $\chi^2(6)$ = 61.58, p < .001, which indicates that the full model with six predictors provides a better fit than the null model with no independent variables in predicting the ordinal response variable. The Pseudo R^2 = 0.0324, which is the likelihood ratio R^2 L, or McFadden's R^2 .

Table 6-14: Summary of regression over participants' likelihood to click on links across the vignettes using unknown sender as reference category

Variable	Value	Odd Ratio	Robust std. err.	[95% conf. interval]	Z value	P- Value
Sender type	Friend	1.438	.1758547	[1.131-1.827]	2.97	0.003
	Family	1.813	.2620236	[1.366-2.407]	4.12	0.000
	Work/Business Colleagues	1.534	.1764742	[1.225-1.922]	3.72	0.000
	Someone known but not friend, family or work/Business Colleague	1.432	.1688902	[1.136 -1.805]	3.05	0.002
Prev Condition	Preview	1.596	.169703	[1.295-1.965]	4.39	0.000

Habitual Usage	Habitual score	1.979	.4831424	[1.227 - 3.194]	2.80	0.005

The significance of the model fit was reported at p<.05

To test **H1b,** which posits that users will be more likely to report the intention to click on links in phishing messages sent by family, friends, or work/business colleagues than from contacts they know who are not family, friends or work/business colleagues, an ordinal logistic regression was fitted using same independent variables as in Table 6-14. However, unlike the previous model, the reference category was the sender type "Someone you know but not family, work or business colleagues". In other words, the model compared the odds of clicking a phishing link in messages sent by family, friends, or work/business colleagues against messages from other contacts who are not family, friends or work/business colleagues. Table 6-15 presents the estimated results for this model. The results show that the odds ratios for the categories' friend' and 'work/business colleagues' are almost equal to one, and their respective p-values are all greater than 0.05. This indicates no significant difference in the participants' likelihood of clicking links in messages sent by any of these categories compared to those sent by other contacts who are not family, friends, or work/business colleagues. Furthermore, the 95% confidence interval of the regression coefficients across the categories' friend' and 'work/business colleagues' contains zero, which indicates the coefficient is not statistically significant at the 5% level.

However, there appears to be a significant difference in the participants' likelihood of clicking links in messages sent by family members and other contacts who are not family, friends or work/business colleagues. The model shows that the odds of clicking a phishing link when a message is from a family member are 1.266 times higher than for phishing links in messages received from other contacts who are not family, friends, or work/business colleagues. This indicates a 26.6% higher likelihood of opening phishing links in messages from a

family contact compared to other contacts who are not family, friends or work/business colleagues, which is not a very high percentage. For the relationship type "unknown", we see that the Z value is negative, indicating that the odds of clicking a phishing link when a message is from contacts who are not family, friends or work/business colleagues decrease by a factor of .698 when compared to links in messages from an unknown sender. This is expected, as the values in Table 6-14 indicate that participants were more likely to click on phishing links in messages from contacts who are not family, friends, or work/business colleagues than from unknown senders. The preview condition and habitual usage results remain the same as those in the model in Table 6-14. Overall, the results of Table 6-15 partially supported **H1b**.

Table 6-15: Summary of regression over participants' likelihood to click on links across the vignettes using other acquaintances as the reference category

Variable	Value	Odd Ratio	Robust std. err.	[95% conf. interval]	Z value	P-Value
Sender Type						
	Friend	1.004	.0956612	[.832-1.209]	0.04	0.968
	Family	1.266	.1331772	[1.0301.555]	2.24	0.025
	Work/Business Colleagues	1.071	.1063564	[.881-1.301]	0.69	0.488
	Unknown	.698	.0823378	[.554 - .879]	-3.05	0.002
Prev Condition						
	Preview	1.595	.169703	[1.2951.965]	4.39	0.000
Habitual Usage	Habitual score	1.979	.4831424	[1.2273.194]	2.80	0.005

The significance of the model fit was reported at $p<.05$

To test hypotheses **H2a**, **H4** and **H6**, a multiple ordinal logistic regression was fitted to examine the impact of the link preview, the relationship with the sender, and habitual WhatsApp usage on participants' trust ratings of phishing messages

in WhatsApp. The results in Table 6-16 show that the participants were more likely to trust phishing messages from known contacts than from unknown contacts. Specifically, participants were more likely to trust phishing messages from a friend, family member, work colleague, or someone they know who isn't a family member, friend, or work colleague than from an unknown contact. For example, the odds of trusting a message received from a friend are 1.550 times higher compared to messages received from an unknown contact. This translates to a 55% higher likelihood of trusting messages from a friend compared to an unknown sender. These results support (**H2a**) of this study**.** Furthermore, the results in Table 6-16 indicate that participants were more likely to trust messages with a link preview than those without one. The odds of trusting a message with a link preview are 1.546 times higher than those of a message without a link preview. This translates to a 54.6% higher likelihood of trusting messages with link previews than those without them. This result supports **H4**. For habitual WhatsApp usage, the results show that the odds of trusting a phishing message increase by 1.721 for every one-unit increase in habitual WhatsApp usage. In other words, the odds of trusting a phishing message increase by 72.1% for a one-unit increase in habitual WhatsApp usage. Thus, (**H6**) is supported. Overall, the model's log-likelihood ratio chi-square test, LR $X^2(6) = 82.84$, $p < .001$, which indicates that the full model with six predictors provides a better fit than the null model with no independent variables in predicting the ordinal response variable. The model's Pseudo $R^2 = 0.0245$, which is the likelihood ratio R^2 L, or McFadden's R^2.

Another ordinal logistic regression was fitted to test **H2b**, which posits that users are more likely to report higher trust ratings for phishing messages sent by family, friends, and work/business colleagues than for those sent by contacts they know who are not family, friends, or work/business colleagues. Table 6-17 shows the estimated results from the model.

The model compared the odds of trusting a phishing message sent by family, friends, or work/business colleagues against messages from other contacts who

are not family, friends or work/business colleagues. Similar to the result for click likelihood, Table 6-17 shows that the odds ratios for the categories' friend' and 'work/business colleagues' are almost equal to one, and their respective p-values are all greater than 0.05. This indicates no significant difference in the participants' trust ratings of messages sent by any of these categories and those sent by other contacts who are not family, friends or work/business colleagues. Furthermore, the 95% confidence interval of the regression coefficients across the categories' friend' and 'work/business colleagues' contains zero, which indicates the coefficient is not statistically significant at the 5% level.

However, there is a significant difference in the participants' trust ratings of messages sent by family members and those sent by other contacts who are not family, friends, or work/business colleagues. The model shows that the odds of trusting a message sent by a family member are 1.377 times more trustworthy compared to messages received from other contacts who are not family, friends, or work/business colleagues. This indicates a 37.7% higher likelihood of trusting phishing messages from a family contact compared to other contacts who are not family, friends or work/business colleagues. Table 6-17 also shows that when compared to messages received from an unknown sender, the odds of trusting a phishing message received from contacts who are not family, friends or work/business colleagues decrease by a factor of .733, implying that messages from known contacts who are not family, friends or work/business colleagues are rated as more trustworthy compared to those from unknown contacts.

Overall, the estimated results in Table 6-17 partially supported **H2b**. Table 6-15 summarises the results concerning the study hypotheses.

Table 6-16: Summary of regression over participants' trust ratings of the phishing messages

Variable	Value	Odd Ratio	Robust std. err.	[95% conf. interval]	Z value	P- Value
Sender Type	Friend	1.550	.1719692	[1.247 -1.926]	3.95	0.000
	Family	1.877	.2457582	[1.453-2.426]	4.81	0.000
	Work/Business Colleagues	1.594	.1697745	[1.294-1.965]	4.38	0.000
	Someone known but not friend, family or work/Business Colleague	1.363	.1653576	[1.075-1.729]	2.55	0.011
Prev Condition	Preview	1.546	.170597	[1.245- 1.919]	3.94	0.000
Habitual Usage	Habitual score	1.721	.4032389	[1.087 - 2.724]	2.32	0.020

Table 6-17: Summary of regression over participants' trust ratings of the phishing messages using other acquaintances as the reference category

Variable	Value	Odd Ratio	Robust std. err.	[95% conf. interval]	Z value	P- Value
Sender Type						
	Friend	1.137	.1142236	[.934 - 1.385]	1.28	0.200
	Family	1.377	.1514359	[1.110-1.708]	2.91	0.004
	Work/Business Colleagues	1.169	.129258	[.941 -1.453]	1.42	0.156
	Unknown	.733	.0889762	[.578 -.930]	-2.55	0.011
Prev Condition						
	Preview	1.545	.170597	[1.244- 1.919]	3.94	0.000
Habitual Usage	Habitual score	1.721	.4032389	[1.087 - 2.724]	2.32	0.020

The significance of the model fit was reported at p<.05

Table 6-18: Summary of findings in relation to study hypotheses

H1a: Users will be more likely to report the intention to click on links in phishing messages sent by contacts they know than by unknown contacts.	**Supported**

H1b: Users will be more likely to report the intention to click on links in phishing messages sent by family, friends, or work/business colleagues than from contacts they know who are not family, friends or work/business colleagues.	**Partially supported**
H2a: Users will be more likely to report higher trust ratings for phishing messages sent by contacts they know than for those sent by unknown contacts.	**Supported**
H2b: Users will be more likely to report higher trust ratings for phishing messages sent by family, friends, and work/business colleagues than from contacts they know who are not family, friends or work/business colleagues.	**Partially supported**
H3: Users will be more likely to report the intention to click on links in phishing messages with a link preview.	**Supported**
H4: Users will be more likely to report higher trust ratings for phishing messages with a preview.	**Supported**
H5: The higher the individual's self-reported habituation to WhatsApp, the more likely they are to report an intention to click on phishing links	**Supported**
H6: Higher self-reported habituation to WhatsApp will lead to higher trust ratings of phishing messages	**Supported**

6.4.5 Participants' reasons for indicating either the likelihood of clicking or not clicking on links in the phishing messages

As mentioned earlier in this chapter, follow-up questions were administered to participants who expressed their likelihood or unlikelihood of responding to the

phishing vignettes. These questions were designed to delve into the reasoning behind the participants' choices and provide a comprehensive view of their decision-making process. All 127 participants included in the analysis responded to the relevant questions. Since respondents were permitted to select multiple options, the frequencies depicted in the graphs represent the number of participants who endorsed each specific response category, rather than mutually exclusive counts, because participants could select multiple responses; hence, they might be counted in more than one category.

6.4.6 Reasons for selecting the likelihood of clicking

Figure 6-8 shows the responses obtained for vignettes with link previews. The most selected reason was that messages referenced a known organisation, followed by the participant's relationship with the sender. However, interestingly, fewer participants indicated that a relationship with the sender influenced their decision when the communicating party was someone they knew but not a family member, friend, or business colleague.

Furthermore, Figure 6-8 also shows that the presence of the logo was one of the reasons many participants indicated they would respond to the messages. The fact that the vignettes referenced a well-known organisation was also selected by many. It can also be seen that many chose the presence of a known organisation as their reason when the vignette was from family, work/business colleagues. Similarly, the sender appeared to have more influence over participants when the message was from either a family or work/business colleague.

It should be noted that for easier comparison, the responses for vignettes depicting an unknown relationship were excluded from Figure 6-8. This was due to the context of the vignettes (i.e., depicting a hypothetical scenario where an unknown contact claiming to be an employee of an organisation contacted the user). Instead, these were analysed separately, and the results show that despite these messages coming from unknown contacts (n=42, 33%) of the participants

were willing to respond to the messages because they came from an organisation they knew, whilst (n=31, 24%) were willing to click because the messages contained a logo.

Based on your previous response, select from the following, the reasons why you are likely or extremely likely to follow the link in the message. (Check all that apply)

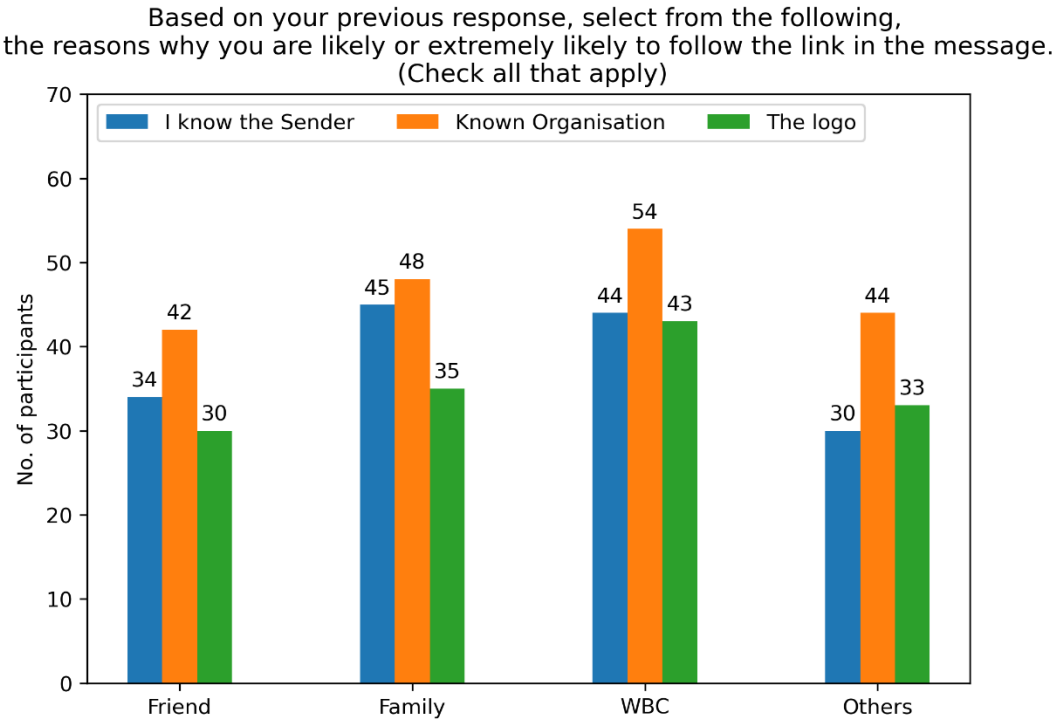


Figure 6-8: Participants' reasons for indicating a likelihood of clicking phishing links in messages with previews ("WBC" stands for Work/business colleagues, and "Others" stands for those the user knows but not friends, family or work/business colleagues)

Figure 6-9 shows the responses obtained for vignettes without previews. The results indicate that many participants were willing to click on the links because they came from known senders. There does not appear to be a significant difference in the number of participants who select different reasons across the various communicating parties. Similarly, although some participants chose the

presence of a known organisation as their reason, there does not seem to be a significant difference in their selection across the different communicating parties. Nevertheless, further analysis reveals that when the messages originated from unknown contacts, some participants (n=36, 28%) were willing to respond because the messages came from an organisation they knew.

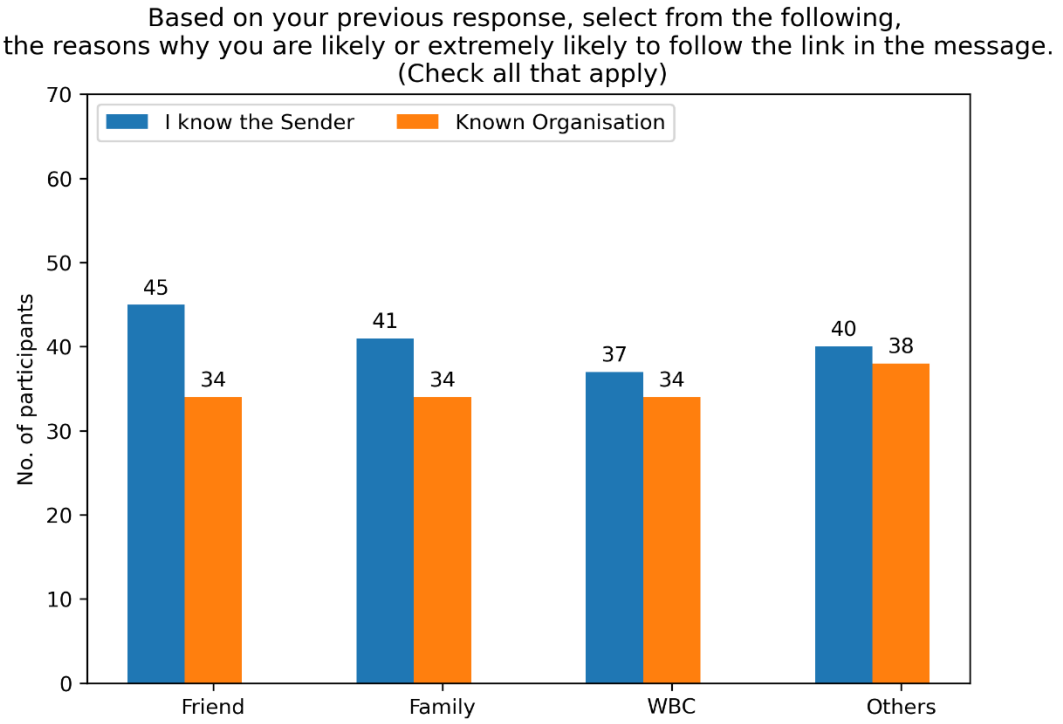


Figure 6-9: Participants' reasons for indicating a likelihood of clicking phishing links in messages with no previews ("WBC" stands for Work/business colleagues and "Others" stands for those the user knows but not friends, family or work/business colleague)

The results in Figure 6-8 and Figure 6-9 suggest that more participants were likely to respond due to the messages referencing a known organisation in vignettes with previews compared to those without previews. This is likely due to the presence of logos in the vignettes with previews, which suggests a potential impact of some components of link previews on users' likelihood of clicking phishing links in WhatsApp.

6.4.7 Reasons for selecting the unlikelihood of clicking

Figure 6-10 shows the participants' reasons for indicating an unlikelihood of clicking links in vignettes with previews. The data does not show any unique pattern. However, (n=24, 19%) of the participants said they do not respond to messages from those they know but who aren't their family, friends or business colleagues, with the reason being "because they know them". Interestingly, (n=18) participants had other reasons for not indicating they were likely to click on links sent by those whom they know but aren't their family, friends or business colleagues. Figure 6-10 does not show participants' responses to the vignette depicting an unknown sender. However, further analysis reveals that when the messages originated from unknown contacts (n=32, 25%) of the participants said they were less likely to respond because they didn't know the sender, confirming the regression analysis results.

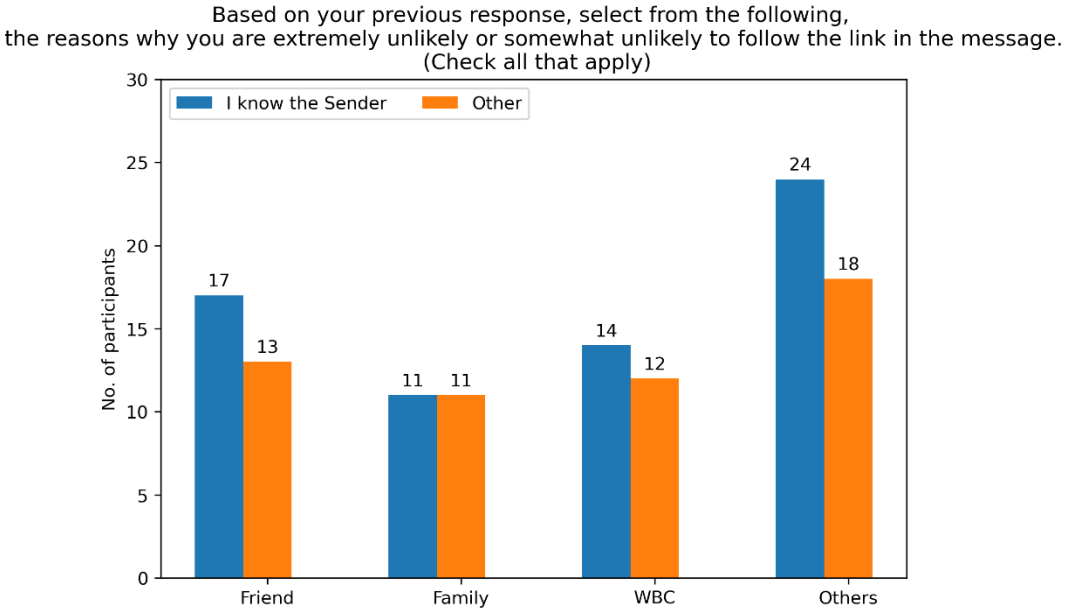


Figure 6-10: Participants' reasons for indicating an unlikelihood of clicking phishing links in messages with previews ("WBC" stands for

Work/business colleagues, and "Others" stands for those the user knows but not friends, family or work/business colleague)

Figure 6-11 shows the participants' reasons for indicating an unlikelihood of clicking links in vignettes with no previews across various relationships, including family, friends, work/business colleagues, and others. Interestingly, the absence of a logo was a significant factor behind the participants' decisions across the four relationship types. Furthermore, knowing the sender was also stated as a reason for not clicking. For messages from unknown contacts (n=47, 37.0%) were unlikely to click on such messages because they don't know the sender. This value is higher than that of the other category, where the message was from an unknown sender, but with a preview (n=32, 25%). Furthermore, (30, 23.6%) were unlikely to click on links in messages from unknown contacts due to the absence of a logo.

