Emerald Information Management and Computer Security

Getting Users to Click: A Content Analysis of Phishers' Tactics and Techniques in Mobile Instant Messaging Phishing

Journal:	*Information and Computer Security*
Manuscript ID	ICS-11-2023-0206.R1
Manuscript Type:	Original Article
Keywords:	Phishing, Persuasion Principles, Mobile Instant Messaging, URL obfuscation

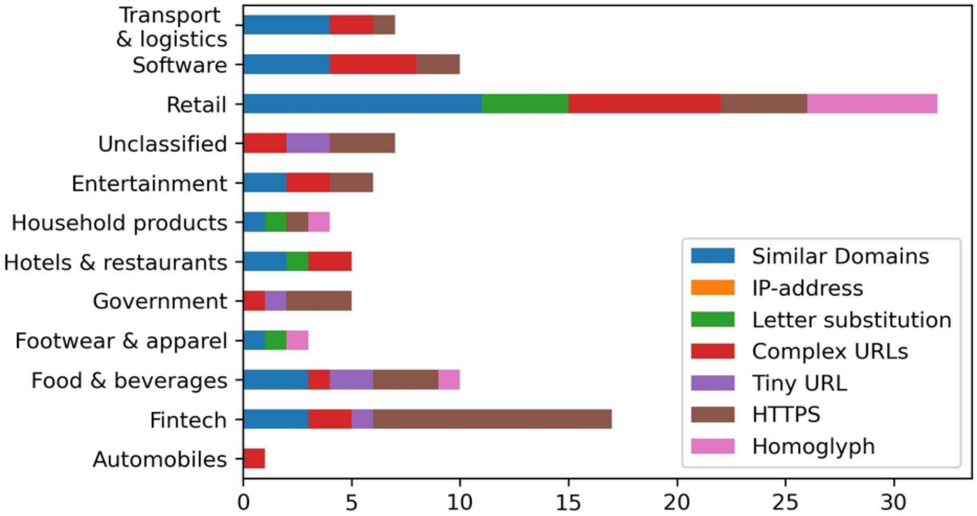

Figure 1: URL construction techniques across company categories (Due to space, the category Interactive media and Internet entertainment is shortened to "Entertainment")

161x87mm (150 x 150 DPI)

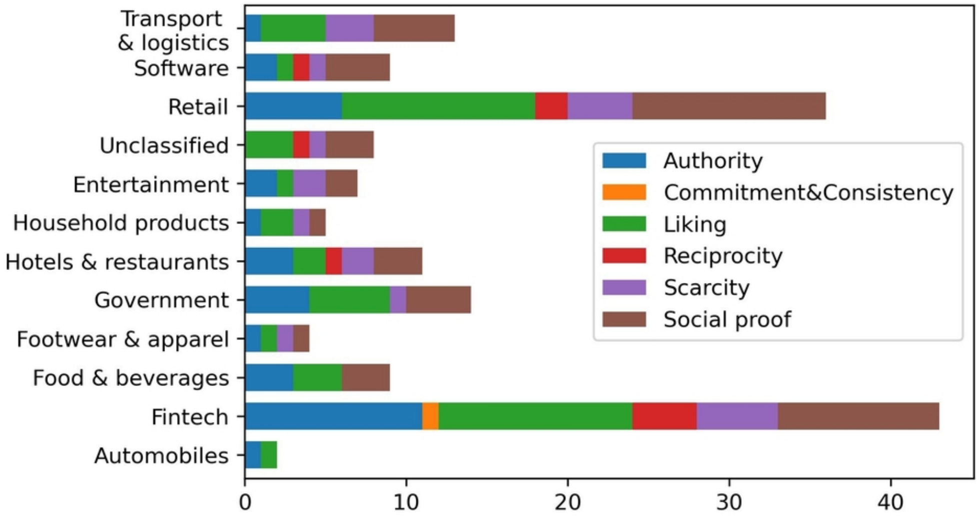

Figure 2. Persuasion principles across company categories (Due to space, the category Interactive media and Internet entertainment is shortened to "Entertainment")

161x87mm (150 x 150 DPI)