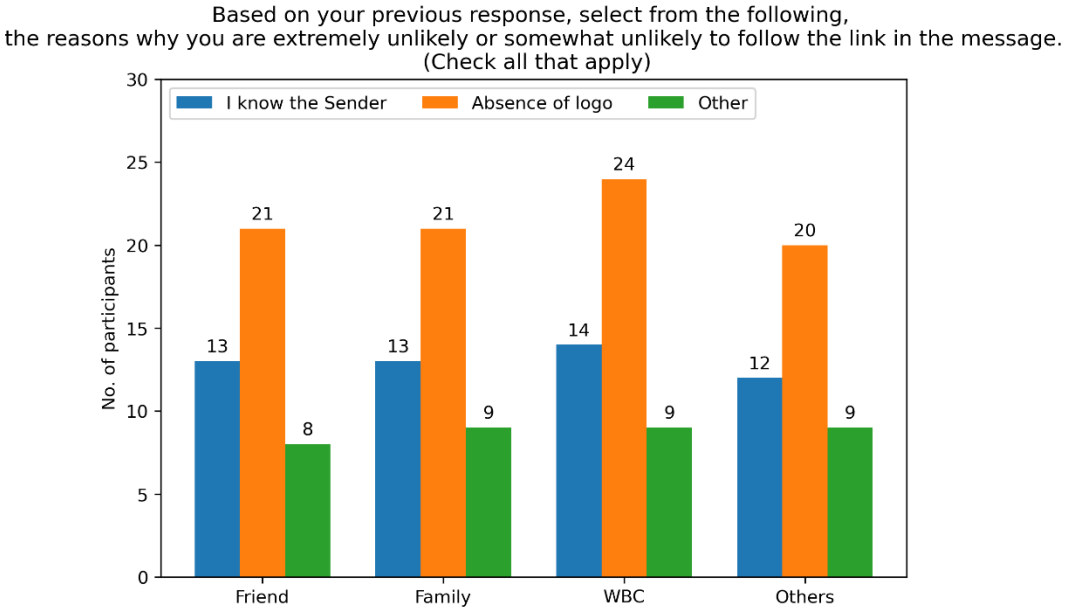


Figure 6-11:Participants' reasons for indicating an unlikelihood of clicking phishing links in messages without previews (" WBC" stands for

Work/business colleagues and "Others" stands for those the user knows but not friends, family or work/business colleagues)

6.4.8 What other factors did participants report as influencing their decision?

Open-ended responses relating to why the participants indicated either the likelihood or the unlikelihood of following the links in the vignettes were included as datasets for qualitative analysis. Deductive thematic analysis was employed to identify the themes that participants reported as influencing their decisions. Details of what deductive analysis is and the justification for adopting it in this research can be found in Chapter 3, Section 3.5 of this thesis. Nevertheless, the data were coded using the themes identified in the study (Williams, Morgan and Joinson, 2017; Williams and Polage, 2019). These included:

1. Use of influence techniques- used to indicate when a reference was made to the persuasive strategies that were utilised in the messages, including the use of authority by referencing known and trusted organisations, requiring individuals to take urgent action (urgency) and loss of some form (loss aversion)

2. Use of authenticity cues- reference to authenticity cues, such as spelling errors, grammatical errors, message inaccuracies, or inconsistencies.

3. Routine- reference to behavioural habits such as always declining some types of messages or always checking on the organisation's website to confirm before clicking links

4. Degree of knowledge- reference to lack of knowledge, such as not being sure whether a phishing message is phishing

In addition to the above, the following new themes were derived from the thematic analysis.

1. Trust in contacts- used to indicate when a participant referred to the phishing self-efficacy of their contacts, for example, a participant responded with "x *would never forward a commercial offer to me, and it looks like spam*", which indicates that the recipient believes the sender has the skills to detect phishing messages.

2. Unusual- used to indicate when reference is made to something not being the standard way of doing things. In this context, the individual may not have noticed other cues but only relied on the fact that the medium used to send the message was unusual. For example, one of the participants responded to a vignette depicting a giveaway scam imitating Qatar Airways with "*Qatar Airways would not send personal message to someone. They will usually communicate with members using authorised channels*"

Since the overarching aim of this phase is to understand what aspect of the vignettes may have influenced the participants' decision-making, the qualitative data were divided into responses relating to the likelihood of following the link in the message and those relating to the unlikelihood. Only (n=24) participants provided qualitative responses. Table 6-19 presents the frequencies (no of times) and percentages of the themes according to the participants' response choices (i.e., likelihood or unlikelihood). It is essential to note that a participant's response could reference multiple themes. Since every participant can provide a qualitative response to each of the ten vignettes, the counts and percentages in Table 6-19 represent the number of times a particular theme was referenced across the entire set of vignettes. Table 6-19 shows that participants were more likely to respond to the open-ended questions when they indicated an unlikelihood of clicking than when they indicated a likelihood. Furthermore, influence techniques were referenced more than other themes as the main reason

for the participants' decisions. Table 6-20 presents the most common categories of responses observed under each theme.

Table 6-19: Count and percentages of themes when respondents indicate a likelihood compared to when they indicate an unlikelihood

		Likelihood	Unlikelihood
Themes	Influence techniques	4 (3.3%)	36 (30.0%)
	Authenticity cues	1 (0.8%)	30 (25.0%)
	Routine/Behavioural habits	2 (1.7%)	21 (17.5%)
	Degree of knowledge	0 (0.0%)	6 (5.0%)
	Unusual	0 (0.0%)	12 (10.6%)
	Trust in contacts	0 (0.0%)	8 (6.7%)
Total		7 (5.8%)	113 (94.2%)

Table 6-20: Most common categories identified within each qualitative theme

Themes					
Influence techniques	**Authenticity cues**	**Routine**	**Knowledge**	**Contacts**	**Unusual**
Potential to win Suspicious of personal information request Suspicious of message	Untrusted links Grammatical errors Message not signed Genuine link	Avoid rewards/offers Respond to messages from known contacts only	Unsure Legitimacy Not interested	Not expecting such messages from contacts	Not a suitable communication method

Suspicious of reward Urgency		Use alternative verification Curiosity			

The use of influence techniques

Sixteen participants referenced the presence of influence techniques as the reason behind their decisions. These influence techniques were referenced 40 times (33.3%). This was either a case of successfully influencing the participants to indicate the likelihood of clicking (4, 3.3%) or failing to persuade them, in which case they indicated an unlikelihood of clicking (36, 30.0%). Furthermore, influence techniques were referenced 25 times in response to vignettes with previews with logos and 15 times in response to vignettes without previews. Time pressure/ urgency was stated as the reason for influencing the participants. For example, one participant wrote," I don't want my booking to be cancelled". This was in response to the vignette depicting a booking cancellation communication from Airbnb. Another participant also stated the opportunity to access a potential benefit as the reason behind their decision: "I have the potential to win". This statement shows that the participant has succumbed to the social proof principle, a known persuasive principle used in phishing scams. However, Table 6-19 also shows that in 36 cases (30.0%), the influence cues could not have the desired effect. For example, one participant responded, "The technique used in this method is a common scam method", and another said, "The mail is trying to force action immediately, which usually signifies a scam". These responses demonstrate how the influence techniques failed to deliver the desired outcome.

The use of authentication cues

Reference to authenticity cues was the second most observed within the qualitative data. Table 6-19 shows that 25.8% of the responses referenced authenticity cues, with 1 (0.8%) leading to a likelihood of clicking and 30 (25.0%)

leading to an unlikely response. The participant who indicated a likelihood to respond wrote "*Genuine link*" as their reason. In contrast, those who indicated a lack of willingness to respond referenced the link as being untrustworthy and fake. For example, one respondent said, "*because a url looks like standard phishing attempts*", and another said, "*Looks like an unusual link*". Others referenced the grammatical errors in the message "*Text and grammar of message- sounds like a scam*". Another participant thinks legitimate messages should be signed. This participant wrote "*Message not signed*" as the reason for not clicking.

Individuals' habits and routines

Reference to routine behaviour and habits was the third most common within the qualitative data. The routine behaviour reflects how the participants typically handled messages in WhatsApp, sharing details about their customs, habits, and behaviour. Table 6-19 shows that 19.2% of the qualitative responses referenced routine behaviour, with 2 (1.7%) leading to a likelihood of clicking phishing links and 21 (17.5%) leading to an unlikelihood. One participant who indicated a likelihood of clicking stated, "*In the past, this kind of activity was also frequent*". This participant may have seen similar rewards/offers messages from other organisations in the past and therefore believes such offers are legitimate. Those participants who indicated a lack of willingness to respond said they tend to avoid rewards/offers of this kind. For example, one participant said, "*I am weary of free things. Most are scams if not all*". Another participant said, "*almost every message with freebies comes with some risks*". Other participants also claimed that they always use alternative verification strategies when they receive such messages, "*I'd follow up on the Binance site to see if it was valid*". Two participants said they don't click on links on WhatsApp: "*I do not click on links on WhatsApp*". An interesting observation shows that some users indicated a likelihood of clicking not because they trusted the messages, but out of curiosity. For example, one user wrote "*inquisitive*" to indicate a likelihood to click. Another user stated an unlikelihood to click but wrote, "*Not sure if genuine but may click on it out of curiosity/because nothing stands out to me as being obviously suspicious*".

Other themes

Although not widely referenced, the data revealed codes indicating participants' knowledge and trust in their contacts. It was observed that some participants lacked the necessary knowledge to discern the fraudulent nature of the messages. For instance, one participant said, "*Not that interested in getting the gift, not sure if it's genuine*". Another participant's response was, "*Ru is Russia, I don't drink cola, seems random*". However, another finding is that some participants believe their contacts should be able to detect spam messages, inferring a strong belief in their contacts' phishing self-efficacy. For example, one participant said, "x *would never forward a commercial offer to me, and it looks like spam*", and another wrote, "*I don't travel. It doesn't sound at all like the sort of thing this person would send*". Responding to the vignettes depicting work/business colleagues, some participants referred to the security culture of their organisation. For example, "*Wouldn't expect this content from a work colleague*" and "*My work colleagues know better than to send spam. I'd raise it to work IT as a security concern*". Another observation was that some participants believed communication on WhatsApp is personal, and therefore, it is unusual for businesses to contact their customers through such apps. For instance, one participant responded, "*Qatar Airways would not send personal message to someone. They will usually communicate with members using authorised channels*". Another participant said, "*It comes through [WhatsApp] instead of Airbnb website messages system or email*".

6.5 Discussion

In this section, the findings of this chapter are discussed, and a connection is drawn between these findings and previous research on phishing susceptibility.

6.5.1 Relationship with the Sender

The findings from this study revealed evidence of an impact of the relationship with the message sender on susceptibility to phishing on WhatsApp. Participants in this study were notably more inclined to click on phishing links when the sender was known to them, aligning with prior literature on email phishing (Jagatic *et al.*, 2007; Moody, Galletta and Dunn, 2017). This can be inferred from the descriptive statistics and the number of significant effects in the regression model. In addition, analysis of the follow-up questions also showed that for most participants, the main reason for indicating a likelihood of clicking was that the message came from someone they knew, supporting **H1a** of this study. These findings align with a study conducted almost two decades ago, which found that internet users were four times more likely to click on phishing links in messages impersonating a known acquaintance (Jagatic *et al.*, 2007). The findings of this study also show that the participants were more likely to trust phishing messages from known contacts than from unknown contacts, supporting **H2a**. The optional follow-up questions also confirmed these findings, as they showed that more participants were willing to click on the links in the messages because they came from those they knew. Specifically, more participants said so when the messages came from family, friends, or work/business colleagues.

Nevertheless, the findings of this study didn't fully support hypotheses **H1b** and **H2b,** where the expectation was that the participants would be more likely to report the intention to click on links in phishing messages sent by family, friends, or work/business colleagues than from contacts they know who are not family, friends or work/business colleagues and that they would find phishing messages sent by family, friends, or work/business colleagues as more trustworthy than contacts they know who are not family, friends or work/business colleagues. Although these hypotheses are not fully supported, the regression results suggest that the participants were more likely to report an intention to click on links in messages or trust messages from family than from contacts who are not family,

friends, or work/business colleagues. This is similar to the findings of (Seng *et al.*, 2019), who conducted a lab study with a mock-up Facebook interface with 30 participants and found that the study participants were more likely to click suspicious posts shared on Facebook by family members than posts shared by other contacts who aren't family or friends. However, Seng *et al.* (2019) also found that the study participants were more likely to click on suspicious posts shared on Facebook by their friends. This finding differs from those of the current study. One possible explanation for this difference is that the authors of Seng *et al.* (2019) controlled for the variation in relationship type by using terms such as "Parent" to represent family members and "Best friend" to represent friends. The current study didn't employ this approach but asked users to provide pseudonyms for each relationship type.

Furthermore, the authors of Seng *et al.* (2019) did not provide details of the kind of messages participants were tested on, unlike in the current study, where participants were exposed to free voucher and gift card scams. Thus, it is likely that the content of the messages used in the current study impacted the participants. For example, the descriptive statistics showed that the vignette, where the message impersonated Amazon and was sent by a work/business colleague, promising users an 80% discount on products, received the highest trust ratings from the participants, despite containing a fraudulent link and cues that should have ignited suspicion. However, the participants likely rated this message as highly trustworthy because it resembles a familiar type of communication from retailers on Amazon.

Furthermore, the vignette, in which the message impersonated IKEA and was sent by a family member, was the second most rated. This vignette used the URL construction technique of homoglyphs. In homoglyph attacks, cybercriminals register Unicode domain names that closely resemble real ones but redirect users to different servers. These attacks are challenging for internet users to detect, as evidenced in the user study conducted by Thao *et al.* (2020).

Nevertheless, the fact that there is a significant effect between certain categories of relationships and participants' decisions to click and trust phishing messages suggests that participants are likely relying on a social confirmation or social proof approach. In this approach, credibility is established through the actions and beliefs of others and depends on endorsement heuristics (Metzger, Flanagin and Medders, 2010). This approach is especially relevant in MIM apps such as WhatsApp since users will likely receive forwarded messages from known and trusted contacts. In fact, evidence from the qualitative content analysis study in Chapter 5 of this thesis has shown that cybercriminals exploit the trust users have in their contacts by asking them to forward phishing messages to them. In this regard, phishers can propagate their campaigns by exploiting users' lack of knowledge. The evidence from this study also confirms that some participants were not paying attention to the obvious fraudulent links, but rather used their relationship with the sender and other cues within the messages to verify the message's legitimacy. This demonstrates that, despite decades of security awareness training that often emphasised the importance of checking links before clicking, typical internet users still do not follow these recommendations. However, it is also likely that the participants were unaware of the tricks used by scammers to construct fraudulent URLs, and therefore, could not parse the URLs correctly or identify the phishing cues.

While the relationship with the sender does appear to influence users to respond to phishing messages in WhatsApp, it can also have the opposite effect (i.e., influencing users not to respond). This was evidenced in the responses to follow-up questions, where knowing the sender was stated as the reason why some participants were unlikely to respond to the phishing messages. The qualitative data also show that some participants were less likely to click, especially when the contact was a work/business colleague, because they do not expect such types of messages from their work colleagues. In this regard, the participants' decision not to click is not an indication that they paid any attention to the suspicious cues in the messages; instead, the messages may have appeared out of

context. This is likely due to the varying communication contexts of different contacts. Therefore, messages out of context are considered suspicious while those in context are deemed legitimate.

6.5.2 The presence of a link preview

As highlighted in section 2.2.2, MIM apps use the link preview mechanism to summarise the content of a shared webpage by displaying the shared page's title, description, logo/images, and domain name. However, section 2.2.2 also pointed out that those with malicious intent can manipulate the content of a link preview. Section 6.3.2 of this thesis highlights that the researcher manipulated the link preview section of the vignettes by inserting both familiar brand names in the domain name section of the link preview and logos/images of familiar brands inserted in the image section of the link preview.

The results of this study showed that participants were more likely to click phishing links in messages with a link preview than in messages without one. Similarly, messages with a link preview were rated as more trustworthy than those without one. However, it is not clear which components of the link preview influenced the participants because, as mentioned in 2.2.2, the preview contains the page's title, description, logo/images, and domain name. However, the follow-up questions show that the presence of the organisation's logo was one of the reasons why participants reported the likelihood of clicking. Furthermore, familiar brand names were also indicated as a reason for expressing the likelihood of clicking phishing links. These findings suggest that the participants evaluated the legitimacy using design cues. This is consistent with existing knowledge in the literature on email phishing. Specifically, several studies have shown that participants rated emails with logos, copyright information, and familiar brand names as more trustworthy (Jakobsson *et al.*, 2007; Williams and Polage, 2019). According to Williams and Polage (2019), a possible explanation for this is that these cues tend to invoke stereotypes and expectations about what

a legitimate message should look like, thereby preventing users from paying attention to the actual suspicious cues. Research conducted a decade ago showed that participants wrongly classified legitimate emails as phishing because they lacked logos (Parsons *et al.*, 2013), indicating that the absence of design cues may violate integrity expectations on the part of users. The current study's findings showed that, irrespective of the communication medium, users' stereotypes and expectations of a legitimate message remain the same. These stereotypes and expectations contribute to users' heuristics, enabling them to avoid cognitively demanding tasks such as checking the URL, making them susceptible to phishing attacks. Indeed, when users encounter phishing messages in WhatsApp, the most important cue is the URL itself, as other aspects of the message, such as the link preview, can be manipulated, allowing cybercriminals to include logos of impersonated brands, titles, and brand names. The manipulation itself is a trivial task, as it requires only a few changes to the HTML tags of the fake webpage or the meta-tags.

While MIM's capability to provide link previews has transformed the presentation of links in such applications by improving the visual aspects of communication, it simultaneously illustrates how a seemingly simple feature can pose security vulnerabilities, as evidenced by this study's findings. The problem is rooted in how the apps generate link previews. Consequently, users can do little apart from being vigilant and correctly parsing URLs. However, we know from the literature that correctly parsing URLs is often challenging, even for tech-savvy users (Albakry, Vaniea and Wolters, 2020). Moreover, the small screen size of smartphones is considered a barrier to users' ability to parse a URL correctly on smartphones (Goel and Jain, 2018). As a result, parsing URLs can be more challenging for users of MIM apps. Thus, it is not logical for the platforms to rely on users to perform such tasks.

6.5.3 Habitual WhatsApp usage

The findings indicate that habitual WhatsApp usage is significantly associated with an increased likelihood of responding to phishing messages on WhatsApp and rating such messages as trustworthy. This supports previous findings by Williams and Polage (2019) and Vishwanath (2015) that email and Facebook usage habits can increase participants' susceptibility to phishing attacks. When using a particular medium becomes habitual, individuals tend to automatically respond to communication in such a medium without reflecting on their decision before enacting it (Vishwanath, 2015). Automatically responding to communication is linked to heuristic processing or mental shortcuts (Vishwanath, 2016). Often, individuals with such habits find the systematic processing of information effortful, resulting in a search for an easier way. This means paying less attention to the obvious phishing cues that should have triggered their suspicions. The findings in this study show that irrespective of the type of media, individuals with such habits are more susceptible to phishing compared to those without such habits.

6.5.4 Qualitative responses

Regarding participants' reasons for indicating a likelihood of responding, there was evidence that the influence technique of time pressure/ urgency successfully convinced some participants to respond. Time pressure/urgency is not new to phishing on WhatsApp, as such methods are widely used in email phishing (Akdemir and Yenal, 2021). Similar to the current study, the findings in Williams, Morgan and Joinson (2017) showed that urgency cues successfully influenced users' responses to phishing emails. However, despite effectiveness, influence techniques sometimes fail to generate the desired impact. In such cases, they are referred to as failed influence techniques (Williams, Morgan and Joinson, 2017). The qualitative data show that some participants were suspicious of the pretext used in the vignettes. This may be because the messages relate to giveaway

campaigns. Typically, these types of messages should raise user suspicion. Yet, many participants thought they were trustworthy and were willing to follow the links in them.

Similar to the influence techniques, the authenticity cues made some participants suspicious, rather than persuading them. Some participants detected the links as suspicious, whilst others pointed to grammatical errors in the messages. Whilst these groups of participants showed some level of knowledge regarding their ability to spot the fraudulent aspects of the messages, others were unable to do so despite the apparent fraudulent cues. These findings suggest that knowledge influences phishing susceptibility, aligning with previous research findings (Graham and Triplett, 2017). Furthermore, the qualitative responses also indicate that some participants employ cyber hygiene practices to protect themselves from phishing. These include using alternative verification to confirm the legitimacy of communication, not responding to messages from unknown contacts and avoiding rewards and offers. This approach aligns with current advice regarding phishing in MIM apps (Kaspersky, 2021). However, the qualitative data suggest that some participants may have responded out of curiosity. These individuals are referred to as curiosity clickers and are more susceptible to phishing attacks, as they are more likely to disclose their details when clicking on fraudulent links (Kuraku, 2022).

6.6 Limitations

One drawback of this study is that the number of participants who passed the screening test was relatively small. As a result, the analysis was based on a small sample size. However, stringent measures were adopted to ensure the quality of the responses. Furthermore, because participants were recruited from specific online groups and forums, most were residing in the US or the UK. This has the potential to impact the generalisability of the study's findings. Furthermore, this study employed an experimental online vignette survey methodology to systematically examine the impact of the user-sender relationship and the

presence of link preview on the user's intention to click on a phishing link. This approach presents a more realistic scenario compared to typical questionnaire items (Aguinis and Bradley, 2014). However, the use of experimental online vignette survey methodology to assess participants' susceptibility also has a limitation in that it relies on hypothetical scenarios, which may not accurately capture real-world scenarios. However, the participants in this study were not informed that the study focused on phishing; therefore, the limitation associated with measuring the natural behaviour of the participants was mitigated. However, gaining deeper insights may necessitate the use of qualitative data collection approaches, such as online observations or interviews with users of MIM apps as they interact with sample MIM phishing attacks. Furthermore, it is possible that the users were influenced by other susceptibility factors, particularly the content of the messages. As a result, future studies should investigate the impact of message content and organisational types on susceptibility to phishing in MIM apps. Another threat to the validity of this study is ecological validity, as it wasn't conducted on mobile devices or during typical mobile use.

6.7 Chapter summary and conclusion

This chapter presents the findings of an experimental vignette survey investigating individuals' susceptibility to phishing in WhatsApp by examining how individuals evaluate suspicious WhatsApp messages, specifically the likelihood of clicking phishing links received from different communicating parties and the extent to which individuals trust these messages. Overall, despite the messages being typical phishing messages that should have ignited suspicion, the self-reported data indicate that more participants believed the messages were trustworthy and indicated a likelihood of clicking the fraudulent links, indicating their susceptibility to such attacks. However, some factors seem to influence the participants' self-reported responses. Specifically, it shows that the relationship between the sender and the recipient significantly shaped the

participants' responses to phishing messages, with participants being more likely to click on phishing links when the sender was known to them. These results are similar to those found in Chapter 4, which show that MIM app users were more inclined to click on links they received from those in their social circle. Overall, the conclusion that can be drawn from these results is that the participants implicitly or explicitly rely on people in their social circle to protect them from phishing. This approach makes them susceptible to phishing, considering the results of Chapter 5, which show that cybercriminals leveraged relationships to enhance phishing credibility.

The self-reported data also shows that the participants were more likely to click phishing links and trust the message when a link preview was displayed, compared to messages without a preview. The participants' reliance on authentication cues such as the link preview, indicates that phishers can easily deceive MIM app users by manipulating the elements of a link preview. Another finding is that habitual WhatsApp usage was significantly related to an increased likelihood of responding to phishing messages on WhatsApp, confirming earlier findings on email phishing, which shows that habituation increases susceptibility to phishing. These findings have both practical and research implications. These implications will be discussed in the next chapter.

Chapter 7

General Discussion and Conclusion

This chapter presents the main conclusions that can be drawn from the previous chapters. The researcher aims to synthesise the findings from the various studies in this thesis to present a comprehensive view of phishing in MIM apps. Subsequently, the chapter expands to include proposals for addressing the problem of phishing in MIM apps, providing advice for security awareness professionals and MIM platform providers. The chapter subsequently discusses the limitations of the research before suggesting avenues for future research.

7.1 Research problem restated

In recent years, MIM apps such as WhatsApp, Viber, Telegram, and Facebook Messenger have emerged as dominant platforms for communication. However, the increasing reliance on these applications has also attracted malicious actors, using them to propagate phishing campaigns. Unlike conventional platforms, MIM applications possess distinctive contextual and technical characteristics that can fundamentally reshape the phishing strategies and techniques used by attackers and the vulnerabilities experienced by users. The unique contextual characteristics of MIM apps justify further investigation into how phishing tactics and user behaviours are shaped and influenced within this environment. This thesis addressed a research gap in that the impact of the unique contextual characteristics of MIM apps on phishing susceptibility has not been rigorously studied in the literature.

7.2 Major findings

This section of the thesis presents the key findings and uses cross-cutting themes to link the three studies together.

7.2.1 Trust as a Central Vulnerability in MIM phishing

Across the three studies, trust emerges as a fundamental vulnerability in MIM phishing. In the first study, the self-reported data indicate that MIM apps users frequently click on links shared by friends, family, and colleagues. They also forward links received from private groups rather than those from public groups. Similarly, the self-reported data from the third study revealed that, despite the messages being typical reward and gain phishing scams, the participants were more likely to trust and click messages sent by friends, family members, work/business colleagues, and other acquaintances than messages from unknown contacts. These findings, therefore, show that the study participants perceived phishing links originating from individuals within their social circle as originating from trusted sources, reflecting a reliance on interpersonal trust as a proxy for legitimacy. This behaviour reflects a critical vulnerability such that when malicious links are disseminated through trusted contacts, users' reliance on source credibility heuristics substantially reduces their likelihood of engaging in critical evaluation of messages, further increasing their risk of phishing. However, findings of the third study also demonstrate that while interpersonal trust remains a dominant heuristic in MIM phishing, its influence is moderated by relational expectations and message context. For example, despite quantitative data revealing a higher likelihood of clicking and trusting messages from known contacts, the qualitative data show that trust can reduce clicking when messages are "out of character" (e.g., work contacts sending reward scams), indicating that trust operates contextually rather than uniformly across social ties.

The qualitative content analysis of sample phishing messages conducted in the second study also shows that phishers actively exploit the user-sender relation and source credibility when crafting phishing messages. The findings revealed that phishers frequently integrate persuasion strategies, such as social proof, liking, and authority, and reinforce them via deceptive technical features, such as manipulated link previews and carefully constructed URLs. These findings demonstrate that cybercriminals exploit user trust and reliance on source credibility heuristics by deliberately constructing URLs with familiar organisational cues, brand names, and socially recognisable terms within complex URL structures. These cues align with users' heuristic processing strategies, whereby legitimacy is inferred from recognisable elements rather than verified through careful examination of the full URL. When combined with deceptive link previews, these URL construction techniques reinforce the perception that the link originates from a trusted and socially endorsed source, deceiving users into responding to fraudulent messages. Furthermore, the findings of the second study also show that phishers continue to exploit the nature of human social interactions, capitalising on the liking principle, which relies on the notion that humans believe and trust those they like or share similarities with. These findings revealed that phishers are aware of the informal nature of MIM apps and are exploiting connections between individuals to propagate their campaigns.

Overall, the findings from the three studies in this thesis revealed that trust, especially trust derived from social relationships, systematically reduces critical evaluation of fraudulent messages.

7.2.2 The Socio-Technical Nature of MIM Phishing

The findings across all three studies demonstrate that phishing in MIM apps is an inherently socio-technical phenomenon, shaped by the interaction between user behaviours, persuasive message design strategies, and platform affordances,

including how link previews are generated and link forwarding or sharing functionalities. The results of the first study revealed that users engage with links in ways shaped by their social relationships and habitual interaction. These behavioural patterns create an environment in which technical elements such as URLs are rarely scrutinised in detail, risking reliance on heuristic processing, where users rely on simple decision-making rules. Findings from the second study illustrate how cybercriminals design URLs to exploit these behavioural patterns, employing complex URL structures and embedded organisational cues that are likely to trigger heuristic processing by users. This technical manipulation of URLs is further amplified by platform features such as link previews, which visually summarise the content of the webpage but can also shift users' attention away from the underlying URL. As shown in the second study, these previews can be manipulated to display the logo of the impersonated organisation and domain name, effectively masking the deceptive nature of the URL itself. The second study further shows that cybercriminals deliberately tailor phishing messages to align with the informal and conversational nature of MIM apps, embedding phrases that evoke a sense of shared risk by suggesting that other users have already undertaken similar actions. This shows that phishers are aware of the tendency of MIM app users to infer the legitimacy or safety of an action based on the behaviour or endorsement of others.

The third study empirically confirms that phishing susceptibility in MIM apps emerges from the interaction between technical obfuscation techniques (e.g. homoglyph URLs and manipulated link previews), social relationships, and habitual usage patterns. These findings reinforce the inherently socio-technical nature of MIM phishing.

Taken together, the findings from the three studies in this thesis indicate that phishing in MIM apps constitutes a socio-technical problem, encompassing both technical factors related to the platform design and social interaction patterns associated with how users engage with the platforms' features.

7.2.3 Users as Unintentional Spreaders of Phishing Attacks Within MIM Apps

A recurring theme across the studies is the role of users as unintentional spreaders of phishing attacks within MIM apps. Analysis of the self-reported data from the first study indicates that participants exhibit a pronounced tendency to forward links, particularly within trusted social networks. This behaviour not only facilitates the amplification and diffusion of phishing campaigns within MIM apps but also extends the attack surface beyond the initial recipient, thereby increasing the likelihood of secondary victimisation among users and their contacts. Importantly, such behaviours do not occur in the absence of external influence; findings from the second study demonstrate that phishers deliberately deploy persuasive strategies, notably the social proof and liking principles, to explicitly encourage users to forward malicious links and to imply collective participation in the purported action. Phisher's inclusion of phrases such as "I have received mine" was a deliberate attempt to mislead users into believing that the senders of the message benefited from the scam, thereby encouraging users to respond and potentially share the scam with others. Furthermore, phishers' ability to create benign-looking link previews may also facilitate onward sharing, since findings from the third study have shown that the participants were more likely to trust and click phishing links in messages with a link preview than in messages without one. Furthermore, the third study also shows that higher habitual WhatsApp usage is associated with a higher likelihood of responding to phishing on WhatsApp. This behavioural pattern can lower the ability of users to scrutinise links, increasing the chances of forwarding such messages to others.

7.3 Contributions

The research contributes to the knowledge by expanding previous phishing research. Specifically, it advances the theory on user susceptibility to phishing by

contributing to a greater understanding of the phenomenon through mobile instant messaging applications. The following paragraphs highlight specific contributions of this thesis.

First, this thesis demonstrates that phishing in MIM apps is an inherently socio-technical problem rather than a predominantly technical or behavioural problem. Whilst prior literature has often examined user susceptibility to MIM phishing from a purely user behavioural perspective (Lee, Gan and Liew, 2023b, 2023a), the findings of this thesis demonstrate that phishing susceptibility in MIM apps emerges from the interaction between user behaviours, persuasive message design strategies, social relationships and platform affordances. This contribution adds to the existing research on phishing susceptibility. These findings also highlight that neither technical defences nor user awareness alone can sufficiently explain or mitigate phishing risks in MIM environments, but a combination of both approaches can.

Second, this thesis contributes theoretically by extending trust-based explanations of phishing susceptibility beyond the binary distinction of known versus unknown senders. The thesis revealed unique empirical results that underscore the user-sender relationship as a significant factor in susceptibility to phishing in MIM apps. It highlights how interpersonal trust serves as a proxy for inferring the legitimacy of phishing messages but also shows that trust is sometimes context-dependent, influenced not only by the sender's identity but also by expectations surrounding communication context (e.g. work-related versus informal interactions). These findings further enhance our understanding of trust exploitation in phishing attacks.

Third, this thesis makes a significant theoretical contribution by demonstrating how platform affordances, particularly link previews, actively shape phishing susceptibility in MIM apps. Apart from serving as interface features, this thesis shows that link previews function as persuasive artefacts that influence trust judgements and behavioural intentions. This contribution highlights how design

features intended to enhance usability can be systematically repurposed by attackers to facilitate deception. Thus, this thesis extends the literature on the impact of authenticity cues on phishing susceptibility (Williams and Polage, 2019) but also highlights how some software design and development practices can undermine the efforts of cybersecurity professionals in the fight against cyber threats.

Fourth, this thesis contributes to the existing research on phishing susceptibility by highlighting habitual platform use as a significant factor shaping phishing susceptibility in MIM apps. The findings demonstrate that frequent, routine engagement with MIM platforms is significantly associated with an increased likelihood of responding to phishing messages and rating such messages as trustworthy. In conjunction with existing research on the effects of habituation on phishing susceptibility in social networking sites such as Facebook (Vishwanath, 2015) and in email-based communication (Williams & Polage, 2019), the findings of this thesis extend this body of work by demonstrating that habitual engagement with messaging platforms similarly heightens users' susceptibility to phishing.

Fifth, this thesis makes a theoretical contribution to the study of persuasion in phishing by demonstrating how platform affordances, technical artefacts, and habitual social interaction patterns within MIM apps are strategically exploited to implement the persuasive principles of social proof, liking, and authority, thereby shaping both user susceptibility and the propagation of phishing attacks. This is a unique finding that shows how the unique characteristics of MIM apps influence the choice of persuasion principles by attackers. These findings further contribute to the theory by demonstrating that persuasive principles play a critical role not only in initial phishing victimisation but also in the propagation of phishing attacks. The use of social proof and liking principles to encourage link forwarding enables the use of persuasion as a mechanism for propagating phishing attacks through trusted social networks.

Finally, this thesis contributes to phishing research by collecting and curating a dataset of mobile instant messaging phishing attacks, which has been made publicly available to support reuse and further investigation by the research community.

7.4 Recommendations for practice

Based on the findings of this thesis, the following recommendations for addressing the problem of phishing in MIM applications are provided.

7.4.1 Recommendations for users and security awareness professionals

Current anti-phishing training materials primarily focus on email, neglecting MIM apps. To reduce susceptibility to MIM phishing, updating existing materials and improving user awareness of the threat are essential.

This thesis recommends that, in addition to existing advice that stresses the importance of users checking links before clicking, phishing awareness facts and advice should reflect the nature of phishing in MIM apps. They should inform users that fraudulent messages can be shared by their contacts who have fallen for such phishing attacks or whose accounts have been hacked. They can also be shared in MIM apps groups by the members of those groups. MIM app users should be advised to always verify the legitimacy of a message before responding by contacting both the sender of the message and the impersonated organisation. They should be aware that shared links in messenger apps can contain a manipulated link preview, which is used by cybercriminals to hide the actual organisation and destination behind another.

7.4.2 Recommendations for MIM app platforms

Findings from the studies conducted in this thesis demonstrate that cybercriminals exploit trust and habitual interaction to implement the social proof principle during MIM phishing by intentionally including phrases that convey the concept of shared risk, alerting the recipient that previous users have taken similar actions. They also show that phishers manipulate the link preview to hide the actual organisation and destination behind another. As a recommendation, this thesis recommends the implementation of automatic phishing detection solutions on MIM apps utilising existing URL and content-based solutions. However, this thesis suggests that such solutions be implemented in a privacy-preserving manner. In terms of content-based solutions, this thesis recommends incorporating relevant social proof indicators into the design of content-based phishing detection tools for MIM applications, particularly through the integration of keywords that signal commonly shared risks. Moreover, the analysis revealed that a substantial proportion of giveaway-themed phishing messages contained celebrity names, suggesting a critical area of focus for software developers. Accordingly, detection mechanisms should prioritise the identification of giveaway messages that reference celebrities, as these represent a recurrent and exploitable phishing strategy. In terms of phishing URL detection, this thesis recommends using existing phishing URL detection techniques. However, such methods should be used in a privacy-preserving manner. An example of a privacy-preserving phishing URL detection technique is the Google safe browsing real-time, privacy-preserving URL protection approach (Jasika Bawa, Xinghui Lu, Jonathan Li, 2024). This approach utilises a hash-based URL look-up, leveraging both a Google Safe Browsing server and a privacy server to ensure users' privacy and provide real-time protection. Another recommendation for MIM platform providers is to limit the ability to manipulate the content of the link preview by ensuring that the organisational cues within the titles, descriptions, images, and domain names of the link preview correspond with the information of the URL the user is visiting.

7.5 Limitation

The limitations of each study were described in their respective sections. To summarise, the quantitative study in Chapter 4 used snowball sampling, which, although effective, frequently leads to the inclusion of people with more interconnectedness compared to what would be observed in the real world. However, an effort was made to mitigate this limitation by sharing the survey link on multiple social media platforms, including r/SampleSize on Reddit, samplesize on Facebook, and SurveyCycle. Furthermore, most participants had completed either an undergraduate or postgraduate degree; therefore, the behaviours of other demographic groups might differ from those observed in the study. The study was also based on self-reported data, which is subject to bias. However, users' susceptibility was extrapolated from the analysis of their self-reported behaviour towards general links and not phishing links. Thus, this approach may have helped in reducing the bias in the study.

The study in Chapter 5 was based on phishing messages collected from publicly accessible websites using the Google search engine. The dataset does not accurately represent all MIM phishing attacks circulated within the period, since MIM phishing attacks that were not reported or shared online were not recorded. Furthermore, the use of deductive content analysis might have hindered the researchers' ability to identify additional contextual characteristics relevant to the phenomenon under investigation (Hsieh and Shannon, 2005), implying that other elements of MIM phishing might not have been identified due to reliance on existing categories from the literature. Additionally, deductive content analysis sometimes suffers from issues related to reliability and trustworthiness. However, this thesis applied the recommended strategies from the literature to enhance the trustworthiness and reliability of the findings, including selecting 10-15% of the data for the training and piloting stage and testing the inter-coder reliability (IRR) and ensuring that the IRR value meets the minimum threshold.

The study in Chapter 6 also recruited participants from specific online groups and forums, most of whom resided in the US or the UK. This has the potential to limit the generalisability of the study's findings. In addition, the use of experimental online vignette survey methodology to assess participants' susceptibility also has a limitation in that it relies on hypothetical scenarios, which may not accurately capture real-world scenarios. However, the participants in this study were not informed that the study focused on phishing; therefore, the limitation associated with measuring the natural behaviour of the participants was mitigated to some extent.

The study in Chapter 6 was based on an online survey. However, despite their advantages, researchers have pointed out that participants may not fully engage in the research when they are not being observed. This study in Chapter 6 employed different attention checks and data cleaning strategies to ensure the quality of the data used for analysis. As a result, only a reasonable number of the participants passed the data quality checks.

The studies in Chapters 4 and 6 rely on self-reported data. Thus, it is important to note that the self-reported data may not accurately represent users' actual behaviour in real-world situations. In addition, this form of data is subject to known limitations, including response bias, which is the tendency of participants to respond to questions on some basis other than the question content. For example, there is a tendency for participants to present themselves in a socially favourable manner, particularly when reporting on personal traits, attitudes, or behaviours, which may lead them to falsify or exaggerate answers. Furthermore, the participants may not be sure how to answer a particular question because they lack knowledge in the area or because they do not understand the question. Although the research has used several techniques, including attention checks and data cleaning strategies, to mitigate other types of bias, such as extreme responding, it is important to note that the results from questionnaire studies in this thesis may not represent the actual behaviour of the participants.

7.6 Future work

The study provides a solid foundation for future research on the secure behaviours of MIM app users, the psychological aspects of phishing in MIM apps and the factors that may impact user susceptibility to phishing in MIM apps.

First, the contributions section of this thesis highlights that phishing in MIM apps is a socio-technical problem rather than a predominantly technical or behavioural problem. In this regard, there is an opportunity for future research on the development of a theoretical model of factors that influence users in clicking on phishing links in MIM apps from a broader socio-technical perspective, integrating theories such as the habit theory and deterrence theory to further investigate the individual technological factors that contribute to susceptibility to phishing in MIM apps.

Second, while this thesis highlights the role of interpersonal trust in shaping phishing susceptibility, future research could further examine how trust interacts with technical cues, such as system-generated warnings or security indicators. Experimental studies could investigate whether and to what extent such technical cues are capable of moderating trust-based heuristics. For instance, future work could explore whether users would continue to express an intention to engage with phishing links originating from contacts within their social circle, even when explicit warnings indicate the fraudulent nature of the links.

Third, the study presented in Chapter 5 could be further strengthened by expanding the dataset of phishing messages and conducting a more detailed analysis to develop a comprehensive taxonomy of MIM phishing attacks.

Fourth, the findings of this thesis depict phishing in MIM apps as a socio-technical problem requiring both technical defences and user awareness. Therefore, future research could look into developing and evaluating anti-phishing facts and advice, applying the user awareness recommendations provided in 7.4.1. This can be in the form of a between-subject study where one group evaluates sample MIM

phishing messages after reading the tips and advice, and another evaluates the same sample MIM phishing messages without reading such messages. Such a study could be used to determine the effectiveness of the recommendations in helping users detect phishing in MIM apps. Furthermore, future research could investigate the potential of Large Language Models (LLMs) to act as cybersecurity assistants for users of MIM apps. The potential of these models to act as cybersecurity assistants for users has recently been documented in the literature (Chataut, Gyawali and Usman, 2024). The recent release and integration of the LLM model, Meta AI, into Facebook, Instagram, WhatsApp, and Messenger demonstrate that these models can be seamlessly integrated into the various MIM apps while preserving the end-to-end encryption nature of communication. For example, Meta ensures users' privacy by ensuring that the model reads only messages that mention @Meta AI (Meta, 2024). This gives users control over what the model can read and respond to, maintaining the pledge to the end-to-end encryption aspect of MIM apps. As a result, future research could explore the development and integration of such models into MIM platforms, with the capacity to act as a cybersecurity assistant for users, helping them detect phishing messages and links. In this regard, MIM app users can request assistance when encountering suspicious messages or links. The LLM can be trained to analyse messages thoroughly, looking for typical phishing indicators such as suspicious links or attachments. It should also be able to examine the language used in the email. The findings of this thesis offer some valuable tips for training such models, as they can be trained to identify relevant social proof cues or words that indicate the concept of shared risks. Furthermore, they could be trained to look for messages emphasising reward and gain, such as giveaway campaigns.

References

Abad, S., Gholamy, H. and Aslani, M. (2023) "Classification of malicious URLs using machine learning," Sensors, 23(18), p. 7760. doi: 10.1080/23742917.2024.2378552

Abroshan, H. *et al.* (2017) 'Phishing attacks root causes', in *International Conference on Risks and Security of Internet and Systems*. Springer, pp. 187–202. doi: 10.1007/978-3-319-76687-4_13

Abroshan, H. *et al.* (2021) 'Phishing happens beyond technology: The effects of human behaviors and demographics on each step of a phishing process', *IEEE Access*, 9, pp. 44928–44949. doi:10.1109/ACCESS.2021.3066383

Abutair, H., Belghith, A. and AlAhmadi, S. (2019) 'CBR-PDS: a case-based reasoning phishing detection system', *Journal of Ambient Intelligence and Humanized Computing*, 10, pp. 2593–2606. doi: 10.1007/s12652-018-0736-0

ActionFraud (2018) *Adidas scam*. Available at: https://www.facebook.com-/actionfraud/posts/this-latest-adidas-whatsapp-scam-is-another-example-of-a-clever-homograph-attack/2021054694578900/ (Accessed: 30 May 2023).

Afane, K. et al. (2024) "Next-generation phishing: How LLM agents empower cyber attackers," in 2024 IEEE International Conference on Big Data (BigData). IEEE, pp. 2558–2567, doi: 10.1109/BigData62323.2024.10825018.

Aguinis, H. and Bradley, K. J. (2014) 'Best practice recommendations for designing and implementing experimental vignette methodology studies', *Organizational research methods*, 17(4), pp. 351–371.doi: 10.1177/1094428114547952

AirBNB Community Center (2018) *New to AirBNB - Communication via Whatsapp*. Available at : https://community.withairbnb.com/t5/Help/New-to-AirBNB-Communication-via-Whatsapp/m-p/627669 (Accessed: 13 August 2023).

Akbar, N. (2014) '*Analysing persuasion principles in phishing emails*'. MSc Thesis University of Twente. Available at: https://essay.utwente.nl/66177/ (Assessed:25 March 2023)

Akdemir, N. and Yenal, S. (2021) 'How phishers exploit the coronavirus pandemic: A content analysis of COVID-19 themed phishing emails', *SAGE Open*, 11(3), p. 21582440211031880. doi: 10.1177/21582440211031879

Alani, M. M. and Tawfik, H. (2022) 'PhishNot: A Cloud-Based Machine-Learning Approach to Phishing URL Detection', *Computer Networks*, 218, p. 109407. doi:10.1016/j.comnet.2022.109407

Albakry, S., Vaniea, K. and Wolters, M. K. (2020) 'What is this URL's Destination? Empirical Evaluation of Users' URL Reading', *Proceedings of Human Factors in Computing Systems - Proceedings*, pp. 1–12. doi: 10.1145/3313831.3376168.