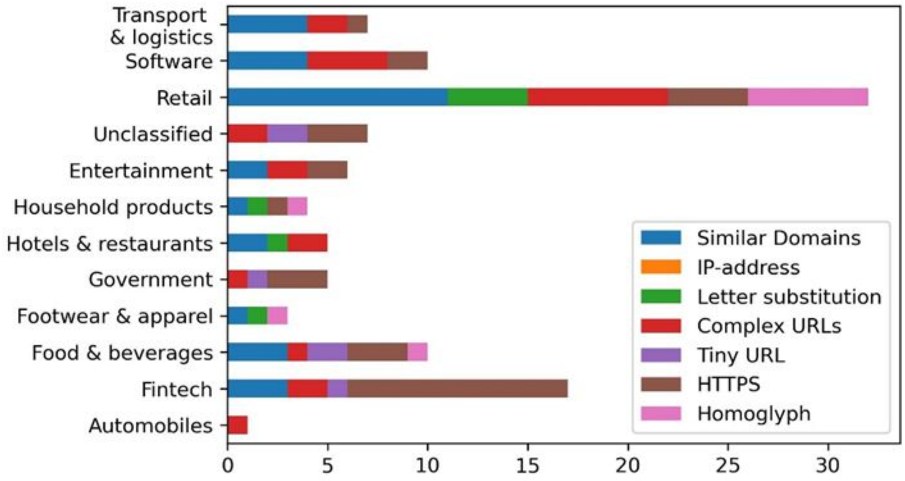

Figure 1*:* URL construction techniques across company categories (Due to space, the category Interactive media and Internet entertainment is shortened to "Entertainment")

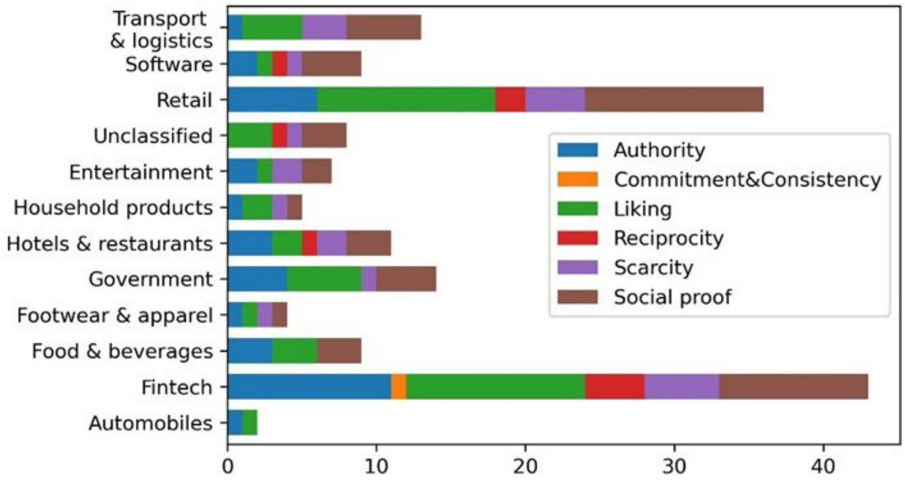

Figure 2. Persuasion principles across company categories (Due to space, the category Interactive media and Internet entertainment is shortened to "Entertainment")

Getting Users to Click: A Content Analysis of Phishers' Tactics and Techniques in Mobile Instant Messaging Phishing

Abstract

Purpose: This study investigates how phishers apply persuasion principles and construct deceptive URLs in mobile instant messaging (MIM) phishing.

Design/methodology/approach: In total, 67 examples of real-world MIM phishing attacks were collected from various online sources. Each example was coded using established guidelines from the literature to identify the persuasion principles and the URL construction techniques used.

Findings: The principles of social proof, liking, and authority were the most widely used in MIM phishing, followed by scarcity and reciprocity. Most phishing examples use three persuasion principles, often a combination of authority, liking, and social proof. In contrast to email phishing but similar to vishing, the social proof principle was the most commonly used in MIM phishing. Phishers implement the social proof principle in different ways, most commonly by claiming that other users have already acted (e.g. crafting messages that indicate the sender has already benefited from the scam). In contrast to email, retail and fintech companies are the most commonly targeted in MIM phishing. Furthermore, phishers created deceptive URLs using multiple URL obfuscation techniques, often using spoofed domains, making the URL complex by adding random characters and using homoglyphs.

Originality/value: The insights from this study provide a theoretical foundation for future research on the psychological aspects of phishing in MIM apps. The study provides recommendations that software developers should consider when developing automated anti-phishing solutions for MIM apps and proposes a set of MIM phishing awareness training tips.

Keywords: Phishing, Persuasion Principles, Mobile Instant Messaging, URL obfuscation

Paper Type: Research paper

1 Introduction

Phishing attempts to trick internet users into clicking malicious links or visiting fraudulent websites (NCSC, 2018). Phishers can also attempt to persuade targets to download malware or divulge sensitive information. Phishing commonly occurs via email, with phishers masquerading as legitimate entities. Increasingly, email is not the only phishing medium. Phishers can also use other media, such as Short Message Service (SMS), voice calling (vishing), and instant messaging. Recently, mobile instant messaging (MIM) applications (apps) such as WhatsApp, Telegram and Viber have been used by cybercriminals for phishing. This was noted in an article published by Kaspersky in 2021 [1]. Kaspersky internet security software for mobile devices detected 91,242 phishing links between December 2020 and May 2021. Most originated from WhatsApp (89.6%), followed by Telegram (5.6%) and Viber (4.7%). Some of these platforms (i.e. Snapchat and WhatsApp) claim to have measures in place to foil phishing [2][3]. Snapchat's anti-phishing functionality is based on Google Safe Browsing, a blacklisting service that can only detect known, reported phishing URLs and is ineffective in detecting zero-day phishing attacks (Sheng *et al.*, 2009). Due to its end-to-end encryption, WhatsApp states it performs automated phishing detection on users' devices [2]. However, it is unclear when WhatsApp and Snapchat introduced these measures since findings from (Stivala and Pellegrino, 2020) reveal that most MIM apps, including WhatsApp and Snapchat, lack or have ineffective automated anti-phishing measures. On the other hand, a number of educational interventions have been developed to educate users about phishing attacks. However, these are generally email-specific and generally neglect other vectors, such as MIM apps (Mossano *et al.*, 2020).