Algarni, A. A. M., Xu, Y. and Chan, T. (2015) 'Susceptibility to social engineering in social networking sites: The case of Facebook', *Proceedings of the 36th International Conference on Information Systems (ICIS 2015)*. Association for Information Systems (AIS), pp. 1–23.

Ali, S. *et al.* (2018) 'Privacy and security issues in online social networks', *Future Internet*, 10(12), p. 114. doi:10.3390/fi10120114

Alnajim, A. and Munro, M. (2009) 'An anti-phishing approach that uses training intervention for phishing websites detection', proceedings of *ITNG 2009 - 6th International Conference on Information Technology: New Generations*. doi: 10.1109/ITNG.2009.109.

Alohali, M. *et al.* (2018) 'Identifying and predicting the factors affecting end-users' risk-taking behavior', *Information & Computer Security*, 26(3), pp. 306–326. doi: 10.1108/ICS-03-2018-0037

Alqhatani, A. (2021) '*Understanding and Designing for Sharing and Privacy in Wearable Fitness Platforms*'. The University of North Carolina at Charlotte. Available at: www.proquest.com/openview/b297bcd8a349c8c67b33af3d1d8c6144/1?pq-origsite=gscholar&cbl=18750&diss=y (Accessed: 17/09/2025)

Alsharnouby, M., Alaca, F. and Chiasson, S. (2015) 'Why phishing still works: User strategies for combating phishing attacks', *International Journal of Human Computer Studies*. doi: 10.1016/j.ijhcs.2015.05.005.

Anderson, J. and Lightfoot, A. (2022) 'Exploratory Survey Research', *Research Methods in Language Teaching and Learning: A Practical Guide*, pp. 182–199. doi: 10.1002/9781394260409.ch12

APWG (2020) *Phishing Activity Trend Report 3rd Quarter*. Available at: https://docs.apwg.org/reports/apwg_trends_report_q3_2020.pdf (25/10/2020)

APWG (2025) *PHISHING ACTIVITY TRENDS REPORT*. Available at: https://apwg.org/trendsreports/ (Accessed: 22 July 2025).

Arachchilage, N. A. G. and Cole, M. (2011) 'Design a mobile game for home computer users to prevent from "phishing attacks"', *Proceedings of the International Conference on Information Society, i-Society 2011*. doi: 10.1109/isociety18435.2011.5978543.

Arsen (2024) *Facebook Phishing: Protecting Your Profile*. Available at: https://arsen.co/en/blog/facebook-phishing#:~:text=Phishers often send messages through,this you in this photo%3F%22. (Accessed: 17/09/2025)

Aung, E. S. and Yamana, H. (2019) 'URL-based phishing detection using the entropy of non- A lphanumeric characters', *Proceedings of the ACM International Conference Proceeding Series*. doi: 10.1145/3366030.3366064. (Accessed 29/02/2020

Auspurg, K. and Hinz, T. (2014) *Factorial survey experiments*. Sage Publications.

Auton, J.C. and Sturman, D. (2025) "Persuasion under pressure: the influence of persuasion principles and time constraints on phishing email susceptibility," Information & Computer Security, 33(5), pp. 845–859. Doi: 10.1108/ICS-07-2024-0163

Avrahami, D. and Hudson, S. E. (2006) 'Communication characteristics of instant messaging: effects and predictions of interpersonal relationships', *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pp. 505–514, November 4 - 8, 2006, Banff Alberta Canada. doi:10.1145/1180875.118095

Bada, M., Sasse, A. M. and Nurse, J. R. C. (2019) 'Cyber security awareness campaigns: Why do they fail to change behaviour?', *International Conference on Cyber Security for Sustainable Society,* csss2015. Available at: https://ora.ox.ac.uk/objects/uuid:cfed4907-d32a-4450-b075-ad37477b10d8 (Accessed 08/02/2026)

Bagui, Sikha *et al.* (2019) 'Classifying phishing email using machine learning and deep learning', *Proceedings of the 2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*. IEEE, pp. 1–2. doi: 10.1109/CyberSecPODS.2019.8885143

Baryshevtsev, M. and McGlynn, J. (2020) 'Persuasive appeals predict credibility judgments of phishing messages', *Cyberpsychology, Behavior, and Social Networking*, 23(5), pp. 297–302. doi:10.1089/cyber.2019.059

Basnet, R. B. and Doleck, T. (2015) 'Towards developing a tool to detect phishing URLs: A machine learning approach', *Proceedings - 2015 IEEE International Conference on Computational Intelligence and Communication Technology, CICT 2015*. doi: 10.1109/CICT.2015.63.

BBC (2023) *Thornaby: Woman targeted in £13k railway station QR code scam*. Available at: https://www.bbc.co.uk/news/uk-england-tees-67335952 (Accessed: 18 March 2025).

Bell, S. and Komisarczuk, P. (2020) 'An Analysis of Phishing Blacklists: Google Safe Browsing, OpenPhish, and PhishTank', *Proceedings of the Australasian Computer Science Week Multiconference*, pp. 1–11. doi :10.1145/3373017.3373020

Binance (2021) *How to Protect Your Binance Account from Scam*. Available at: https://www.binance.com/en/support/faq/how-to-protect-your-binance-account-from-scam-be30cd51c32a4e1f9b2ab60807ce0240. (Accessed: 13 August 2023).

Blythe, M., Petrie, H. and Clark, J. A. (2011) 'F for fake: four studies on how we fall for phish', *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 3469–3478. Available from: ACM Portal: ACM Digital Library

Borges-Tiago, M. T., Tiago, F. and Cosme, C. (2019) 'Exploring users' motivations to participate in viral communication on social media', *Journal of business research*, 101, pp. 574–582. Available from: ACM Portal: ACM Digital Library. doi 10.1145/1978942.1979459

Bradley, N. A. and Dunlop, M. D. (2005) 'Toward a multidisciplinary model of context to support context-aware computing', *Human-Computer Interaction*, 20(4), pp. 403–446. doi:10.1207/s15327051hci2004_2

Brant, R. (1990) 'Assessing proportionality in the proportional odds model for ordinal logistic regression', *Biometrics*, pp. 1171–1178. Available at :https://www.jstor.org/stable/2532457

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative research in psychology*, 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa

Bryman, A. (2016) *Social research methods*. Oxford university press.

Bujang, Y. R. and Hussin, H. (2013) 'Should we be concerned with spam emails? A look at its impacts and implications', *Proceedings of the 2013 5th International Conference on Information and Communication Technology for the Muslim World (ICT4M)*. IEEE, pp. 1–6. doi: 10.1109/ICT4M.2013.6518877

Bulkgate (2021) *Viber for Business*. Available at: https://www.bulkgate.com/en/solutions/viber-for-business/ (Accessed: 8 June 2021).

Butavicius, M. *et al.* (2016) 'Breaching the human firewall: Social engineering in phishing and spear-phishing emails', *Proceedings of ACIS 2015*. Association for Information Systems. Available at: https://aisel.aisnet.org/acis2015/98/. (Accessed 18/09/2025)

Buxton, O. (2024) *What is smishing? 6 reliable SMS phishing detection tips*. Available at: https://lifelock.norton.com/learn/internet-security/what-is-smishing (Accessed: 4 May 2025).

Cain, A. A., Edwards, M. E. and Still, J. D. (2018) 'An exploratory study of cyber hygiene behaviors and knowledge', *Journal of Information Security and Applications*, 42, pp. 36–45. doi: 10.1016/j.jisa.2018.08.002.

Canfield, C. I., Fischhoff, B. and Davis, A. (2019) 'Better beware: comparing metacognition for phishing and legitimate emails', *Metacognition and learning*, 14, pp. 343–362. doi: 10.1007/s11409-019-09197-5

Cao, H. *et al.* (2021) 'You Recommend, I Buy: How and Why People Engage in Instant Messaging Based Social Commerce', *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), pp. 1–25. doi: 10.1145/3449141.

Caputo, D. D. *et al.* (2014) 'Going spear phishing: Exploring embedded training and awareness', *IEEE Security and Privacy, Page(s):28-38*.doi: 10.1109/MSP.2013.106.

Chaiken, S. (1980) 'Heuristic versus systematic information processing and the use of source versus message cues in persuasion.', *Journal of personality and social psychology*, 39(5), p. 752.

Chataut, R., Gyawali, P. K. and Usman, Y. (2024) 'Can ai keep you safe? a study of large language models for phishing detection', *Proceedings of the 2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, pp. 548–554. doi: 10.1109/CCWC60891.2024.10427626

Checkpoint (2021) *Autoreply attack! New Android malware found in Google Play Store spreads via malicious auto-replies to WhatsApp messages*. Available at: https://blog.checkpoint.com/2021/04/07/autoreply-attack-new-android-malware-found-in-google-play-store-spreads-via-malicious-auto-replies-to-whatsapp-messages/ (Accessed: 14 June 2021).

Chhabra, S. *et al.* (2011) 'Phi. sh/$ ocial: the phishing landscape through short urls', *Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*, pp. 92–101. doi: 10.1145/2030376.20303

Churchill Jr, G. A. (1979) 'A paradigm for developing better measures of marketing constructs', *Journal of marketing research*, 16(1), pp. 64–73. doi: 10.1177/002224377901600110

Cialdini, R. B. (2006) 'Influence: the psychology of persuasion, revised edition', *New York: William Morrow*.

CIO (2006a) *McAfee Warns of 'SMiShing' Attacks*. Available at: https://www.cio.com/article/2444862/mcafee-warns-of--smishing--attacks.html (Accessed: 4 June 2021).

CIO (2006b) *Move Over Phishing, Here Comes 'Vishing'*. Available at: https://www.cio.com/article/258747/voice-over-ip-move-over-phishing-here-comes-vishing.html.

CISCO (2021) *What is phishing?* Available at: https://www.cisco.com/c/en_uk/products/security/email-security/what-is-phishing.html#~how-phishing-works (Accessed: 25 January 2021).

Clasen, M., Li, F. and Williams, D. (2021) 'Friend or foe: An investigation into recipient identification of sms-based phishing', *Proceedings of the International Symposium on Human Aspects of Information Security and Assurance*. Springer, pp. 148–163. doi: 10.1007/978-3-030-81111-2_13

Clifford, S. and Jerit, J. (2014) 'Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies', *Journal of Experimental Political Science*, 1(2), pp. 120–131. doi:10.1017/xps.2014.5

Cloudflare (2025) *What is email? | Email definition*. Available at: https://www.cloudflare.com/en-gb/learning/email-security/what-is-email/ (Accessed: 13 April 2025).

Cofense (2021) *History of Phishing*. Available at: https://cofense.com/knowledge-center/history-of-phishing/ (Accessed: 26 January 2021).

Cofense (2023) *What is Smishing?* Available at: https://cofense.com/knowledge-center/what-is-smishing#:~:text=Though it%27s been around since,use and increasing technological capabilities.

Creswell, J. W. and Clark, V. P. (2007) 'Mixed methods research', *Thousand Oaks, CA*.

ctinsider (2025) *Watertown police warn residents of new Facebook scam using fake accounts*. Available at: https://www.ctinsider.com/waterbury/article/watertown-pd-warn-new-facebook-scam-20169487.php? (Accessed: 29 April 2025).

Cuddeford, D. (2018) *WhatsApp: Mobile Phishing's Newest Attack Target*. Available at: https://www.darkreading.com/endpoint/whatsapp-mobile-phishing-s-newest-attack-target (Accessed: 11 March 2022).

Dassanayake, D. (2021) *WhatsApp users warned not to trust fake Amazon anniversary free gift message*. Available at: https://www.express.co.uk/life-style/science-technology/1415675/WhatsApp-message-warning-Amazon-free-gift-scam (Accessed: 15 September 2022).

della Porta, D. and Keating, M. (2008) *Approaches and methodologies in the social sciences: A pluralist perspective*. Cambridge University Press.

DeSimone, J. A., Harms, P. D. and DeSimone, A. J. (2015) 'Best practice recommendations for data screening', *Journal of Organizational Behavior*, 36(2), pp. 171–181. doi: 10.1002/job.1962

Desolda, G., Greco, F. and Viganò, L. (2025) "APOLLO: A GPT-based tool to detect phishing emails and generate explanations that warn users," Proceedings of the ACM on Human-Computer Interaction, 9(4), pp. 1–33. Doi: 10.1145/373304

Devriendt, E. *et al.* (2012) 'Content validity and internal consistency of the Dutch translation of the Safety Attitudes Questionnaire: an observational study', *International journal of nursing studies*, 49(3), pp. 327–337. doi: 10.1016/j.ijnurstu.2011.10.002.

Dhamija, R., Tygar, J. D. and Hearst, M. (2006) 'Why phishing works', *Proceedings of the Conference on Human Factors in Computing Systems, pp. 581-590*. doi: 10.1145/1124772.1124861.

Diaz, A., Sherman, A. T. and Joshi, A. (2020) 'Phishing in an academic community: A study of user susceptibility and behavior', *Cryptologia*, 44(1), pp. 53–67. doi:10.1080/01611194.2019.1623343

Distler, V. *et al.* (2021) 'A Systematic Literature Review of Empirical Methods and Risk Representation in Usable Privacy and Security Research', *ACM Transactions on Computer-Human Interaction (TOCHI)*, 28(6), pp. 1–50. doi:10.1145/3469845

Distler, V. (2023) 'The influence of context on response to spear-phishing attacks: an in-situ deception study', *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–18. doi:doi.org/10.1145/3544548.3581170

Dixon, N. (2018) 'Stranger-ness and belonging in a neighbourhood WhatsApp group', *Open Cultural Studies*, 1(1), pp. 493–503. doi: 10.1515/culture-2017-0046

Dixon, S. J. (2024) *Most popular global mobile messenger apps as of April 2024, based on number of monthly active users*. Available at: https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/ (Accessed: 11 July 2024).

Dror, Y. (2024) *Your 2024 guide to WhatsApp scams: How to avoid and handle them*. Available at: https://ironvest.com/blog/whatsapp-scams/ (Accessed: 25 July 2025).

Dunlop, M., Groat, S. and Shelly, D. (2010) 'GoldPhish: Using images for content-based phishing analysis', *proceedings of the 5th International Conference on Internet Monitoring and Protection, ICIMP 2010*. doi: 10.1109/ICIMP.2010.24.

Duò, M. (2022) *Alexa Rank: Everything You Need to Know About It*. Available at: https://kinsta.com/blog/alexa-rank/ (Accessed: 22 March 2023).

Edwards, M. *et al.* (2023) 'SMiShing attack vector: Surveying end-user behavior, experience, and knowledge', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications Sage CA: Los Angeles, CA, pp. 1911–1915.

Egress (2024a) *Key takeaways from the 2024 Phishing Threat Trends Report*. Available at:https://www.egress.com/blog/company-news/takeaways-from-the-phishing-threat-trends-report (Accessed: 22 July 2025).

Egress (2024b) *The evolution of QR code phishing: Unmasking new 'quishing' tactics*. Available at: https://www.egress.com/blog/phishing/the-evolution-of-qr-code-phishing (Accessed: 18 March 2025).

Elliot Volkman (2019) *Why Social Media is Increasingly Abused for Phishing Attacks*, *PhishLabs Blog*. Available at: https://info.phishlabs.com/blog/how-social-media-is-abused-for-phishing-attacks (Accessed: 24 June 2020).

Elo, S. *et al.* (2014) 'Qualitative content analysis: A focus on trustworthiness', *SAGE open*, 4(1). doi:10.1177/2158244014522633

Etikan, I., Musa, S. A. and Alkassim, R. S. (2016) 'Comparison of convenience sampling and purposive sampling', *American journal of theoretical and applied statistics*, 5(1), pp. 1–4.doi:10.11648/j.ajtas.20160501.11

FBI (2007) *Something Vishy Be Aware of a New Online Scam*. Available at: https://archives.fbi.gov/archives/news/stories/2007/february/vishing_022307.

Feng, F. *et al.* (2018) 'The application of a novel neural network in the detection of phishing websites', *Journal of Ambient Intelligence and Humanized Computing*, pp.1865–1879 . doi:10.1007/s12652-018-0786-3

Fernando, M. and Arachchilage, N. A. G. (2020) 'Why Johnny can't rely on anti-phishing educational interventions to protect himself against contemporary phishing attacks?'*Proceedings of the Australasian Conference on Information Systems* , Australia. Available at: https://aisel.aisnet.org/acis2019/42 .

Ferreira, A., Coventry, L. and Lenzini, G. (2015) 'Principles of persuasion in social engineering and their use in phishing', *Proceedings of the Human Aspects of Information Security, Privacy, and Trust: Third International Conference, HAS 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015.* . Springer, pp. 36–47.doi:10.1007/978-3-319-20376-8_4

Ferreira, A. and Jakobsson, M. (2016) 'Persuasion in scams', in *Understanding social engineering based scams*. Springer, pp. 29–47.doi:10.1007/978-1-4939-6457-4_4

Ferreira, A. and Lenzini, G. (2015) 'An analysis of social engineering principles in effective phishing', in *2015 Workshop on Socio-Technical Aspects in Security and Trust*. IEEE, pp. 9–16. doi: 10.1109/STAST.2015.10

Ferreira, A. and Teles, S. (2019) 'Persuasion: How phishing emails can influence users and bypass security measures', *International Journal of Human-Computer Studies*, 125, pp. 19–31. doi:10.1016/j.ijhcs.2018.12.004

Fimberg, K. and Sousa, S. (2020) 'The Impact of Website Design on Users' Trust Perceptions', *Proceedings of the International Conference on Applied Human Factors and Ergonomics*. Springer, pp. 267–274. doi: 10.1145/3419249.3420086

fortinet (2024) *What Is Email Spoofing?* Available at: https://www.fortinet.com/resources/cyberglossary/email-spoofing (Accessed: 13 April 2025).

Franz, A. and Croitor, E. (2021) 'Who Bites the Hook? Investigating Employees' Susceptibility to Phishing: A Randomized Field Experiment.', in *ECIS* 2021 Research Papers. 125.

Frauenstein, E. D. and Flowerday, S. (2020) 'Susceptibility to phishing on social network sites: A personality information processing model', *Computers & security*, 94, p. 101862. doi:10.1016/j.cose.2020.101862

Frauenstein, E. D. and Flowerday, S. V (2016) 'Social network phishing: Becoming habituated to clicks and ignorant to threats?', *2016 Information Security for South Africa (ISSA)*. IEEE, pp. 98–105. doi: 10.1109/ISSA.2016.7802935.

FTC (2019) *How to Recognize and Avoid Phishing Scams*. Available at: https://www.consumer.ftc.gov/articles/how-recognize-and-avoid-phishing-scams (Accessed: 16 March 2020).

Furnell, S. M., Bryant, P. and Phippen, A. D. (2007) 'Assessing the security perceptions of personal Internet users', *Computers & Security*, 26(5), pp. 410–417. doi:10.1016/j.cose.2007.03.001

Gan, C. L., Lee, Y. Y. and Liew, T. W. (2024) 'Fishing for phishy messages: predicting phishing susceptibility through the lens of cyber-routine activities theory and heuristic-systematic model', *Humanities and Social Sciences Communications*, 11(1152), pp. 1–17. doi:10.1057/s41599-024-04083-1

Garfinkel, S. and Lipford, H. R. (2014) 'Usable security: History, themes, and challenges', *Synthesis Lectures on Information Security, Privacy, and Trust*, 5(2), pp. 1–124.

Garimella, K. and Tyson, G. (2018) 'WhatsApp, doc? A first look at whatsapp public group data', Proceedings of the *International AAAI Conference on Web and Social Media, ICWSM 2018*. doi: 10.1609/icwsm.v12i1.14989

Ge, N. A. G. and Love, S. (2014) 'Security awareness of computer users: A phishing threat avoidance perspective', *Computers in Human Behavior*, 38, pp. 304–312. doi: 10.1016/j.chb.2014.05.046.

Gelo, O., Braakmann, D. and Benetka, G. (2008) 'Quantitative and qualitative research: Beyond the debate', *Integrative psychological and behavioral science*, 42, pp. 266–290. doi: 10.1007/s12124-008-9078-3

Gillin, P. (2023) *The history of phishing*. Available at: https://www.verizon.com/business/resources/articles/s/the-history-of-phishing/ (Accessed: 8 May 2023).

Gillis, A. and Jacksin, W. (2002) 'Research for nurses: Methods and interpretation', *(No Title)*. Available at: https://cir.nii.ac.jp/crid/1970023484893996304

Gjertsen, E. G. B. *et al.* (2017) 'Gamification of Information Security Awareness and Training.', *Proceedings of the 3rd International Conference on Information Systems Security and Privacy*, pp. 59–70. doi: 10.5220/0006128500590070

Goel, D. and Jain, A. K. (2018) 'Mobile phishing attacks and defence mechanisms: State of art and open research challenges', *computers & security*, 73, pp. 519–544. doi: 10.1016/j.cose.2017.12.006.

Goenka, R., Chawla, M. and Tiwari, N. (2024) 'A comprehensive survey of phishing: Mediums, intended targets, attack and defence techniques and a novel taxonomy', *International Journal of Information Security*, 23(2), pp. 819–848. doi: 10.1007/s10207-023-00768-x

Gold, J. (2023) *Email phishing still the main way in for hackers: report*. Available at: https://www.csoonline.com/article/649551/email-phishing-still-the-main-way-in-for-hackers-report.html (Accessed: 19 March 2025).

Gordon, W. J. *et al.* (2019) 'Assessment of employee susceptibility to phishing attacks at US health care institutions', *JAMA network open*, 2(3), pp. e190393. doi:10.1001/jamanetworkopen.2019.0393

Gragg, D. (2003) 'A multi-level defense against social engineering', *SANS Reading Room*, 13.

Graham, R. and Triplett, R. (2017) 'Capable guardians in the digital environment: The role of digital literacy in reducing phishing victimization', *Deviant Behavior*, 38(12), pp. 1371–1382.

Green, J. and Thorogood, N. (2018) 'Qualitative methods for health research'.Sage Publications

Greitzer, F. L. *et al.* (2021) 'Experimental investigation of technical and human factors related to phishing susceptibility', *ACM Transactions on Social Computing*, 4(2), pp. 1–48. doi:10.1145/3461672

Grilli, M. D. *et al.* (2021) 'Is this phishing? Older age is associated with greater difficulty discriminating between safe and malicious emails', *The Journals of Gerontology: Series B*, 76(9), pp. 1711–1715. doi: 10.1093/geronb/gbaa228

Guba, E.G. (1990) "The paradigm dialog.," in Alternative paradigms conference, mar, 1989, indiana u, school of education, san francisco, ca, us. Sage Publications, Inc.