The literature has acknowledged that solving the problem of phishing requires both user awareness and technical countermeasures (Stojnic, Vatsalan and Arachchilage, 2021). Thus, the first step towards curbing the success of phishing in MIM apps is to understand the nature of phishing messages on these platforms by investigating the tactics and techniques that phishers use to defraud targets. Phishers employ several strategies to fool victims, including URL obfuscation techniques to create deceptive URLs (Lin *et al.*, 2011) and persuasion principles to create convincing messages (Hadnagy, 2018). The use of persuasion principles in both email phishing and vishing is well-documented in the literature, with findings showing that phishers use different persuasion principles depending on the attack vector (i.e., email or voice calling) (Ferreira, Coventry and Lenzini, 2015; Jones *et al.*, 2020). However, it has not been established whether these differences extend to MIM phishing (i.e., whether phishers use different persuasion principles and URL construction techniques in MIM apps).

The current research conducts a deductive content analysis of real-world MIM phishing attacks to examine what persuasion principles and URL construction techniques were applied. The study was based on 67 images of real-world mobile instant messaging phishing messages targeting the users of WhatsApp, Telegram and Viber. The images were collected from the Google search engine, similar to the approach in (Akdemir and Yenal, 2021). The images were analysed based on Cialdini's persuasion principles (Cialdini, 2006) and existing URL obfuscation methods from the literature.

This paper extends the work of (Ahmad, Terzis and Renaud, 2023) to include a detailed analysis of the industries most targeted and the URL construction techniques employed during MIM phishing. The results suggest that retail companies and fintech platforms were the most targeted, with the most messages about giveaways and fake lotteries. The social proof and liking principles were used the most, followed by authority and scarcity, with reciprocity, commitment, and consistency less frequently used. Most phishing messages combine three persuasion principles: authority, liking, and social proof. Phishers widely use HTTPS in MIM phishing. However, they combine HTTPS with similar/spoofed domains, complex URLs or homoglyphs to deceive users. These findings provide insights into how cybercriminals carry out phishing attacks on MIM apps.

The rest of the paper begins with a survey of previous research on the use of URL obfuscation techniques and persuasion principles in phishing in the background section. The methods section outlines the methodological approach for the study. The results section presents our findings, followed by a discussion and recommendations, together with an overview of the study's limitations. The final section concludes the paper and sets out avenues for future work.

2 Background

Phishers employ various strategies to persuade victims to click on malicious links. These links could compromise the victims' online accounts or lead to the victims downloading malware on their devices. As phishing aims to convince unsuspecting users to divulge their personal and financial information, individuals' social, psychological, and cognitive vulnerabilities play an essential role in its success. Often, cybercriminals exploit these weaknesses by deploying persuasion techniques during phishing. Persuasion techniques have their roots in marketing, where they are used to influence buyers to buy goods and services (Cialdini, 2006). Cialdini identified six persuasion principles that are very effective in marketing. Researchers have found these principles widely used in social engineering, phishing attacks and general scams (Uebelacker and Quiel, 2014; Jones *et al.*, 2020). Gragg identified seven psychological triggers used in social engineering attacks (Gragg, 2003), and more recently, Stajano and Wilson (Stajano and Wilson, 2011) identified seven principles of general scams. However, these principles share many similarities. For example, the diffusion of responsibility and moral duty by Gragg, the herd principle by Stajano and Wilson and the social proof by Cialdini all focus on the feeling of shared risk on the victim's part (Ferreira, Coventry and Lenzini, 2015).

Cialdini's six principles of persuasion are currently considered the most widely accepted in the literature (Butavicius *et al.*, 2016). The six principles are:

1) Authority: This principle states that people tend not to question those in authority because of the fear of losing privileges, condemnation, or humiliation.
2) Commitment and consistency state that individuals want to be consistent with their previous stand. They want to show an agreement between their words and their actions.
3) The principle of liking states that humans are more likely to comply with requests from people they like or with whom they share similarities.
4) The scarcity principle is based on the idea that when things become scarce, their value increases.
5) Reciprocity is based on humans' belief that they owe a debt to those who do kind things or give them something they enjoy.
6) The principle of social proof states that when humans fear making mistakes, they act according to most others.