Guba, E.G. and Lincoln, Y.S. (1994) "Competing paradigms in qualitative research," Handbook of qualitative research, 2(163–194), p. 105.

Gupta, B. B. *et al.* (2017) 'Fighting against phishing attacks: state of the art and future challenges', *Neural Computing and Applications*. 28, 3629–3654 (2017). doi:10.1007/s00521-016-2275-y

Gupta, S. *et al.* (2015) 'Abusing phone numbers and cross-application features for crafting targeted attacks', *arXiv preprint.* 10.48550/arXiv.1512.07330

Hammarberg, K., Kirkman, M. and De Lacey, S. (2016) 'Qualitative research methods: when to use them and how to judge them', *Human reproduction*, 31(3), pp. 498–501. 10.1093/humrep/dev334

Harborth, D. and Pape, S. (2021) 'Investigating privacy concerns related to mobile augmented reality apps–A vignette based online experiment', *Computers in Human Behavior*, 122, p. 106833. doi: 10.1016/j.chb.2021.106833

Harrison, B., Svetieva, E. and Vishwanath, A. (2016) 'Individual processing of phishing emails: How attention and elaboration protect against phishing', *Online Information Review*, 40(2), pp. 265–281. doi:10.1108/OIR-04-2015-0106

Hatfield, J. M. (2018) 'Social engineering in cybersecurity: The evolution of a concept', *Computers & Security*, 73, pp. 102–113. doi:10.1016/j.cose.2017.10.008

Henderson, K.A. (2011) "Post-positivism and the pragmatics of leisure research," *Leisure sciences*, 33(4), pp. 341–346.

Holbrook, A. L., Green, M. C. and Krosnick, J. A. (2003) 'Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias', *Public opinion quarterly*, 67(1), pp. 79–125. doi: 10.1086/346010

Hong, J. (2012) 'The state of phishing attacks', *Communications of the ACM*. Volume 55(1), pp 74 - 81. doi: 10.1145/2063176.2063197.

House, D. and Raja, M. K. (2020) 'Phishing: message appraisal and the exploration of fear and self-confidence', *Behaviour & Information Technology*, 39(11), pp. 1204–1224.doi:10.1080/0144929X.2019.1657180

Hovland, C.I., Janis, I.L. and Kelley, H. . (1953) *Communication and Persuasion: Psychological Studies of Opinion Change*. New Haven, CT: Yale University Press.

Hsieh, H.-F. and Shannon, S. E. (2005) 'Three approaches to qualitative content analysis', *Qualitative health research*, 15(9), pp. 1277–1288. doi: 10.1177/1049732305276687

Hughes, R. and Huby, M. (2002) 'The application of vignettes in social and nursing research', *Journal of advanced nursing*, 37(4), pp. 382–386. doi: 10.1046/j.1365-2648.2002.02100.x

Indiana University (2022) *Phishing Education & Training*. Available at: https://phishing.iu.edu/tips-and-strategies/index.html. (Accessed 25/10/2022)

iRadio (2018) *Tyto Park Scam*. Available at: https://m.facebook.com/thisisiradio/posts/1927459280648472/?comment_id=1927565900637810 (Accessed: 30 May 2023).

Jagatic, T. N. *et al.* (2007) 'Social phishing', *Communications of the ACM*, 50(10), pp. 94–100.

Jakobsson, M. *et al.* (2007) 'What instills trust? a qualitative study of phishing': *Proceedings of the International Conference on Financial Cryptography and Data Security*. Springer, pp. 356–361. doi:10.1007/978-3-540-77366-5_32

Jamieson, S. (2004) 'Likert scales: How to (ab) use them?', *Medical Education*, 38(12), pp. 1217–1218.

Jasika Bawa, Xinghui Lu, Jonathan Li, A. W. (2024) *Real-time, privacy-preserving URL protection*. Available at: https://security.googleblog.com/2024/03/blog-post.html (Accessed: 8 November 2024).

Jayatilaka, A., Arachchilage, N. A. G. and Babar, M. A. (2021) 'Falling for Phishing: An Empirical Investigation into People's Email Response Behaviors', *Proceedings of 42 International Conference on Information Systems*, Austin, Texas, USA. Association for Information Systems. Available at: https://aisel.aisnet.org/icis2021/cyber_security/cyber_security/1/

Jones, K. S. *et al.* (2020) 'How social engineers use persuasion principles during vishing attacks', *Information & Computer Security*. doi:

https://doi.org/10.1108/ICS-07-2020-0113 doi: 10.1108/ICS-07-2020-01130113.

Jr., T. H. (2019) *How this scammer used phishing emails to steal over $100 million from Google and Facebook*. Available at: https://www.cnbc.com/2019/03/27/phishing-email-scam-stole-100-million-from-facebook-and-google.html (Accessed: 17 March 2025).

Kankam, P.K. (2019) "The use of paradigms in information research," *Library & Information Science Research*, 41(2), pp. 85–92. doi: 10.1016/j.lisr.2019.04.003

Kaspersky (2021) *Phishing in messenger apps – what's new?* Available at: https://www.kaspersky.com/about/press-releases/2021_phishing-in-messenger-apps-whats-new (Accessed: 4 January 2022).

Kävrestad, J. *et al.* (2022) 'Evaluation of contextual and game-based training for phishing detection', *Future Internet*, 14(4), p. 104. doi:10.3390/fi14040104

keepnetlabs (2024a) *How to Run Phishing Simulations: A Step-by-Step Guide*. Available at: https://keepnetlabs.com/blog/how-to-run-phishing-simulations-a-step-by-step-guide (Accessed: 11 May 2025).

keepnetlabs (2024b) *What Are Vishing Statistics in 2025? What Are Real-life Vishing Examples?* Available at: https://keepnetlabs.com/blog/rise-of-voice-phishing-a-deep-dive-into-vishing-statistics-in-2023 (Accessed: 29 April 2025).

Keith, M. J. *et al.* (2015) 'The role of mobile-computing self-efficacy in consumer information disclosure', *Information Systems Journal*, 25(6), pp. 637–667. doi: 10.1111/isj.12082

Kelle, U. (2006) 'Combining qualitative and quantitative methods in research practice: purposes and advantages', *Qualitative research in psychology*, 3(4), pp. 293–311. doi: 10.1177/1478088706070839

Khonji, M., Iraqi, Y. and Jones, A. (2013) 'Phishing detection: A literature survey', *IEEE Communications Surveys and Tutorials*, pp. 2091–2121. doi: 10.1109/SURV.2013.032213.00009.

Kietzmann, J. H. *et al.* (2011) 'Social media? Get serious! Understanding the functional building blocks of social media', *Business Horizons*. doi: 10.1016/j.bushor.2011.01.005.

Killam, L. (2013) Research terminology simplified: Paradigms, axiology, ontology, epistemology and methodology. Laura Killam.

Kim, D. and Kim, J. H. (2013) 'Understanding persuasive elements in phishing e-mails: A categorical content and semantic network analysis', *Online Information Review*. doi: 835-850.10.1108/OIR-03-2012-0037.

Kjeldskov, J. *et al.* (2004) 'Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field', in *International Conference on Mobile Human-Computer Interaction*. Springer, pp. 61–73. doi: /10.1007/978-3-540-28637-0_6

Klakegg, O.J. and Tvedt, I.M. (2024) "Ontology, epistemology, and axiology: Understand your philosophy and approach," in Design methods and practices for research of project management. Routledge, pp. 3–18.

Klütsch, J. *et al.* (2024) 'Friend or phisher: how known senders and fear of missing out affect young adults' phishing susceptibility on social media', *Humanities and Social Sciences Communications*, 11(1), pp. 1–14. doi: 10.1057/s41599-024-03412-8

Kondracki, N. L., Wellman, N. S. and Amundson, D. R. (2002) 'Content analysis: Review of methods and their applications in nutrition education', *Journal of nutrition education and behavior*, 34(4), pp. 224–230. doi: 10.1016/S1499-4046(06)60097-3

Kovacs, G. T. *et al.* (2012) 'The Australian community overwhelmingly approves IVF to treat subfertility, with increasing support over three decades', *Australian and New Zealand Journal of Obstetrics and Gynaecology*, 52(3), pp. 302–304. doi:10.1111/j.1479-828X.2012.01444.x

Krippendorff, K. (2018) *Content analysis: An introduction to its methodology*. Sage publications.

Krisam, C. *et al.* (2021) 'Dark patterns in the wild: Review of cookie disclaimer designs on top 500 german websites', *Proceedings of the 2021 European Symposium on Usable Security*, pp. 1–8. doi:10.1145/3481357.3481516

Kritika, E. (2025) "A comprehensive literature review on phishing URL detection using deep learning techniques," Journal of Cyber Security Technology, 9(4), pp. 315–343. doi: 10.1080/23742917.2024.2378552

Krombholz, K. *et al.* (2015) 'Advanced social engineering attacks', *Journal of Information Security and Applications*, pp.113-122. doi: 10.1016/j.jisa.2014.09.005.

Kumaraguru, P. (2009) *PhishGuru: A System for Educating Users about Semantic Attacks*. Carnegie Mellon University. Available at: http://reports-archive.adm.cs.cmu.edu/anon/isr2009/CMU-ISR-09-106.pdf.

Kumaraguru, P. *et al.* (2010) 'Teaching johnny not to fall for phish', *ACM Transactions on Internet Technology*., 10(2), pp.1-31. doi: 10.1145/1754393.1754396.

Kuraku, S. (2022) *Curiosity Clicks: The Need for Security Awareness*. University of the Cumberlands. Available at: https://www.proquest.com/openview/715fac6ca9ff0be644e582ebefcb10ee/1?pq-origsite=gscholar&cbl=18750&diss=y.

Kustiawan, Y.A. and Ghauth, K.I. (2025) “Feature Engineering for Phishing Website Detection Using Machine Learning: A Systematic Review,” IEEE Access, 13, pp. 192080–192104. doi: 10.1109/ACCESS.2025.3630334

Kyngäs, H. (2020) ‘Qualitative research and content analysis’, *The application of content analysis in nursing science research*, pp. 3–11. doi: https://doi.org/10.1007/978-3-030-30199-6_1

Lain, D., Kostiainen, K. and Čapkun, S. (2022) ‘Phishing in organizations: Findings from a large-scale and long-term study’, in *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, pp. 842–859. doi: 10.1109/SP46214.2022.9833766

Lallemand, C. and Koenig, V. (2020) ‘Measuring the contextual dimension of user experience: development of the user experience context scale (UXCS)’, *Proceedings of the 11th nordic conference on human-computer interaction: shaping experiences, shaping society*, pp. 1–13. doi: h10.1145/3419249.3420156

Lambert, V. A. and Lambert, C. E. (2012) ‘Qualitative descriptive research: An acceptable design’, *Pacific Rim international journal of nursing research*, 16(4), pp. 255–256.

LaRose, R. and Eastin, M. S. (2004) ‘A social cognitive theory of Internet uses and gratifications: Toward a new model of media attendance’, *Journal of broadcasting & electronic media*, 48(3), pp. 358–377.doi: 10.1207/s15506878jobem4803_2

Laura Ceci (2024) *Mobile messenger and communication apps - Statistics & Facts*. Available at: https://www.statista.com/topics/1523/mobile-messenger-apps/#topicOverview (Accessed: 25 August 2024).

Lecun, Y., Bengio, Y. and Hinton, G. (2015) ‘Deep learning’, *Nature*. 521, 436–444. doi: 10.1038/nature14539.

Lee, Y. Y., Gan, C. L. and Liew, T. W. (2023a) ‘Susceptibility to instant messaging phishing attacks: does systematic information processing differ between

genders?', Crime Prevention and Community Safety, 25(2), pp. 179–203. doi: /10.1057/s41300-023-00176-2

Lee, Y. Y., Gan, C. L. and Liew, T. W. (2023b) 'Thwarting instant messaging phishing attacks: the role of self-efficacy and the mediating effect of attitude towards online sharing of personal information', *International Journal of Environmental Research and Public Health*, 20(4), p. 3514. doi:10.3390/ijerph20043514

Lei, W., Hu, S. and Hsu, C. (2023) 'Unveiling the process of phishing precautions taking: The moderating role of optimism bias', *Computers & Security*, 129, p. 103249. doi:10.1016/j.cose.2023.103249

Li, L. *et al.* (2014) 'Does explicit information security policy affect employees' cyber security behavior? A pilot study', in *2014 Enterprise Systems Conference*. IEEE, pp. 169–173. doi: 10.1109/ES.2014.66.

Liang, H. and Xue, Y. L. (2010) 'Understanding security behaviors in personal computer usage: A threat avoidance perspective', *Journal of the association for information systems*, 11(7), p. 1. doi: 10.17705/1jais.00232.

Limayem, M., Hirt, S. G. and Cheung, C. M. K. (2007) 'How habit limits the predictive power of intention: The case of information systems continuance', *MIS quarterly*, pp. 705–737. doi:10.2307/25148817

Lin, E. *et al.* (2011) 'Does domain highlighting help people identify phishing sites?', *Proceedings of the Conference on Human Factors in Computing Systems*. doi: 10.1145/1978942.1979244.

Lin, T. *et al.* (2019) 'Susceptibility to spear-phishing emails: Effects of internet user demographics and email content', *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(5), pp. 1–28. doi: https://doi.org/10.1145/3336141

Loraine, B., Wick, W. and Christoph, G. (2020) 'How to use and assess qualitative research methods', *Neurological Research and Practice*, 2(1). doi: 10.1186/s42466-020-00059-z

Loxdal, J. *et al.* (2021) 'Why Phishing Works on Smartphones: A Preliminary Study, *Proceedings of 2021 Hawaii International Conference on System Sciences (HICSS)* , pp. 1–10. Available at: https://aisel.aisnet.org/hicss-54/st/security_and_privacy_of_hci/5/

Maksimovic, J. and Evtimov, J. (2023) "Positivism and post-positivism as the basis of quantitative research in pedagogy," Research in Pedagogy, 13(1), pp. 208–218.

Maroofi, S. *et al.* (2021) 'Adoption of email anti-spoofing schemes: a large scale analysis', *IEEE Transactions on Network and Service Management*, 18(3), pp. 3184–3196. doi: 10.1109/TNSM.2021.3065422

Martinho, C. and Zejnilovic, S. (no date) *Goodbye, Alexa. Hello, Cloudflare Radar Domain Rankings*, *2022*. Available at: https://blog.cloudflare.com/radar-domain-rankings/ (Accessed: 25 April 2023).

Mathis, F., Vaniea, K. and Khamis, M. (2021) 'Replicueauth: Validating the use of a lab-based virtual reality setup for evaluating authentication systems', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–18. doi: 10.1145/3411764.344547

McAfeeSiteAdvisor (2023) *McAfee WebAdvisor*. Available at: https://www.mcafee.com/en-gb/safe-browser/mcafee-webadvisor.html (Accessed: 9 May 2023).

McHugh, M. L. (2012) 'Interrater reliability: the kappa statistic', *Biochemia medica*, 22(3), pp. 276–282.

McKibben, W. B. *et al.* (2020) 'How to conduct a deductive content analysis in counseling research', *Counseling Outcome Research and Evaluation*, pp. 1–13. doi: 10.1080/21501378.2020.1846992

McLachlan, S. (2021) *Experiment: What Kind of Promoted Tweet Gets a Higher Click-Through Rate?* Available at: https://blog.hootsuite.com/experiment-promoted-tweets/ (Accessed: 13 June 2021).

Medvet, E., Kirda, E. and Kruegel, C. (2008) 'Visual-similarity-based phishing detection', in *Proceedings of the 4th International Conference on Security and Privacy in Communication Networks, SecureComm'08*. doi: 10.1145/1460877.1460905.

Meléndez, R., Ptaszynski, M. and Masui, F. (2024) 'Comparative Investigation of Traditional Machine-Learning Models and Transformer Models for Phishing Email Detection', *Electronics*, 13(24), p. 4877. doi: 10.3390/electronics13244877

Merriam, S. B. (2002) 'Introduction to qualitative research', *Qualitative research in practice: Examples for discussion and analysis*, 1(1), pp. 1–17.

Meta (2024) *About AIs from Meta*. Available at: https://faq.whatsapp.com/2257017191175152 (Accessed: 6 November 2024).

Methods (2024) *Be cautious of QR codes: Quishing is the new Phishing*. Available at: https://blog.methods.co.uk/en/all-insights/quishing-is-the-new-phishing (Accessed: 20 April 2025).

Metzger, M. J., Flanagin, A. J. and Medders, R. B. (2010) 'Social and heuristic approaches to credibility evaluation online', *Journal of communication*, 60(3), pp. 413–439. doi: 10.1111/j.1460-2466.2010.01488.x

Microsoft (2023a) *A brief history of QR codes*. Available at: https://www.microsoft.com/en-us/microsoft-365-life-hacks/privacy-and-safety/brief-history-qr-codes. (Accessed 02 May 2023)

Microsoft (2023b) *Different types of phishing attacks*. Available at: https://www.microsoft.com/en-gb/security/business/security-101/what-is-phishing?ef_id=_k_EAIaIQobChMIgp2WnMeI_wIVh-vtCh0kIQE4EAAYASAAEgJ70vD_BwE_k_&OCID=AIDcmmao55x8o7_SEM__k_EAIaIQobChMIgp2WnMeI_wIVh-vtCh0kIQE4EAAYASAAEgJ70vD_BwE_k_&gclid=EAIaIQobChMIgp2 (Accessed: 22 May 2023).

Microsoft (2023c) *How can SmartScreen help protect me in Microsoft Edge?* Available at: https://support.microsoft.com/en-us/microsoft-edge/how-can-smartscreen-help-protect-me-in-microsoft-edge-1c9a874a-6826-be5e-45b1-67fa445a74c8 (Accessed: 11 May 2020).

Microsoft (2025) *Use Link Preview in Outlook.com and Outlook on the web*. Available at: https://support.microsoft.com/en-gb/office/use-link-preview-in-outlook-com-and-outlook-on-the-web-ebbfd8ce-d38e-40ef-bb8c-a5362e881163 (Accessed: 13 April 2025).

Microsoft Threat Intelligence (2023) *Detecting and mitigating a multi-stage AiTM phishing and BEC campaign*. Available at: https://www.microsoft.com/en-us/security/blog/2023/06/08/detecting-and-mitigating-a-multi-stage-aitm-phishing-and-bec-campaign/ (Accessed: 14 July 2024).

Mimecast (2025) *What is phishing simulation?* Available at: https://www.mimecast.com/content/phishing-simulation/ (Accessed: 11 May 2025).

Mirror (2023) *Netflix Discount Codes August 2023*. Available at: https://discountcode.mirror.co.uk/discount-code/netflix (Accessed: 12 August 2023).

Mishra, S. and Soni, D. (2020) 'Smishing Detector: A security model to detect smishing through SMS content analysis and URL behavior analysis', *Future*

Generation Computer Systems, 108, pp. 803–815. doi:10.1016/j.future.2020.03.021

Mohajan, H. K. (2018) 'Qualitative research methodology in social sciences and related subjects', *Journal of economic development, environment and people*, 7(1), pp. 23–48. Available at: https://www.ceeol.com/search/article-detail?id=640546

Moody, G. D., Galletta, D. F. and Dunn, B. K. (2017) 'Which phish get caught? An exploratory study of individuals' susceptibility to phishing', *European Journal of Information Systems*, 26(6), pp. 564–584. doi: 10.1057/s41303-017-0058-x

Murashka, U. (2018) *WhatsApp: Newest Attack Target for Mobile Phishing*. Available at: https://www.infosecurity-magazine.com/next-gen-infosec/whatsapp-attack-mobile-phishing/ (Accessed: 6 July 2020).

Mustafa, H. *et al.* (2016) 'End-to-end detection of caller ID spoofing attacks', *IEEE Transactions on Dependable and Secure Computing*, 15(3), pp. 423–436.doi: 10.1109/TDSC.2016.2580509

Naiakshina, A. *et al.* (2017) 'Why do developers get password storage wrong? A qualitative usability study', *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 311–328. doi: 10.1145/3133956.3134082

NCSC (2018) *Phishing attacks: defending your organisation*. Available at: https://www.ncsc.gov.uk/guidance/phishing (Accessed: 25 January 2021).

NCSC (2024) *QR Codes - what's the real risk?* Available at: https://www.ncsc.gov.uk/blog-post/qr-codes-whats-real-risk (Accessed: 18 March 2025).

Netcraft (2020) *Phishing protection, in your favourite browse*. Available at: https://netcraft.app/browser/ (Accessed: 15 March 2020).

Nicholson, J., Coventry, L. M. and Briggs, P. (2017) 'Can we fight social engineering attacks by social means? Assessing social salience as a means to improve phish detection.', in *SOUPS*, pp. 285–298.

Nieto, J. G. (2018) *What is the character limit that a WhatsApp message can have?* Available at:

Norman, G. (2010) 'Likert scales, levels of measurement and the "laws" of statistics', *Advances in Health Sciences Education*, 15(5), pp. 625–632. doi: 10.1007/s10459-010-9222-y

Noy, C. (2008) 'Sampling knowledge: The hermeneutics of snowball sampling in qualitative research', *International Journal of social research methodology*, 11(4), pp. 327–344. doi: 10.1080/13645570701401305

O'Connell, B. and Curry, B. (2022) *Stock Market Sector*. Available at: https://www.forbes.com/advisor/investing/stock-market-sectors/#financials-sector (Accessed: 8 April 2023).

Oh, S. et al. (2025) "Uncovering the Dark Patterns of Phishing Emails: An Eye-Tracking Analysis," IEEE Access, vol. 13, pp. 197634-197644, doi: 10.1109/ACCESS.2025.3633616.

O'Hara, K. *et al.* (2014) 'Everyday dwelling with WhatsApp', *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. doi: 10.1145/2531602.2531679.

Oliveira, D. *et al.* (2017) 'Dissecting spear phishing emails for older vs young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing', *Proceedings of the 2017 chi conference on human factors in computing systems*, pp. 6412–6424. doi: 10.1145/3025453.3025831

Onwuegbuche, F.C. et al. (2025) “Securing the Weakest Link: Exploring Affective States Exploited in Phishing Emails With Large Language Models,” ol. 13, pp. 173460-173486, 2025, doi: 10.1109/ACCESS.2025.3615788.

Orbell, S. *et al.* (2001) ‘The theory of planned behavior and ecstasy use: Roles for habit and perceived control over taking versus obtaining substances’, *Journal of Applied Social Psychology*, 31(1), pp. 31–47. doi:10.1111/j.1559-1816.2001.tb02480.x

Orunsolu, A. A. *et al.* (2017) ‘An empirical evaluation of security tips in phishing prevention: A case study of nigerian banks’, *International Journal of Electronics and Information Engineering*, 6(1), pp. 25–39.

Paas, L. J. and Morren, M. (2018) ‘Please do not answer if you are reading this: Respondent attention in online panels’, *Marketing Letters*, 29, pp. 13–21. doi: 10.1007/s11002-018-9448-7

Paganini, P. (2016) *Hackers find a way to send massive messages on Telegram*. Available at: https://securityaffairs.com/48371/hacking/send-massive-messages-telegram.html (Accessed 19 Sep. 25)

Pallant, J. (2007) *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows*. 3rd edn. New York: McGraw Hill Open University Press.

Park, G. (2018) *Towards Ontology-Based Phishing Detection*. Purdue University. Available at: https://docs.lib.purdue.edu/dissertations/AAI10831309/.

Park, G. and Taylor, J. M. (2015) ‘Using syntactic features for phishing detection’, *arXiv preprint arXiv:1506.00037*. doi: 10.48550/arXiv.1506.00037

Parker, H. J. and Flowerday, S. V (2020) ‘Contributing factors to increased susceptibility to social media phishing attacks’, *South African Journal of Information Management*, 22(1), pp. 1–10. Available at: https://journals.co.za/doi/abs/10.4102/sajim.v22i1.1176

Parsons, K. *et al.* (2013) 'Phishing for the truth: A scenario-based experiment of users' behavioural response to emails', *proceedings of the IFIP international information security conference*. Springer, pp. 366–378. doi:10.1007/978-3-642-39218-4_27

Parsons, K. *et al.* (2015) 'The design of phishing studies: Challenges for researchers', *Computers & Security*, 52, pp. 194–206. doi:10.1016/j.cose.2015.02.008

Parsons, K. *et al.* (2019) 'Predicting susceptibility to social influence in phishing emails', *International Journal of Human-Computer Studies*, 128, pp. 17–26. doi: 10.1016/j.ijhcs.2019.02.007

Pattinson, M. *et al.* (2012) 'Why do some people manage phishing e-mails better than others?', *Information Management & Computer Security*. 20 (1): 18–28. 10.1108/09685221211219173

Petelka, J., Zou, Y. and Schaub, F. (2019) 'Put your warning where your link is: Improving and evaluating email phishing warnings', *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–15. doi: 10.1145/3290605.330074

Phishing (2021) *History of Phishing*. Available at: https://www.phishing.org/history-of-phishing (Accessed: 25 January 2021).

PhishTank (2021) *What is phishing?* Available at: http://phishtank.org/what_is_phishing.php?view=website (Accessed: 26 January 2021).

Picchi, A. (2020) *Tribune workers got an email dangling a bonus — but it was a hoax from their employer*. Available at: https://www.cbsnews.com/news/tribune-bonus-email-hoax-cybersecurity-test/ (Accessed: 11 May 2025).

Piwek, L. and Joinson, A. (2016) '"What do they snapchat about?" Patterns of use in time-limited instant messaging service', *Computers in human behavior*, 54, pp. 358–367. doi: 10.1016/j.chb.2015.08.026

Prabakaran, M.K., Meenakshi Sundaram, P. and Chandrasekar, A.D. (2023) "An enhanced deep learning-based phishing detection mechanism to effectively identify malicious URLs using variational autoencoders," *IET Information Security*, 17(3), pp. 423–440. doi: 10.1049/ise2.12106

Punch, K. F. (2013) 'Introduction to social research: *Quantitative and qualitative approaches*'. Sage Publications

Quan-Haase, A. and Young, A. L. (2010) 'Uses and gratifications of social media: A comparison of Facebook and instant messaging', *Bulletin of science, technology & society*, 30(5), pp. 350–361. doi: 10.1177/027046761038000

Rahimnia, F. and Hassanzadeh, J. F. (2013) 'The impact of website content dimension and e-trust on e-marketing effectiveness: The case of Iranian commercial saffron corporations', *Information & Management*, 50(5), pp. 240–247. doi: 10.1016/j.im.2013.04.003

Rahman, M. L. *et al.* (2023) 'Users really do respond to smishing', *Proceedings of the thirteenth ACM conference on data and application security and privacy*, pp. 49–60. doi:10.1145/3577923.358364

Rao, R. S. and Pais, A. R. (2019a) 'Detection of phishing websites using an efficient feature-based machine learning framework', *Neural Computing and Applications*, 31, pp. 3851–3873. doi: 10.1007/s00521-017-3305-0

Rao, R. S. and Pais, A. R. (2019b) 'Jail-Phish: An improved search engine based phishing detection system', *Computers & Security*, 83, pp. 246–267.doi: 10.1016/j.cose.2019.02.011

Rao, R. S., Vaishnavi, T. and Pais, A. R. (2020) 'CatchPhish: detection of phishing websites by inspecting URLs', *Journal of Ambient Intelligence and Humanized Computing*, 11(2), pp. 813–825. doi: 10.1007/s12652-019-01311-4

Rashidi, Y., Vaniea, K. and Camp, L. J. (2016) 'Understanding Saudis' privacy concerns when using WhatsApp', *Proceedings of the Workshop on Usable Security (USEC'16)*, pp. 1–8. Available at: https://www.research.ed.ac.uk/en/publications/understanding-saudis-privacy-concerns-when-using-whatsapp

Reeder, R. W., Ion, I. and Consolvo, S. (2017) '152 simple steps to stay safe online: Security advice for non-tech-savvy users', *IEEE Security & Privacy*, 15(5), pp. 55–64. doi: 10.1109/MSP.2017.3681050.

Reinheimer, B. *et al.* (2020) 'An investigation of phishing awareness and education over time: When and how to best remind users', in *Sixteenth symposium on usable privacy and security (SOUPS 2020)*, pp. 259–284.

Rekouche, K. (2011) 'Early phishing', *arXiv preprint arXiv:1106.4692*. doi: 10.48550/arXiv.1106.4692

Renaud, K. and Flowerday, S. (2017) 'Contemplating human-centred security & privacy research: Suggesting future directions', *Journal of Information Security and Applications*, 34, pp. 76–81. doi: 10.1016/j.jisa.2017.05.006

Resende, G. *et al.* (2019) '(Mis)information dissemination in WhatsApp: Gathering, analyzing and countermeasures', in *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*. doi: 10.1145/3308558.3313688.

Reuning, K. and Plutzer, E. (2020) 'Valid vs. invalid straightlining: The complex relationship between straightlining and data quality', in *Survey Research Methods*, pp. 439–459. doi: 10.18148/srm/2020.v14i5.7641

Ribeiro, L., Guedes, I. S. and Cardoso, C. S. (2024) 'Which factors predict susceptibility to phishing? An empirical study', *Computers & Security*, 136, p. 103558. doi: 10.1016/j.cose.2023.103558

Rizzoni, F. *et al.* (2022) 'Phishing simulation exercise in a large hospital: A case study. DIGITAL HEALTH, 8, . doi: 10.1177/2055207622108171

Roberts, R. *et al.* (2019) 'You are who you appear to be: A longitudinal study of domain impersonation in tls certificates', *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pp. 2489–2504. doi: 10.1145/3319535.3363188.

Robila, S. A. and Ragucci, J. W. (2006) 'Don't be a phish: Steps in user education', in *Working Group Reports on ITiCSE on Innovation and Technology in Computer Science Education 2006*. doi: 10.1145/1140124.1140187.

Roy, S.S. et al. (2024) "From chatbots to phishbots?: Phishing scam generation in commercial large language models," in 2024 IEEE Symposium on Security and Privacy (SP). IEEE, pp. 36–54, doi: 10.1109/SP54263.2024.00182.

Safa, N. S. *et al.* (2015) 'Information security conscious care behaviour formation in organizations', *Computers & Security*, 53, pp. 65–78. doi: 10.1016/j.cose.2015.05.012.

Saleem Khawaja (2022) *Why email is still the number one threat vector*. Available at: https://www.verdict.co.uk/why-email-is-still-the-number-one-threat-vector/ (Accessed: 11 April 2025).

Sameen, M., Han, K. and Hwang, S. O. (2020) 'PhishHaven—an efficient real-time ai phishing URLs detection system', *IEEE Access*, 8, pp. 83425–83443. doi: doi: 10.1109/ACCESS.2020.2991403

Sandelowski, M. (2000) 'Whatever happened to qualitative description?', *Research in nursing & health*, 23(4), pp. 334–340.doi:10.1002/1098-240X(200008)23:4<334::AID-

Sarno, D. M. *et al.* (2017) 'Who are phishers luring?: a demographic analysis of those susceptible to fake emails',*Proceedings of the human factors and ergonomics society annual meeting*. SAGE Publications Sage CA: Los Angeles, CA, pp. 1735–1739. doi: 10.1177/15419312136019

Sarno, D. M. *et al.* (2020) 'Which phish is on the hook? Phishing vulnerability for older versus younger adults', *Human factors*, 62(5), pp. 704–717. doi: 10.1177/0018720819855570

Schiller, K. *et al.* (2024) 'Employees' Attitudes towards Phishing Simulations:" It's like when a child reaches onto the hot hob"', *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pp. 4167–4181. doi: 10.1145/3658644.3690212

Schmidt, W. C. (1997) 'World-Wide Web survey research: Benefits, potential problems, and solutions', *Behavior research methods, instruments, & computers*, 29(2), pp. 274–279. doi: 10.3758/BF03204826

Schonlau, M. and Toepoel, V. (2015) 'Straightlining in Web survey panels over time', in *Survey Research Methods*, pp. 125–137.doi: https:10.18148/srm/2015.v9i2.6128

Scotland, J. (2012) "Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms.," English language teaching, 5(9), pp. 9–16.

SecureOps (2023) FraudGPT' Malicious Chatbot Now for Sale on Dark Web. Available at: https://www.secureops.com/blog/ai-attacks-fraudgpt (Accessed: January 18, 2026).

Seng, S. *et al.* (2019) 'Poster: Understanding user's decision to interact with potential phishing posts on facebook using a vignette study', in *Proceedings of the ACM Conference on Computer and Communications Security*. doi: 10.1145/3319535.3363270.

Seng, S., Al-Ameen, M. N. and Wright, M. (2018) 'Understanding users' decision of clicking on posts in facebook with implications for phishing', in *Workshop on Technology and Consumer Protection (ConPro 18)*.

SentinelOne (2024) *What is Email Spoofing? Types & Examples*. Available at: https://www.sentinelone.com/cybersecurity-101/threat-intelligence/email-spoofing/ (Accessed: 23 April 2025).

Seufert, M. *et al.* (2015) 'Analysis of group-based communication in whatsapp', in *International conference on mobile networks and management*. Springer, pp. 225–238. doi:10.1007/978-3-319-26925-2_17

Shah, P. and Agarwal, A. (2020) 'Cybersecurity behaviour of smartphone users in India: an empirical analysis', *Information & Computer Security*, 28(2), pp. 293–318. doi: 10.1108/ICS-04-2019-0041.

Shalini, L. *et al.* (2022) 'Detection of phishing emails using machine learning and deep learning', in *2022 7th International conference on communication and electronics systems (ICCES)*. IEEE, pp. 1237–1243. doi: 10.1109/ICCES54183.2022.9835846.

Shamy, N. El, Xie, W. and Zhong, C. (2024) 'Smishing: Exploring How Different Persuasion Techniques Influence Users Emotions, Cognitions, and Identification Accuracy', in *NeuroIS Retreat*. Springer, pp. 345–355. doi: 10.1007/978-3-031-71385-9_30

Sharevski, F. *et al.* (2022) 'Phishing with malicious QR codes', *Proceedings of the 2022 European Symposium on Usable Security*, pp. 160–171. doi: 10.1145/3549015.3554172

Shenoy, N. and Mbaziira, A. v (2024) "An extended review: LLM prompt engineering in cyber defense," in 2024 International Conference on Electrical, Computer and Energy Technologies (ICECET. IEEE, pp. 1–6, doi: 10.1109/ICECET61485.2024.10698605.

Sheng, S. *et al.* (2009) 'An empirical analysis of phishing blacklists', proceedings of the *6th Conference on Email and Anti-Spam*.

Sheng, S. *et al.* (2010) 'Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions', *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 373–382. doi: 10.1145/1753326.1753383

Sheridan, K. (2021) *North Korean Attackers Target Security Researchers via Social Media: Google*, *Dark Reading*. Available at: https://www.darkreading.com/endpoint/north-korean-attackers-target-security-researchers-via-social-media-google/d/d-id/1339988 (Accessed: 29 January 2021).

Shimoga, S. V, Erlyana, E. and Rebello, V. (2019) 'Associations of social media use with physical activity and sleep adequacy among adolescents: Cross-sectional survey', *Journal of medical Internet research*, 21(6), p. e14290.doi: 10.2196/14290

Shirazi, H., Bezawada, B. and Ray, I. (2018) '" Kn0w Thy Doma1n Name" Unbiased Phishing Detection Using Domain Name Based Features', *Proceedings of the 23nd ACM on symposium on access control models and technologies*, pp. 69–75.doi: 10.1145/3205977.3205992

Sillence, E. *et al.* (2006) 'A framework for understanding trust factors in web-based health advice', *International Journal of Human-Computer Studies*, 64(8), pp. 697–713. doi: 10.1016/j.ijhcs.2006.02.007

Singkeruang, A. W. T. F., Susanto, S. E. and Saeni, N. (2025) 'Mitigating the Risk of Qushing Threats (QR Phishing) using the Security Behavior Intentions Scale (SeBIS) in supporting digital economic security', *Paradoks: Jurnal Ilmu Ekonomi*, 8(2), pp. 685–696. doi: 10.57178/paradoks.v8i2.1196

Sjouwerman, S. (2025) *Email Remains the Top Attack Vector for Cyberattacks*. Available at: https://blog.knowbe4.com/email-remains-the-top-attack-vector-for-cyberattacks (Accessed: 22 July 2025).

Smadi, S., Aslam, N. and Zhang, L. (2018) 'Detection of online phishing email using dynamic evolving neural network based on reinforcement learning', *Decision Support Systems*. doi: 10.1016/j.dss.2018.01.001.

Smadi, S. M. (2017) *Detection of Online Phishing Email using Dynamic Evolving Neural Network Based on Reinforcement Learning*. Northumbria University. Available at: http://nrl.northumbria.ac.uk/36119/.

Smith, K. L. (2020) *Employment Phishing Scam: Example*. Available at: https://www.linkedin.com/pulse/employment-phishing-scam-example-kevin-leo-smith/.

Sowerby, C. (2023) *BEWARE OF THIS BOOKING.COM PHISHING SCAM*. Available at: https://insideflyer.co.uk/2023/01/beware-of-this-booking-com-phishing-scam/ (Accessed: 13 August 2023).

Sreejesh, S. *et al.* (2014) 'Business research design: Exploratory, descriptive and causal designs', *Business research methods: An applied orientation*, pp. 25–103. doi: 10.1007/978-3-319-00539-3_3

Stajano, F. and Wilson, P. (2011) 'Understanding scam victims: Seven principles for systems security', *Communications of the ACM*. doi: 10.1145/1897852.1897872.

Su, M.-Y. and Su, K.-L. (2023) “Bert-based approaches to identifying malicious urls,” Sensors, 23(20), p. 8499. doi: 10.3390/s23208499

Sullivan, G.M. and Artino Jr, A.R. (2013) “Analyzing and interpreting data from Likert-type scales,” Journal of graduate medical education, 5(4), pp. 541–542. doi: 10.4300/JGME-5-4-18

Di Stasio, V. (2014) *Why education matters to employers: a vignette study in Italy, England and the Netherlands*. Universiteit van Amsterdam [Host].

Statista (2024) *Number of e-mail users worldwide from 2018 to 2027*. Available at: https://www.statista.com/statistics/255080/number-of-e-mail-users-worldwide/ (Accessed: 11 April 2025).

Statista (2025) *Number of internet and social media users worldwide as of February 2025*. Available at: https://www.statista.com/statistics/617136/digital-population-worldwide/ (Accessed: 19 March 2025).

Stemler, S. (2000) ‘An overview of content analysis’, *Practical assessment, research, and evaluation*, 7(1), p. 17. doi:10.7275/z6fm-2e34

Stevens, L. and Wrenn, C. (2013) ‘Exploratory (qualitative) research’, *Concise encyclopedia of church and religious organization marketing*, 53.

Stivala, G. and Pellegrino, G. (2020) ‘Deceptive Previews: A Study of the Link Preview Trustworthiness in Social Platforms’, in *27th Annual Network and Distributed System Security symposium, February 2020, San Diego. Conference: NDSS Network and Distributed System Security Symposium*.

Stojnic, T., Vatsalan, D. and Arachchilage, N. A. G. (2021) ‘Phishing email strategies: understanding cybercriminals’ strategies of crafting phishing emails’, *Security and privacy*, 4(5), p. e165. doi: 10.1002/spy2.165

Sturman, D. *et al.* (2024) 'The roles of phishing knowledge, cue utilization, and decision styles in phishing email detection', *Applied Ergonomics*, 119, p. 104309. doi: 10.1016/j.apergo.2024.104309

Sultan, A. J. (2014) 'Addiction to mobile text messaging applications is nothing to "lol" about', *Social Science Journal*, 51(1), pp.57-69. doi: 10.1016/j.soscij.2013.09.003.

Sun, J. C.-Y. *et al.* (2016) 'The mediating effect of anti-phishing self-efficacy between college students' internet self-efficacy and anti-phishing behavior and gender difference', *Computers in Human Behavior*, 59, pp. 249–257. doi: 10.1016/j.chb.2016.02.004

Swedberg, R. (2020) 'Exploratory research', *The production of knowledge: Enhancing progress in social science*, 2(1), pp. 17–41.

Tabassum, S., Faklaris, C. and Lipford, H. R. (2024) 'What Drives {SMiShing} Susceptibility? A {US}. Interview Study of How and Why Mobile Phone Users Judge Text Messages to be Real or Fake', in *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*, pp. 393–411.

Taherdoost, H. (2022) 'What are different research approaches? Comprehensive Review of Qualitative, quantitative, and mixed method research, their applications, types, and limitations', *Journal of Management Science & Engineering Research*, 5(1), pp. 53–63. doi: ff10.30564/jmser.v5i1.4538ff.ffhal-0374184

Taib, R. *et al.* (2019) 'Social engineering and organisational dependencies in phishing attacks', *Proceedings of Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Part I*. Springer, pp. 564–584. doi:10.1007/978-3-030-29381-9_35

Taivaskero, T. (2017) *The history of mobile instant messaging by Jongla*. Available at: https://www.linkedin.com/pulse/history-mobile-instant-messaging-jongla-timo-mäkelä/ (Accessed: 16 March 2025).

Tashakkori, A. and Creswell, J. W. (2007) 'The new era of mixed methods', *Journal of mixed methods research*. Sage Publications, pp. 3–7. doi: 10.1177/23456789062930

Tatango (2020) *Best URL Shorteners for SMS Marketing*. Available at: https://www.tatango.com/blog/best-url-shorteners-for-sms-marketing/ (Accessed: 9 June 2021).

techradar (2025) 'Phishing campaign targets prominent X users, accounts at risk'. Available at: https://www.techradar.com/pro/security/phishing-campaign-targets-prominent-x-users-accounts-at-risk (Accessed: 29 April 2025).

Thao, T. P. *et al.* (2020) 'Human factors in homograph attack recognition', *Proceedings of the Applied Cryptography and Network Security: 18th International Conference, ACNS 2020, Rome, Italy, October 19–22, 2020, Proceedings, Part II 18*. Springer, pp. 408–435. doi: 10.1007/978-3-030-57878-7_20

Tjostheim, I. and Waterworth, J. A. (2020) 'Predicting personal susceptibility to phishing', in *Information Technology and Systems: Proceedings of ICITS 2020*. Springer, pp. 564–575. doi: 10.1007/978-3-030-40690-5_54

Torten, R., Reaiche, C. and Boyle, S. (2018) 'The impact of security awarness on information technology professionals' behavior', *Computers & Security*, 79, pp. 68–79. doi: 10.1016/j.cose.2018.08.007.

Trend Micro (2024) *What Is Social Media Phishing?* Available at: https://www.trendmicro.com/en_gb/what-is/phishing/types-of-phishing/social-media-phishing.html (Accessed: 19 March 2025).

Tschakert, K. F. and Ngamsuriyaroj, S. (2019) 'Effectiveness of and user preferences for security awareness training methodologies', *Heliyon*, 5(6), p. e02010. doi: 10.1016/j.heliyon.2019.e02010

Ugwu, C.I., Ekere, J.N. and Onoh, C. (2021) "Research paradigms and methodological choices in the research process," *Journal of Applied Information Science and Technology*, 14(2), pp. 116–124.

Vaismoradi, M., Turunen, H. and Bondas, T. (2013) 'Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study', *Nursing & health sciences*, 15(3), pp. 398–405. doi: 10.1111/nhs.12048

Valdez, J. (2016) 'safebrowsing.py - Google Safe Browsing Lookup API (v4)'. Available at: https://github.com/junv/safebrowsing. (Accessed 20 Sep. 25)

Valecha, R., Mandaokar, P. and Rao, H. R. (2021) 'Phishing email detection using persuasion cues', *IEEE transactions on Dependable and secure computing*, 19(2), pp. 747–756. doi: 10.1109/TDSC.2021.3118931

Van Selm, M. and Jankowski, N. W. (2006) 'Conducting online surveys', *Quality and quantity*, 40, pp. 435–456. doi: 10.1007/s11135-005-8081-8

Varshney, U. *et al.* (2002) 'Voice over IP', *Communications of the ACM*, 45(1), pp. 89–96.