Past studies have investigated how phishers apply Cialdini's six principles in phishing attacks. For example, (Akbar, 2014) investigated how these principles were used in phishing emails targeting individuals in the Netherlands by analysing 207 unique English phishing emails collected from a Netherlands database. Results from this study revealed that authority was the most used influence principle, followed by scarcity. In (Ferreira, Coventry and Lenzini, 2015), the authors investigated persuasion principles in phishing emails by analysing 52 emails written in English, collected from their mailboxes and the Internet. They found that the principle of liking was the most used, followed by scarcity and authority. In (Zielinska *et al.*, 2016), the authors analysed persuasion principles used in phishing emails over time. They collected 887 emails from 3 major US universities from January 2010 through June 2015. Their results revealed an increase in the use of principles of commitment/consistency and scarcity over time and a decrease in reciprocation and social proof. Authority and liking show both a decrease and an increase. More recently, Akdemir and Yenal (Akdemir and Yenal, 2021) collected and investigated the persuasion principles used in 208 coronavirus-themed phishing emails. The emails were collected as images from 3 search engines, Google, Bing and Yandex, between April 1 and April 16, 2020. Results from this study show that the principle of authority was the most used, followed by commitment/consistency and liking.

Phishing messages often contain URLs that direct users to fraudulent pages. Phishers construct these URLs using different techniques, although the ultimate goal remains to deceive victims into thinking they interact with legitimate sites. These techniques, often called URL obfuscation techniques, have been well-documented in the literature (Fernando and Arachchilage, 2020). The techniques were first categorised into four types (Garera *et al.*, 2007), namely: 1) Type I, where phishers use an IP address as the hostname; 2)Type II, where phishers use a domain name that looks similar to the imitated one but as a subdomain or in the path component of the URL, 3) Type III, where phishers extend the length of the URL with illogical characters that make reading the URL difficult for users also called complex URLs and, 4) Type IV, misspelling the genuine domain by substituting a letter from the actual domain with another one e.g., Paypa1 instead of Paypal.

In 2020, Fernando and Arachchilage (2020) extended Garera et al.'s study by analysing 10078 verified phishing URLs collected from Phishtank. The researchers found that phishers no longer use IP addresses in phishing URLs. However, they found wide use of HTTPS to reassure users and homoglyphs to spoof URLs. HTTPS is a cryptographic protocol used to secure connections on the web (Felt *et al.*, 2017). Homoglyphs are Unicode characters similar to ASCII characters. Fernando and Arachchilage (2020) recommended that current anti-phishing training methods should be updated to reflect the new strategies.

Research has not examined how phishers construct phishing URLs and how they apply persuasion principles during MIM phishing attacks. There may be differences in the modus operandi of phishers during MIM phishing compared to other vectors. MIM is often between individuals with some offline connection (O'Hara *et al.*, 2014), and researchers in (Ahmad and Terzis, 2022), have shown that MIM users are more likely to click on links when they come from those they know. Moreover, according to source credibility theory, people are more likely to accept and respond to a message when the source presents itself as reliable (Hovland, C.I., Janis, I.L. and Kelley, 1953). The source can be either the message's sender or the organisation being impersonated (Kim and Kim, 2013). Thus, in the context of MIM apps, phishers' chances of defrauding their victims can be increased by impersonating trusted organisations and crafting messages that exploit the trust between the sender and the receiver. This act could change which persuasion principles phishers use. Similarly, the small screen size of smartphones, already a barrier to users' ability to parse a URL correctly (Goel and Jain, 2018), could affect how phishers construct URLs during MIM phishing. Furthermore, most MIM apps use link previews to render links to users. The link preview gives users an idea of the pages they are visiting by providing information such as the logo of the organisation, the domain name and a short page description. However, findings from (Stivala and Pellegrino, 2020) have shown that phishers can manipulate the component of a link preview to direct users to fraudulent pages. Due to these factors and the differences between MIM and other communication media, this study seeks to provide insights into phishing in MIM apps by answering the following questions:

RQ1: What techniques do phishers use to construct URLs during MIM phishing?

RQ2: Does the prevalence of the techniques differ by the category of the company being impersonated?

RQ3: Which persuasion principles were used by phishers in MIM phishing, and how frequently were they used?

RQ4: How many persuasion principles do most phishing messages contain?

RQ5: How were the various persuasion principles implemented?

3 Methods

This study was conducted in three phases. The first phase entailed collecting images of real-world MIM phishing attacks. Following the approach used in (Jones *et al.*, 2020; Akdemir and Yenal, 2021), images of MIM phishing messages were collected with the Google search engine between August 8 and August 15, 2022. These were images of MIM phishing posted online by users and organisations. The terms "phishing messages", "phishing frauds", and "phishing scams" were searched in conjunction with either "WhatsApp", "Telegram", or "Viber". Specifically, the image search section of Google was used to access the actual images of the phishing messages. The study focused on WhatsApp, Telegram and Viber since their users are currently the most targeted [1]. A backdated search from 2018 was performed to cover a longer period.

Using the Google image search function, phishing images published within the previous year were searched using the advanced search operators "Before" and "After". For example, the search started with the query "phishing scams WhatsApp After: 2018-01-01 before 2019-12-01" to search for all images published between 2018-01-01 and 2019-12-01. This approach was used for subsequent years until August 15, 2022, when the data collection started. The returned images were scanned for every search result, selecting those with phishing content written in English. Some phishing messages were reported by many websites, leading to images appearing multiple times. Thus, the dataset was checked, and duplicates were removed. Furthermore, since this study focuses on phishing and not general social engineering scams, only messages with links were selected. The focus on messages with links was based on the current definition of phishing, where links or attachments are considered essential components of such attacks (NCSC, 2018). This process led to 67 images of MIM phishing messages, which the current study is based on.

The second stage focused on answering research questions RQ1 and RQ2, which aim to investigate the URL construction techniques used in the sample phishing messages. The sample was manually checked, and 62 out of the 67 messages were identified as having one or more of the URL construction techniques identified by (Garera *et al.*, 2007; Fernando and Arachchilage, 2020). The excluded messages contained links to an APK file (n=1), Google Docs (n=1), WhatsApp group invitation (n=2) and another message where the link was masked by the reporting website (n=1). RQ1 was answered by checking for the presence of any of the URL construction techniques in each phishing message.

Answering RQ2 required categorising the companies impersonated according to their sectors. The companies were classified using classifications from (Zielinska *et al.*, 2016) and the Global Industry Classification Standard (GICS) developed by MSCI and S&P Dow Jones in 1999 (O'Connell and Curry, 2022). The names of the companies referenced in the URLs or the content of the phishing messages were used to identify the categories the companies belong to.

The third phase involved identifying the persuasion principles in the collected phishing messages. In contrast to the second phase, all 67 example phishing messages were used. A deductive content analysis approach was used. A coding scheme was developed to determine the presence or absence of Cialdini's six persuasion principles in the messages. The scheme was developed using guidelines and examples from (Ferreira and Jakobsson, 2016; Jones *et al.*, 2020; Akdemir and Yenal, 2021). The scheme contained 21 statements, each designed to capture a particular principle of persuasion. Each instance was analysed and coded based on the persuasion principles used.

The first author and another coder who is a native English speaker and currently working within the Cybersecurity industry coded each of the 67 phishing examples. For each phishing example, the coders checked each statement in the codebook to see if it applied. If any statement assigned to a principle is deemed true, the principle associated with that statement is marked as present in the phishing example. The value 1 was used to indicate the presence of a principle, and 0 to indicate the absence.

Inter-coder reliability was assessed using Cohen's kappa. Reliability between coders was substantial (k =0.76), falling within recommended values (McHugh, 2012). One code set was randomly selected to serve as the final code for analysis.

Data analysis was conducted using Python programming language and Jupyter Notebook. Ethical approval was not needed for this study because it was based on publicly available data.

4 Results

Of the 67 phishing images analysed, we found that (29, 43.3%) included a logo as part of the link preview. Almost half of the messages contained a manipulated link preview (32, 47.8%).

Initial analysis also revealed that most messages (52, 77.6%) were sent via WhatsApp, followed by Telegram (14, 20.9%), and Viber (1, 1.5%).

Analysis of the most common words in the phishing samples shows that the five most common words used are 'free', 'get', 'click, 'given', and 'anniversary, as can be seen in Table I. Inspection of this table, suggests that these scams relate to giveaways and frequently referenced the term "click".

Table I: Most frequently used words in MIM phishing samples

Cybercriminals impersonate organisations and individuals in phishing scams. The companies impersonated in the sample messages were classified using the classification method outlined in Section 3. The most frequently spoofed organisations were those in the retail sector (n=15), followed by fintech (n=13). Table 2 shows the categories of companies most impersonated and their frequencies.

Table II: Categories of companies impersonated and their frequency.

How frequently were the various URL construction techniques used in MIM phishing?

To find which URL construction techniques were frequently used by phishers during MIM phishing, each URL in the dataset was analysed using the approach outlined in section 3. Analysis shows that the technique of using a similar domain was the most used (n=31), followed by using HTTPS (n=30) and making URLs complex/challenging to parse (n=24). The use of homoglyphs, which makes detecting spoofed domains challenging, was also seen in (n=9) phishing examples. Other techniques are tiny URL (n=6) and letter substitution (n=7). The technique of using an IP address in place of the hostname was not seen in any of the samples analysed.