Vasiliu, C. *et al.* (2023) 'Exploring the Advantages of Using Social Media in the Romanian Retail Sector', *Journal of Theoretical and Applied Electronic Commerce Research*, 18(3), pp. 1431–1445. doi:10.3390/jtaer18030072

Vishwanath, A. *et al.* (2011) 'Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model', *Decision Support Systems*, 51(3), pp. 576–586. doi:10.1016/j.dss.2011.03.002

Vishwanath, A. (2015) 'Habitual Facebook use and its impact on getting deceived on social media', *Journal of Computer-Mediated Communication*, 20(1), pp. 83–98. doi: 10.1111/jcc4.12100

Vishwanath, A. (2016) 'Mobile device affordance: Explicating how smartphones influence the outcome of phishing attacks', *Computers in Human Behavior*, 63, pp. 198–207. doi: 10.1016/j.chb.2016.05.035

Vishwanath, A. (2017) 'Getting phished on social media', *Decision Support Systems*, 103, pp. 70–81. doi:10.1016/j.dss.2017.09.004

Vishwanath, A., Harrison, B. and Ng, Y. J. (2018) 'Suspicion, cognition, and automaticity model of phishing susceptibility', *Communication research*, 45(8), pp. 1146–1166. doi: 10.1177/009365021562748

Voit, A. *et al.* (2019) 'Online, VR, AR, Lab, and In-Situ: Comparison of Research Methods to Evaluate Smart Artifacts', *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. doi: 10.1145/3290605.330073

Volkamer, M., Renaud, K. and Reinheimer, B. (2016) 'Torpedo: tooltip-powered phishing email detection', in *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer, pp. 161–175. doi: 10.1007/978-3-319-33630-5_12

Volkamer, M., Sasse, M. A. and Boehm, F. (2020) 'Analysing simulated phishing campaigns for staff', in *Computer Security: ESORICS 2020 International Workshops, DETIPS, DeSECSys, MPS, and SPOSE, Guildford, UK, September 17–18, 2020, Revised Selected Papers 25*. Springer, pp. 312–328. doi: 10.1007/978-3-030-66504-3_19

Wadhwa, V. (2018) *WhatsApp public groups can leave user data vulnerable to scraping*. Available at: https://venturebeat.com/2018/04/03/whatsapp-public-groups-can-leave-user-data-vulnerable-to-scraping/ (Accessed: 20 June 2020).

Wallander, L. (2009) '25 years of factorial surveys in sociology: A review', *Social science research*, 38(3), pp. 505–520. doi: 10.1016/j.ssresearch.2009.03.004

Wang, J., Li, Y. and Rao, H. R. (2017) 'Coping responses in phishing detection: an investigation of antecedents and consequences', *Information Systems Research*, 28(2), pp. 378–396. doi:10.1287/isre.2016.0680

van de Weijer, S. G. A., Leukfeldt, R. and Zee, S. van der (2021) 'Cybercrime reporting behaviors among small-and medium-sized enterprises in the Netherlands', in *Cybercrime in Context*. Springer, pp. 303–325. doi:10.1007/978-3-030-60527-8_17

Wetzig, C. (2022) *15 Alarming Cybersecurity Facts And Statistics*. Available at: https://thrivedx.com/resources/article/cyber-security-facts-statistics?referrer=cybint (Accessed: 16 May 2023).

Williams, E. J., Hinds, J. and Joinson, A. N. (2018) 'Exploring susceptibility to phishing in the workplace', *International Journal of Human-Computer Studies*, 120, pp. 1–13. doi: 10.1016/j.ijhcs.2018.06.004

Williams, E. J., Morgan, P. L. and Joinson, A. N. (2017) 'Press accept to update now: Individual differences in susceptibility to malevolent interruptions', *Decision Support Systems*, 96, pp. 119–129. doi:10.1016/j.dss.2017.02.014

Williams, E. J. and Polage, D. (2019) 'How persuasive is phishing email? The role of authentic design, influence and current events in email judgements', *Behaviour & Information Technology*, 38(2), pp. 184–197. doi: 10.1080/0144929X.2018.1519599

Windels, J. (2018) *New family of mobile phishing attacks preying on WhatsApp users*. Available at: https://www.wandera.com/malware-family-whatsapp/ (Accessed: 20 November 2020).

Wired (2020) *The Attack That Broke Twitter Is Hitting Dozens of Companies*. Available at: https://www.wired.com/story/phone-spear-phishing-twitter-crime-wave/ (Accessed: 20 March 2025).

Witte, K. (1992) 'Putting the fear back into fear appeals: The extended parallel process model', *Communications Monographs*, 59(4), pp. 329–349. doi: 10.1080/03637759209376276

WOT (2020) *Protect Your Browsing with Web Of Trust*. Available at: https://www.mywot.com (Accessed: 11 May 2020).

Wright, R. T. *et al.* (2014) 'Research note—influence techniques in phishing attacks: an examination of vulnerability and resistance', *Information systems research*, 25(2), pp. 385–400. doi: 10.1287/isre.2014.0522

Wu, M., Miller, R. C. and Garfinkel, S. L. (2006) 'Do security toolbars actually prevent phishing attacks?', in *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/1124772.1124863.

Yeboah-Boateng, E. O. and Amanor, P. M. (2014) 'Phishing, SMiShing & Vishing: an assessment of threats against mobile devices', *Journal of Emerging Trends in Computing and Information Sciences*, 5(4), pp. 297–307.

Yu, X. and Khazanchi, D. (2017) 'Using embedded mixed methods in studying is phenomena: Risks and practical remedies with an illustration', *Using Embedded Mixed Methods in Studying IS Phenomena: Risks and Practical Remedies with an Illustration*, 41.doi:

ZELTSER, L. (2021) *Free Blocklists of Suspected Malicious IPs and URLs*. Available at: https://zeltser.com/malicious-ip-blocklists/ (Accessed: 30 January 2021).

von Zezschwitz, E., Chen, S. and Stark, E. (2022) '" It builds trust with the customers"-Exploring User Perceptions of the Padlock Icon in Browser UI', in

2022 IEEE Security and Privacy Workshops (SPW). IEEE, pp. 44–50. doi: 10.1109/SPW54247.2022.9833869

Zhang, C. and Conrad, F. (2014) 'Speeding in web surveys: The tendency to answer very fast and its association with straightlining', in *Survey research methods*, pp. 127–135. doi: 10.18148/srm/2014.v8i2.5453

Zhang, Y., Hong, J. I. and Cranor, L. F. (2007) 'Cantina: A content-based approach to detecting phishing web sites', in *16th International World Wide Web Conference, WWW2007*. doi: 10.1145/1242572.1242659.

Zhuo, S. *et al.* (2023) 'SoK: Human-centered phishing susceptibility', *ACM Transactions on Privacy and Security*, 26(3), pp. 1–27. doi: 10.1145/3575797

Zielinska, O. A. *et al.* (2016) 'A temporal analysis of persuasion principles in phishing emails', *Proceedings of the human factors and ergonomics society annual meeting*. SAGE Publications Sage CA: Los Angeles, CA, pp. 765–769. doi: 10.1177/154193121360117

Zikmund, W. G. *et al.* (2000) *Business research methods*. Dryden Press Fort Worth, TX.

Appendix A: Information sheet and questionnaire for study on the behaviours of mobile instant messaging

Appendix A.1

Information sheet

Dear Participant,

Thank you for your interest in participating in this study. Before you decide to participate in this study, it is important that you understand why the research is being done and what it will involve. Please read the following information carefully. You can also ask the researcher if there is anything that is not clear or if you need further information. Once you have read the information, and have decided to participate in the study, please click 'I agree to participate in this study' to start the questionnaire.

Information and Consent Form

You are invited to participate in research conducted by Rufai Ahmad who is a PhD student with the Department of Computer and Information Sciences University of Strathclyde, Glasgow, UK. The research is about the behaviours of mobile instant messaging (MIM) applications (apps) users towards links shared on such platforms. Participation should take approximately about 15 minutes.

Purpose of the Research

The overarching goal of the study is to understand whether the behaviours of users of mobile instant messaging applications expose them to security risks such as phishing. The survey hopes to provide insights about users' attitudes towards links that are shared

on such platforms and the factors that influence their decision to perform certain actions such as clicking and forwarding of these links. The findings from this research are expected to identify potential gaps in current strategies for mitigating security threats in instant messaging applications.

Participation

Your participation in this research study is voluntary. You may refuse to take part in the research or exit the survey at any time without penalty. You may skip any question you do not wish to answer for any reason. This survey does not link your identity to your responses. Therefore, it will be impossible to delete your responses once they are submitted (at the end of the survey). To avoid this, you can withdraw from the survey by closing your browser (tab) at any point of the survey.

What will you do in the project?

This study requires you to respond to questions in an online survey.

Why have you been invited?

You have been chosen to take part in this study because you are 18 years or older and you are a current user of an instant messaging application.

Potential Harm or Risks

There are no physical, psychological or social risks that may result from participating in this research project.

Data Collection Procedures

The survey consists of questions relating to your demographic, general IT knowledge and skills. In addition, the survey also consists of questions relating to your behaviours towards links that are shared on mobile instant messaging applications. The survey will be hosted and administered by Qualtrics.com.

Privacy and Confidentiality

The data collected in this survey will be kept confidential and respondents will not be asked for any personally identifiable information. Responses from this survey will initially be stored on Qualtrics in a password protected electronic format. Data will later be downloaded and stored digitally in a password-protected excel file on the principal's researcher's computer.

Conservation of Data

After completion of the research, data will be deposited into the University of Strathclyde, institutional data repository, Pure for a minimum period of 10 years. Data will also be stored on the principal's researcher's H and I: drive and Strathcloud which is a cloud service run by the University of Strathclyde and has multiple layers of security.

Potential Benefits

You will receive no direct benefit from participating in this research study. However, your responses may help us to learn more about the behaviours of users towards links shared on mobile instant messaging applications. In addition, your participation will help us to come up with effective measures to protect users against security threats on mobile instant messaging applications.

Results of the Research Project

You can always request a summary of the findings after the research has been completed. If you wish to receive a summary, please send an email to the principal researcher whose email is given below. In addition, the results of this study may be published in academic and professional journals and presented at conferences.

Researcher contact details:

Rufai Ahmad

PhD Student

Department of Computer and Information Sciences

Livingstone Tower

Richmond Street Glasgow G1 1XH

Email: rufai.ahmad@strath.ac.uk.

This research was granted ethical approval by the University of Strathclyde Department of Computer Science Ethics Committee. If you have any questions/concerns, during or after the research, or wish to contact an independent person to whom any questions may be directed or further information may be sought, please contact:

Secretary to the Departmental Ethics Committee

Department of Computer and Information Sciences,

Livingstone Tower Richmond Street Glasgow G1 1XH

email: ethics@cis.strath.ac.uk

Consent ELECTRONIC CONSENT: Please select your choice below. You may print a copy of this consent form for your records. Clicking on the "Agree" button indicates that

. You have read the above information

. You voluntarily agree to participate

. You are 18 years of age or older

○ **I AGREE to participate in this study (1)**

○ **I DECLINE to participate in this study (2)**

Appendix A.2

Survey on the behaviours of mobile instant messaging app users towards shared links

Q1. Age_group Please specify your age group.

o 18-30

o 31-45

o 46-65

o 65 and above

Q2. Please specify your gender.

o Male

o Female

o Prefer not to disclose

Q3. Please specify your level of education.

o Secondary

o Further education (i.e. colleges)

o University undergraduate

o Postgraduate (e.g. Masters, Doctorate)

Q4. In which country do you currently reside?

Q5. Please indicate how frequently you used any of the following devices

	Never (1)	Rarely (2)	At least once a month (3)	At least once a week (4)	Multiple times a week (5)	Everyday (6)
Desktop (1)	○	○	○	○	○	○
Laptop (2)	○	○	○	○	○	○
Smartphone (3)	○	○	○	○	○	○
Tablet (5)	○	○	○	○	○	○
Other ICT devices (e.g. smart watch) (4)	○	○	○	○	○	○

Q6. The following questions measures your perceptions of your ability to use mobile phones to accomplish certain tasks. Please indicate whether you agree or disagree with the following statements.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
I believe I have the ability to make purchases using a mobile device (1)	○	○	○	○	○
I believe I have the ability to identify and correct common problems with mobile devices (2)	○	○	○	○	○
I believe I have the ability to install and remove features and applications from mobile devices (3)	○	○	○	○	○
I believe I have the ability to use the productivity features offered by mobile devices (e.g., calendar, email, task scheduling, etc.). (4)	○	○	○	○	○

MIM usage section This section ask about your current mobile instant messaging applications usage. Mobile instant messaging applications are applications that allow their users to communicate with other users on their smartphones in real time. Examples include WhatsApp, Telegram, Signal, Messenger, etc

Q7. Which of the following mobile instant messaging applications do you currently have an active account with and use? (Check all that apply)

- ▢ WhatsApp
- ▢ Line
- ▢ Viber
- ▢ Telegram
- ▢ Slack
- ▢ Signal

☐ ⊗None

Q8. How frequently do you use the following applications?

	Never (1)	Rarely (2)	Sometimes (3)	Very often (4)	Always (5)
WhatsApp (1)	○	○	○	○	○
Line (2)	○	○	○	○	○
Viber (3)	○	○	○	○	○
Telegram (4)	○	○	○	○	○
Slack (5)	○	○	○	○	○
Signal (6)	○	○	○	○	○

Q9. Who do you use these applications to communicate with? (Check all that apply)

☐ Family

☐ Friends

☐ Work/business colleagues

☐ Contact you know, but aren’t friends and family or work/business colleagues

☐ Contact you don't know

☐ ⊗Not applicable

Link clicking behaviour

For the mobile instant messaging application you use, answer the following questions relating to your link clicking behaviour.

Q10. How frequently do you click links when communicating in these Applications?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q 11. How frequently do you consider the sender of a message before clicking the link in that message?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q12. How frequently do you click on links sent to you by your friends?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q13. How frequently do you click on links sent to you by your family members?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q14. How frequently do you click on links sent to you by your work/business colleagues?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q15. How frequently do you click on links sent to you by contact you know, but are not your friends, family or work/business colleagues?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q16. How frequently do you click on links sent to you by contact you do not know?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q17. How often do you check the actual links sent during chats before clicking them?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q18. Preview Most mobile instant messaging applications use link previews to create a visual preview and a brief description of a link that is shared by users. The preview contains both the actual URL shared by users and other information, including an image, hostname, title and summary of the content of the Webpage. How frequently do you check the link preview before clicking a link?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Group_behaviour For the mobile instant messaging application you use, answer the following questions relating to your membership of groups and your link clicking behaviours in those groups.

Q19. Are you a member of any group?

o For all the applications

o For most of the application

o For some of the application

o For none of the applications

Groups_definition There are two types of groups in mobile instant messaging applications, private and public ones. Private groups are groups that can only be joined by invitation, and their members have some offline connections (i.e groups of students enrolled in a particular course, employees of an organisation), while public groups are groups that are openly accessible and can be join by anyone with an interest to join. Links to joining public groups are often

Q20. publicised on well-known public websites and are themed around various topics such as politics, sports, programming, etc.

Do you know whether the groups you are a member of are private or public?

o For all of them

o For most of them

o For some of them

o For none of them

Q21. For the groups you are part of, how frequently do you click on links sent by other members?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q22. For the groups you are part of, how frequently do you click on links that are sent in such groups by your friends?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q23. For the groups you are part of, how frequently do you click on links that are sent in such groups by your family members?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q24. For the groups you are part of, how frequently do you click on links that are sent in such groups by your work/business colleagues?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q25. For the groups you are part of, how frequently do you click on links that are sent in such groups by contacts you know, but aren't friends and family or work/business colleagues?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q26. For the groups you are part of, how frequently do you click on links that are sent in such groups by contacts you don't know?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

For the mobile instant messaging application you use, answer the following questions relating to your links forwarding behaviour.

Q27. How frequently do you forward links to others?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q28. Do you follow links before forwarding them?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q29. How frequently do you consider who the sender of a link is before forwarding it to others?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q30. How frequently do you forward links that are shared with you by others?

o Never (1)

o Rarely (2)

o Sometimes (3)

o Very often (4)

o Always (5)

Q31. How frequently do you forward links that are:

	Never (1)	Rarely (2)	Sometimes (3)	Very often (4)	Always (5)
Shared in public groups (3)	○	○	○	○	○
Shared in private groups (4)	○	○	○	○	○

The following questions ask about your knowledge of phishing scams, your concern about such scams in mobile instant messengers and how you deal with such scams.

Q32. Please indicate whether you agree or disagree with the following.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
I know what phishing is. (1)	○	○	○	○	○

Q33. Are you aware of phishing scams?

o Yes (1)

o No (2)

Q34. Are you aware of any mobile instant messaging scams?

o Yes (1)

o No (2)

Q35. How concerned are you about phishing scams on mobile instant messaging applications?

o Not at all (1)

o Slightly (2)

o Somewhat (3)

o Quite (4)

o Extremely (5)

Q36. Which of these statements best reflect how you respond to the problem of phishing scams on mobile instant messaging applications? (Select all that apply)

▢ I look for suspicious links

▢ I check the link preview

▢ I do not click links from unfamiliar senders

▢ I search google for further information on the website before given out my details

▢ ⊗I do nothing

The following questions measures your perceptions of your ability to protect yourself from phishing scams in mobile instant messaging applications. It also measures your perception of the ability of others (i.e your contacts and the platforms) to protect you from such scams.

Q37. Please indicate whether you agree or disagree with the statements below.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
I believe I have the knowledge and skills to identify phishing scams when they are presented to me during communication on mobile instant messengers. (1)	○	○	○	○	○
I believe that I can protect my personal information from being stolen by phishers. (2)	○	○	○	○	○
I believe I have the skills to implement security measures to stop people from getting my confidential information. (3)	○	○	○	○	○
My contacts have the required skills and knowledge to detect phishing links and therefore can protect me. (4)	○	○	○	○	○
I believe mobile instant messengers have mechanisms in place to	○	○	○	○	○

Q38. Please select which of the following statements best reflects your views of whose responsibility is to protect users from phishing scams on mobile instant messaging apps

o I believe it is the responsibility of mobile instant messaging platforms to protect their users from phishing scams.

o I believe mobile instant messengers' users should be responsible for protecting themselves from phishing scams.

o Protecting users from phishing scams should be a shared responsibility between mobile instant messaging platforms and their users.

Appendix B: Codebook for the study "Content analysis of mobile instant messaging phishing attacks "

Appendix B.1

Codebook for content analysis

Category Name	Description	Example
Authority	• Words expressing authority • Words expressing fear • Words expressing tight deadlines/urgency • Graphics, logos, contact and copyright information from well-known reputable companies or banks • Above were adapted from	"Regards Amazon Security Teams" (Baryshevtsev and McGlynn, 2020) "CONFIRM YOUR PERSONAL INFORMATION PAYPAL" (Ferreira and Teles, 2019) "your account will be blocked." (Ferreira and Lenzini, 2015)

	(Ferreira and Lenzini, 2015)	
Commitment and Consistency	• Leverage an individual's previous commitment or consistency with a previous situation • The content implies that the victim has already committed a similar call (Jones *et al.*, 2020) • The content implies that the victim is committed to helping them based on their role or membership (Jones *et al.*, 2020)	"As you have done before, please update your password using the following link "(Wright *et al.*, 2014)
Liking	• Content appears to be helping the recipient. (Baryshevtsev and McGlynn, 2020) • The content appears polite. (Baryshevtsev and McGlynn, 2020)	"a scam email from a supposed friend of the recipient asking him/her to visit a website that is really interesting."(Ferreira and Jakobsson, 2016)

	• Content seems to come from those the recipient like (i.e. friends or those they share similarities with, such as group members, same gender etc.).	
Reciprocity	• The content shows the recipient has to perform an action in response to a favour • The content of the message tries to obligate the victim to reciprocate a kind gesture (Jones *et al.*, 2020) • The content of the message state or imply that the victim has already committed to helping them (the scammer)? (Jones *et al.*, 2020)	"If you respond, you will have a chance to upgrade to our new version."(Akdemir and Yenal, 2021)
		"You can save 20% on your next electric bill by filling out our online survey within the next three days."(Oliveira *et al.*, 2017)

Scarcity	• The content of the message shows that the recipient can miss out on offers by not acting immediately (Ferreira and Lenzini, 2015) • Words expressing tight deadlines (Ferreira and Lenzini, 2015) • Graphics, logos, pictures, animated graphics, copyright information from well-known reputable companies or banks, or security organisations, as well as different fonts, sizes, colours and links (Ferreira and Lenzini, 2015)	”your account will be blocked” (Ferreira and Lenzini, 2015)
Social proof	• The content contains a reference to known people/colleagues	“I would like to invite you to join the thousands of other clients who have experienced our vacation excursions” (Oliveira *et al.*, 2017)

	(Ferreira and Lenzini, 2015) • The content shows that other users have acted in this manner in the past (Jones *et al.*, 2020) • Does the content imply that complying with the scammer's request will have benefits (Jones *et al.*, 2020) • The content implies that if the victim does not comply with the request, the victim will be "left out" in some way (Jones *et al.*, 2020)	
Fear	• The content induces fear or worries through the threat of some sort of consequences	"Your account will be permanently closed if you do not respond."(Baryshevtsev and McGlynn, 2020)
Source Credibility	• The message contains domain names that imitate the target	Messages contain logos of known reputable organisations

	organisations, official company logos, telephone numbers and email addresses of credible sources	
Urgency	• The message creates a sense of urgency and the need for the user to take immediate action	"you need to update your account before the link expires, after 24 hours"(Park and Taylor, 2015). "You have 24 hours to respond ." (Baryshevtsev and McGlynn, 2020)
URL/domain obfuscation techniques Similar Domains IP-address Letter substitution Complex URLs	• The URL contains a domain that is similar to known legitimate addresses • The URL displays an IP address	"amazon.checkingoutbooksonline.ca" (Lin *et al.*, 2011) http//192.168.111.112/login (Lin *et al.*, 2011) www.uca1gary.ca rather than www.ucalgary.ca (Lin *et al.*, 2011) http://www.amazon.ca.checkingoutbookonline.ca/Golden-Mean-Annabel-Lyon/ (Lin *et al.*, 2011)

	rather than the domain name • One or more letters/characters in the domain are substituted with one or more characters that look similar to those in actual domains • The length of the URL is extended with illogical characters making it difficult to read	
Tiny URL	Is the link a tiny url? Adapted from (Sameen, Han and Hwang, 2020)	

Appendix B.2

Sample Coding Sheet

S/N	Input Data/Image	Excerpts from data	Category1: Reciprocity	Category 2: Authority	Category 3	Category4
1		"Free Voucher"	✔			

Appendix B.3

Below redirects to the University of Strathclyde data repository containing the phishing messages used in Chapter 5

https://doi.org/10.15129/2a93dde8-f088-4762-ba34-c8215498b697

Appendix C: EXPERIMENTAL VIGNETTES AND SURVEY ON THE EFFECT OF LINK PREVIEW AND MESSAGE SENDER ON SUSCEPTIBILITY TO PHISHING IN MIM APPS

APPENDIX C.1

Full list of experimental vignettes

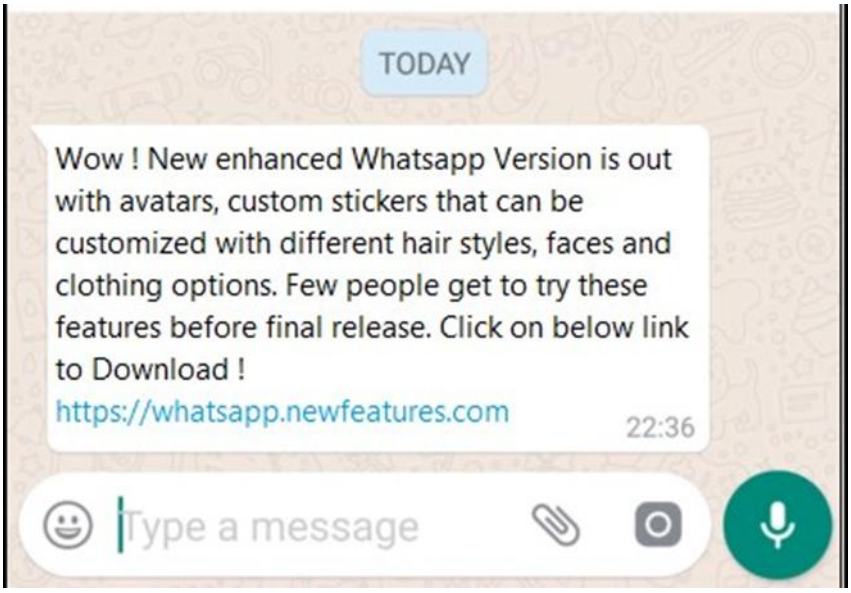


Fig 1. WhatsApp enhancement

Fig 2. Qatar Airways giveaway

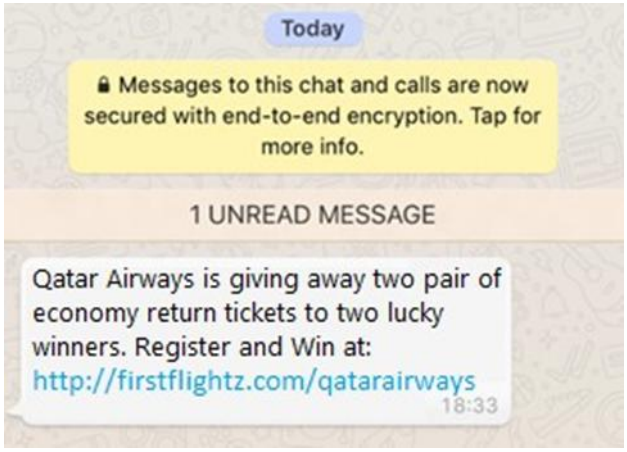


Fig3. Netflix Discount

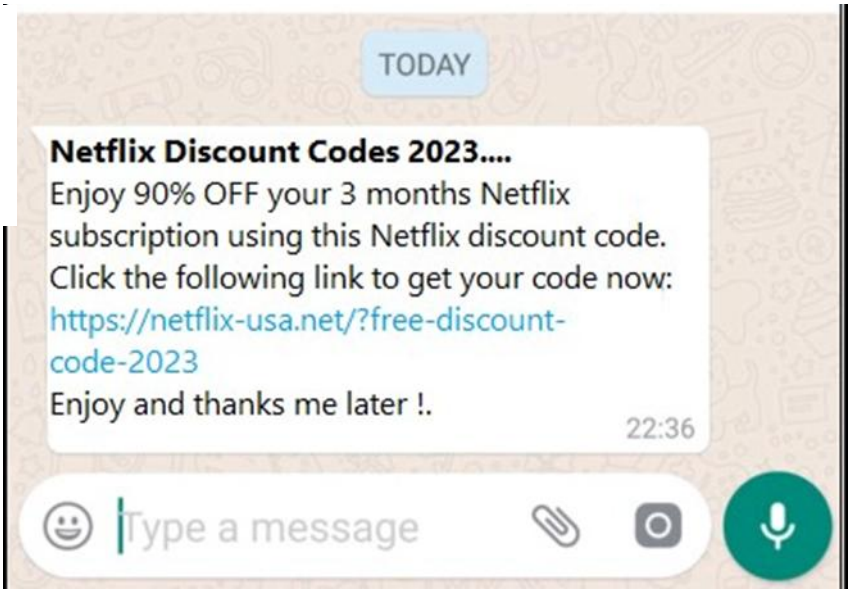


Fig 4. Nettle giveaway

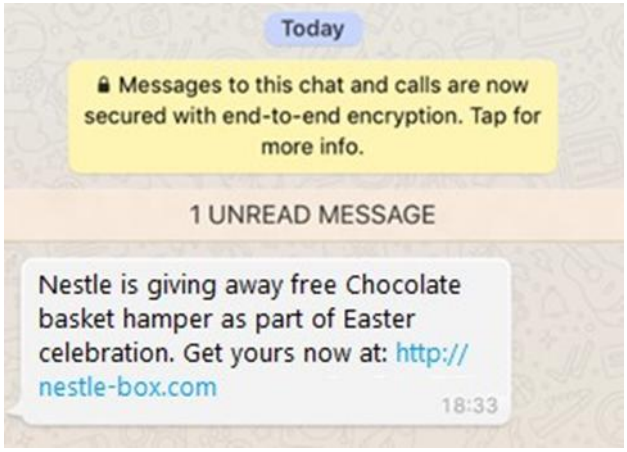

Today
Messages to this chat and calls are now secured with end-to-end encryption. Tap for more info.
1 UNREAD MESSAGE
Nestle is giving away free Chocolate basket hamper as part of Easter celebration. Get yours now at: http://nestle-box.com
18:33

Fig 5. Binance account security

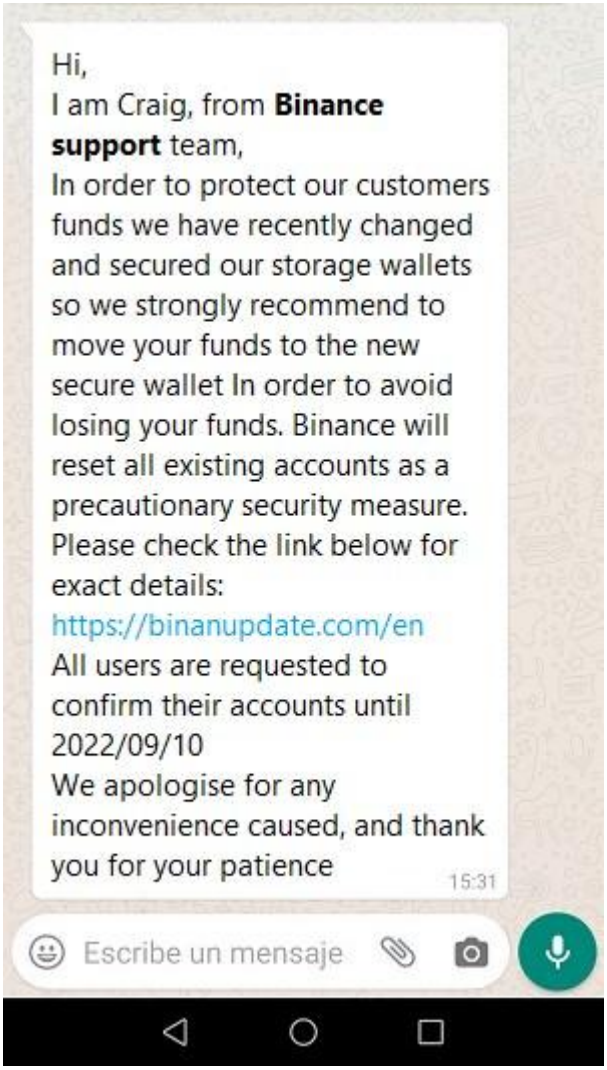


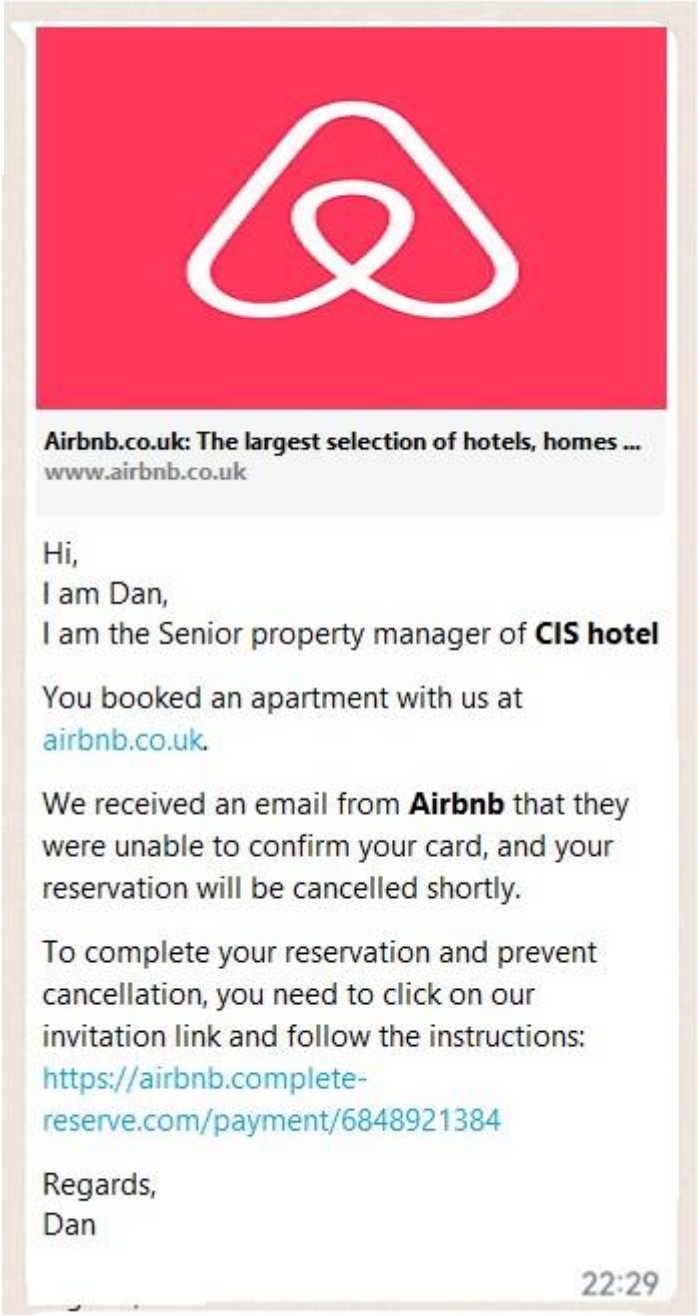
Fig 6. IKEA giveaway

Fig 7. Google Play gift cards

Fig 8. Coca Cola giveway

Fig 9. Amazon discount

Fig 10. Arbnb reservation email



APPENDIX C.2

Information and Consent Form

Dear Participant,

Thank you for your interest in participating in this study. Before you decide to participate in this study, it is important that you understand why the research is being done and what it will involve. Please read the following information carefully. You can also ask the researcher if anything is unclear or need further information. Once you have read the information and have decided to participate in the study, please click 'I agree to participate in this study' to start the questionnaire.

Information and Consent Form

You are invited to participate in research conducted by Rufai Ahmad, a PhD student at the Department of Computer and Information Sciences University of Strathclyde, Glasgow, UK. The study is about the behaviours of WhatsApp users towards information shared by others during instant messaging. Specifically, it aims to understand how WhatsApp users react to information shared by different types of people during instant messaging. Participation should take about 15 minutes.

Purpose of the Research

The study's overarching goal is to understand whether the type of communicating parties influences how users respond to messages during communication on WhatsApp. The survey hopes to provide insights about

users' attitudes towards information that are shared on such platforms and the factors that influence their decision to perform certain actions.

Participation

Your participation in this research study is voluntary. You may refuse to take part in the research or exit the survey at any time without penalty. This survey does not link your identity to your responses. Therefore, deleting your responses once they are submitted (at the end of the survey) will be impossible. To avoid this, you can withdraw from the survey by closing your browser (tab) at any point.

What will you do in the project?

This study requires you to respond to questions in an online survey. Specifically, you will be asked to answer demographic questions (age, gender, educational level and country of residence) and to respond to items of a construct that measures your habitual WhatsApp use. You would then be asked to provide a unique pseudonym for your friend, family, work/business colleague, and someone you know who isn't a family, friend, or work/business colleague. Thereafter, you will be presented with questions that are personalised using the pseudonyms you have provided. Each of these questions contain a screenshot of a WhatsApp message. You will be required to indicate how likely you are to respond to each of these messages.

Why have you been invited?

You have been chosen to participate in this study because you are 18 years or older and are a current user of WhatsApp Messenger.

Potential Harm or Risks

No physical, psychological or social risks may result from participating in this research project.

Data Collection Procedures

The survey consists of questions relating to your demographic. In addition, the survey also consists of questions relating to your behaviours towards messages shared by different communicating parties during instant messaging. The survey will be hosted and administered by Qualtrics.com.

Privacy and Confidentiality

The data collected in this survey will be kept confidential, and respondents will not be asked for any personally identifiable information. Responses will be fully anonymised and Qualtrics will be prevented from collecting any identifiable information including location information, IP addresses, etc. Participants will only be asked to provide their email address if they express willingness to participate in the draw to win one of five £20 Amazon vouchers. This draw is optional and was set-up to appreciate the participants for their time.

Responses from this survey will initially be stored on Qualtrics in a password-protected electronic format. Data will later be downloaded and stored digitally in a password-protected SPSS file on the principal's researcher's computer.

How will the data be processed?

The data collected from the study will be analysed using SPSS software for statistical analysis and R programming language. Specifically, we will employ different quantitative data analysis techniques to analyse the data. The study results will be presented in the form of numbers and statistics using tables, charts and figures. As the study does not collect personally identifiable data, none of the data labels will be anonymised, and the analysis will not attempt to identify any individual.

Conservation of Data

After completion of the research, data will be deposited into the University of Strathclyde institutional data repository, Pure, for a minimum period of 10 years. Data will also be stored on the principal's researcher's H and I: drive and Onedrive, a cloud service run by the University of Strathclyde with multiple security layers.

Potential Benefits

Participation and completion of all the questions in the survey will qualify you for a draw to win one of five £20 Amazon vouchers. You will be asked to provide your email address if you wish to be entered into this draw. Furthermore, your responses may help us learn more about users' behaviours towards different information shared on Whatsapp during instant messaging.

Results of the Research Project

You can always request a summary of the findings after the research has been completed. If you wish to receive a summary, please send an email to either the principal researcher, or Dr Sotirios Terzis whose emails are given below. In addition, the results of this study may be published in academic and professional journals and presented at conferences.

Researchers contact details:

Rufai Ahmad

PhD Student

Department of Computer and Information Sciences

Livingstone Tower

Richmond Street Glasgow G1 1XH

Email: rufai.ahmad@strath.ac.uk.

Dr Sotirios Terzis

Lecturer

Department of Computer and Information Sciences

Livingstone Tower Richmond Street Glasgow G1 1XH

Email: sotirios.terzis@strath.ac.uk

The University of Strathclyde Department of Computer Science Ethics Committee granted this research ethical approval. If you have any questions/concerns during or after the study or wish to contact an independent person to whom any questions may be directed or further information may be sought, please contact:

Secretary to the Departmental Ethics Committee

Department of Computer and Information Sciences,

Livingstone Tower Richmond Street Glasgow G1 1XH

email: ethics@cis.strath.ac.uk

ELECTRONIC CONSENT: Please select your choice below. You may print a copy of this consent form for your records. Clicking on the “Agree” button indicates that .

You have read the above information. You voluntarily agree to participate. You are 18 years of age or older.

- ○ I AGREE to participate in this study (1)
- ○ I DECLINE to participate in this study (5)

APPENDIX C.3

Experimental Vignette study Survey (EVM)

Do you currently use WhatsApp?

o Yes

o No

Group Please specify your age group

o 18-29

o 30-39

o 40-49

o 50–59

o 60 and above

Please specify your gender

o Male

o Female

o Prefer not to disclose

Please specify your level of education

o Did not finish school

o Secondary

o Further education (i.e. colleges)

o University undergraduate

o Postgraduate (e.g. Masters, Doctorate)

In which country do you currently reside?

▼ Afghanistan (1) ... Zimbabwe (1357)

Please indicate whether you agree or disagree with the following statements.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
Checking WhatsApp is part of my daily routine (1)	○	○	○	○	○
I feel my WhatsApp use has gotten out of control (2)	○	○	○	○	○
I would miss checking my WhatsApp if I could no longer do it (3)	○	○	○	○	○
I find myself checking my WhatsApp (updating my status or checking other people's status updates, replying to messages) around the same time each day. (4)	○	○	○	○	○

You will be asked to provide four pseudonyms relating to your friend, a member of your family, a work/business colleague and someone you know but not a friend, family or work/business colleague. A pseudonym is a fake name that a person or group uses for a special purpose. In this survey, the purpose of the pseudonym is to mimic real-world situations based on the assumption that you are likely to receive messages from those whose pseudonyms you have provided. **Please use a unique**

pseudonym for each of the following communicating parties. **For example, if you choose the name John as the pseudonym for your friend, then ensure you do not use the same name for any of the other acquaintances.**

Please enter a pseudonym relating to a friend

_____
_

Please enter a pseudonym relating to a family member

_____
_

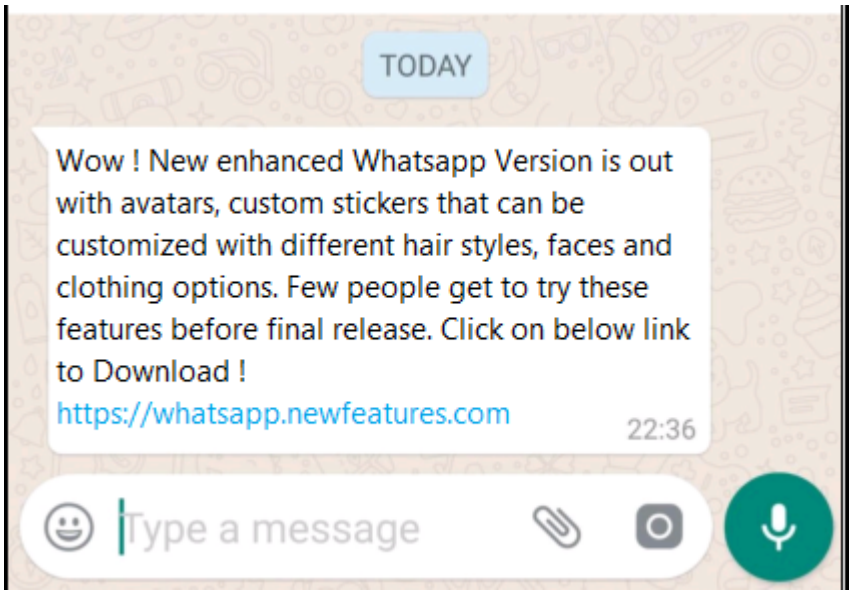
Please enter a pseudonym relating to your work/business colleague

_____
_

Please enter a pseudonym relating to someone you know but not friend, family or work/business colleague

_____
_

Imagine receiving the following message from your friend, ${Friend/ChoiceTextEntryValue} via WhatsApp



How likely are you to follow the link in this message?

- o Extremely unlikely
- o Somewhat unlikely
- o Neither likely nor unlikely
- o Somewhat likely
- o Extremely likely

How trustworthy do you consider this message?

- o not at all trustworthy
- o slightly trustworthy
- o somewhat trustworthy

o very trustworthy

o extremely trustworthy

Based on your previous response, select from the following, the reasons why you are likely or extremely likely to follow the link in the message. (Check all that apply)

▢ Because I know the sender

▢ Message referenced well-known organisation

▢ Other (Please Specify below)

Based on your previous response, select from the following, the reasons why you are extremely unlikely or somewhat unlikely to follow the link in the message. (Check all that apply)

▢ Because I know the sender

▢ Because of the absence of a logo

▢ Other (Please Specify below)

Imagine receiving the following message from your family relation, ${Family/ChoiceTextEntryValue} via WhatsApp.

How likely are you to follow the link in this message?

o Extremely unlikely

o Somewhat unlikely

o Neither likely nor unlikely

o Somewhat likely

o Extremely likely

How trustworthy do you consider this message?

o not at all trustworthy

o slightly trustworthy

o somewhat trustworthy

o very trustworthy

o extremely trustworthy

Based on your previous response, select from the following, the reasons why you are likely or extremely likely to follow the link in the message. (Check all that apply)

▢ Because I know the sender

▢ Message referenced a well-known organisation

▢ Because of the logo

▢ Other (Please specify below)

Based on your previous response, select from the following, the reasons why you are extremely unlikely or somewhat unlikely to follow the link in the message. (Check all that apply)

▢ Because I know the sender

▢ Other (please specify below)

Which of the following mobile instant messaging applications do you currently have an active account with and use? (Check all that apply)

▢ WhatsApp

▢ Line

▢ Viber

▢ Telegram

▢ Slack

▢ Signal

▢ ⊗None