Often, phishers combine many techniques to deceive their targets. This is the case in MIM phishing, with findings revealing eight unique sets containing more than one URL construction technique. Similar domains and complex URLs tend to be used together frequently (n=8), while complex URLs and HTTPS were used together in 7 messages. For example, the URL: http[://]play[.]google[.]store[.]apps[.]details[.]settings[.]pw/play?=1, which was imitating Google Play used both complex URL and similar domain. Table III shows other URL construction techniques that were seen used together.

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

Does the prevalence of the URL construction techniques differ by the category of the company being impersonated?

Analysis of the prevalence of the different URL construction techniques across each company category shows that phishers frequently use similar domains when impersonating organisations in the retail category. The use of HTTPS to deceive users was prevalent in scams targeting Fintech companies. The highest usage of homoglyphs was also seen in the retail category. Figure 1 presents the prevalence of these techniques across other categories of organisations.

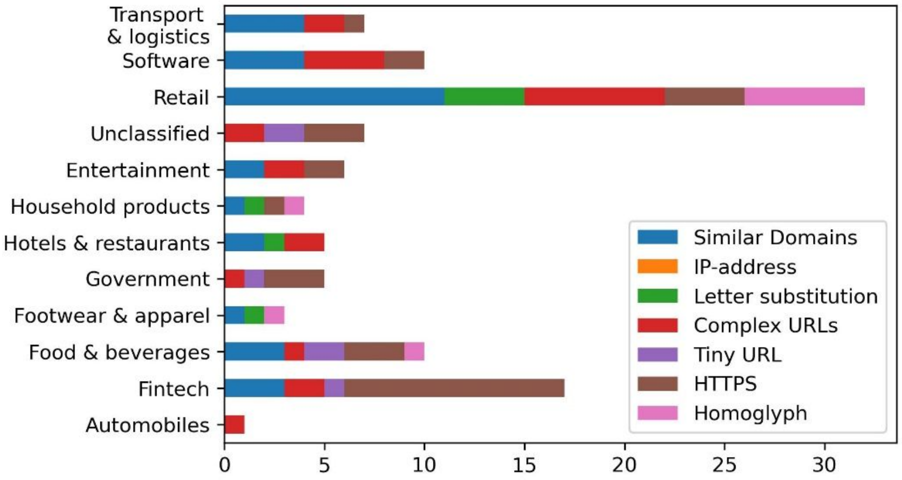

Figure 1: URL construction techniques across company categories (Due to space, the category Interactive media and Internet entertainment is shortened to "Entertainment")

Which persuasion principles were used by phishers in MIM phishing, and how frequently were they used?

The principle of social proof was the most used in the sample (n=51, 76.1%), followed by liking (n=50, 74.6%), authority (n=37, 55.2%), scarcity (n=23, 34.3%), reciprocity (n=11, 16.4%) and commitment and consistency (n=1, 1.4%).

When analysed across each category of impersonated organisations, the results suggest that the principles of authority, liking, scarcity, and social proof were used in most organisation types (See Figure 2). However, the results also indicate that phishers widely apply the liking and social proof principles in phishing scams targeting the retail sector. Figure 2 also suggests that phishers use the principle of reciprocity in scams targeting fintech, hotels and restaurants, retail, and software companies.

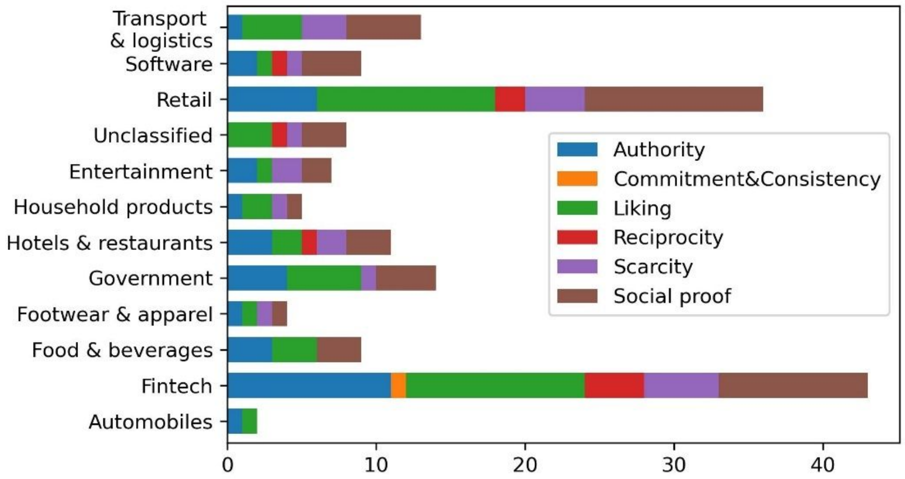

Figure 2. Persuasion principles across company categories (Due to space, the category Interactive media and Internet entertainment is shortened to "Entertainment")

How many persuasion principles do most phishing messages contain?

The dataset was converted into a Pandas DataFrame, where every row represents a phishing message. A Python script was written to check the number of times the value 1 appears in every row of the Pandas DataFrame whilst saving this value in a list. This approach allowed counting the number of persuasion principles in each message. Using the Python library Scipy, the most frequent score in the list was calculated.

Analysis shows that most of the phishing messages (n=21) contained three persuasion principles, followed by those having two (n=16). Ten messages contained four persuasion principles. In five messages, phishers were seen using five persuasion principles.

Further analysis of each of the combinations shows that most of the messages using two principles contained the liking and social proof principles (n=8), followed by authority and liking (n=6), and authority and social proof (n=2). Most messages using three persuasion principles tend to contain authority, liking, and social proof (n=8), followed by liking, scarcity, and social proof (n=7). Other combinations are liking, reciprocity and social proof (n=4) and authority, scarcity and social proof (n=2). Phishing messages using four persuasion principles tend to contain authority, liking, scarcity and social proof (n=8). At the same time, (n=2) messages were based on authority, liking, reciprocity and social proof. Messages with the highest number of deception techniques contained five principles. These messages were cryptocurrency scams distributed via the Telegram platform. Specifically, it was observed that phishers simultaneously applied the principles of authority, liking, reciprocity, scarcity and social proof in four example messages. Phishers applied the principles of authority, commitment and consistency, liking, scarcity and social proof in one phishing example.

Some messages (n=13) used only one persuasion principle to entice their targets. Most of these messages were based on social proof (n=5), followed by authority (n=4), liking (n=2), scarcity (n=1), and reciprocity (n=1).

How were the various persuasion principles implemented?

Analysis shows that phishers enforce authority by using logos and names of well-known organisations, social proof by referencing famous people and stressing the benefit of visiting the fraudulent page. For example, in one cryptocurrency scam, phishers referenced the CEO of Coinbase. In some cases, phishers apply the principle of social proof by portraying to the recipient of the message that the sender has already benefited from complying and, therefore, the receiver should comply by visiting the fraudulent page. For example, most giveaways and voucher scams contained the sentence "I have received mine". This approach tricks the receiver into believing the sender already benefits from complying.

Phishers apply the principle of liking by appearing to be helping the recipient. Scarcity was demonstrated by stressing that the recipient of the phishing message must act immediately or risk losing out on offers. Phishers enforce scarcity by including phrases such as "free air tickets to 200 customers". Messages with the reciprocity principle ask users to act in response to a favour. These messages were mainly related to cryptocurrency scams. For example, phishers told users they had won a certain amount of BTC and could only claim it after sending a specific amount of Bitcoin to a cryptocurrency wallet.

5 Discussion and Recommendations

5.1 Overview of findings

This study shows how phishers use MIM apps to propagate phishing campaigns. Most campaigns were conducted via WhatsApp, agreeing with the findings from [1]. The campaigns relate to giveaways and contain known phishing terms such as free', 'get', 'click, similar to what was found by (Stojnic, Vatsalan and Arachchilage, 2021). The results revealed that phishers frequently combine domain spoofing, HTTPS and complex URLs to create deceptive URLs during MIM phishing. While this is a well-known phisher technique (Bijmans *et al.*, 2021), applying it to the smartphone context suggests that phishers are aware of the hardware limitations of smartphones (i.e. small screen size) and are likely exploiting that limitation to make correct parsing of URLs difficult for users. HTTPS was widely used to create deceptive URLs during MIM phishing. Users have been encouraged to look for HTTPS when sharing sensitive information on the web (Roberts *et al.*, 2019). However, previous research has shown that users misunderstood the true meaning of HTTPS by assuming it indicates the trustworthiness of a website (von Zezschwitz, Chen and Stark, 2022). Interestingly, the URLs of most scams impersonating fintech companies had HTTPS. A recent survey of fintech users revealed that many have higher educational qualifications (Hu *et al.*, 2019). Thus, fintech users may likely be tech-savvy and, therefore, can be suspicious of URLs lacking HTTPS. Perhaps this is the reason for using HTTPS in fintech phishing scams.

Findings from this study reveal the prevalent use of persuasion principles in MIM phishing. The results suggest that social proof, liking, and authority were the most widely used, followed by scarcity, reciprocity, commitment and consistency. Furthermore, most phishing messages have three persuasion principles, often a combination of authority, liking, and social proof. In most cases, phishers enforce authority by using logos and names of well-known organisations. Using logos and names of well-known organisations enforces source credibility and convinces the recipient to respond (Kim and Kim, 2013), aligning with the source credibility theory (Hovland, C.I., Janis, I.L. and Kelley, 1953). However, in MIM apps, the source can be either the user sending the message or the organisation being impersonated. When phishing users of MIM apps, cybercriminals often ask users to forward the message to others (Cuddeford, 2018). This approach adds another layer to the source's credibility by allowing phishers to exploit the trust between the sender and the receiver. Another observation is that most giveaway campaigns contain the term "I have received mine" to deceive users into thinking that those sending the message have acted on it. This approach can be very effective when the receiver trusts the sending party. Phishers have persuaded many users to respond to MIM phishing messages using this approach (ActionFraud, 2018; iRadio, 2018). Furthermore, this approach is in line with the social proof principle, which is based on the concept of shared risk (i.e. by informing the target that other users have acted in this manner in the past, so it is okay for them to do the same). While the social proof principle is widely used in MIM phishing, the literature on email phishing revealed that it is not widely used (Ferreira, Coventry and Lenzini, 2015). This suggests a potential difference between the two vectors. However, the results of this study are similar to what was found in vishing scams (Jones *et al.*, 2020) in that phishers rely heavily on stressing the benefit of complying, suggesting the reliance on the social proof principle.

Findings from the studies by (O’Hara *et al.*, 2014; Ahmad and Terzis, 2022) revealed that communication in MIM apps tends to be informal and often between those that share some interest (i.e. friends or those they share similarities with, such as MIM app group members etc.). This makes it easier for phishers to implement the liking principle since messages will likely come from those the recipient likes. Once phishers deceive a user or compromise an account, they can move laterally to other users by accessing the compromised user’s contacts. It is likely the reason why most of the messages analysed in this study contained the liking principle.

The results also suggest that phishers used the principles of reciprocity, commitment and consistency less frequently. This is similar to the literature on vishing, where phishers used the principles of reciprocity, commitment and consistency less (Jones *et al.*, 2020). In contrast, the authors in (Ferreira, Coventry and Lenzini, 2015) found reciprocity, commitment and consistency widely used in emails offering large sums of money. The authors found that emails using these principles tend to be personalised and informal and often ask for a favour from the target. It may be challenging for phishers to implement these principles in MIM phishing. For example, it may be difficult for phishers to demonstrate an individual's previous commitment or consistency with an earlier situation in MIM phishing, unless the attack is targeted and trust has already been established between the phisher and the target. Similarly, reciprocity may be easier to implement in a targeted attack that requires the phisher to establish some trust with the target.

In terms of organisations most targeted, the results suggest that retail companies and fintech platforms were the most targeted, contradicting earlier studies on email phishing, where education and health institutions were impersonated the most (Zielinska *et al.*, 2016; Akdemir and Yenal, 2021). However, this is likely because retail and fintech organisations use MIM apps to communicate with customers. For example, the Telegram platform is considered the de facto platform for cryptocurrency (Smuts, 2019). Similarly, the literature has shown that retail companies frequently use MIM apps to communicate with customers (Vasiliu *et al.*, 2023).

Finally, the findings of this study revealed that most phishing examples include manipulated link previews. A manipulated link preview can deceive users into thinking they are visiting the right page. The authors in (Stivala and Pellegrino, 2020) have raised concerns about the potential of using link previews to deceive users into visiting fraudulent pages. This study confirms the exploitation of link previews to direct users to fraudulent pages.

In the next section, we will provide a set of recommendations for tackling the problem of phishing in MIM apps.

5.2 Recommendations

Automatic detection: Currently, most MIM apps do not have automated technical countermeasures for detecting phishing. Although some platforms, such as Snapchat, use phishing blacklists, this approach is ineffective. WhatsApp mentions performing on-device checks but doesn't explain what these checks entail. Notwithstanding, current automated phishing detection solutions will likely work in MIM apps. However, software developers must consider the end-to-end communication encryption of these apps. The SafetyNet Safe Browsing API for Android powered by Google Play Service is an example of a privacy-preserving approach for phishing URL detection [4]. This API can be useful for software developers of instant messaging applications. This approach preserves users' privacy by performing a hash-based lookup against a locally stored list of unsafe URL hashes, updated every 30 minutes. One drawback of this API is that it is based on a phishing blacklist. Therefore, its effectiveness depends on its update speed and frequency. Effective detection of MIM phishing may require more advanced techniques that use artificial intelligence, URL features and message content. The current URL and content-based solutions are likely to work. However, they need to be used in a privacy-preserving manner. When building automated phishing detection tools and security crawlers for MIM apps, we recommend using the keywords in Table I to refine the detection logic of these tools. Software developers should also pay attention to complex URLs as they are widely used in MIM phishing.

From the perspective of persuasion principles, the findings of this study show that phishers frequently use the social proof principle when conducting phishing campaigns in MIM apps. Incorporating persuasion cues in phishing detection models has improved accuracy (Valecha, Mandaokar and Rao, 2021). Therefore, we recommend using relevant social proof cues when building content-based phishing detection tools for MIM apps by integrating keywords that indicate shared risks. Furthermore, we observed that most of the giveaway phishing messages tend to contain the names of celebrities. Thus, this can be an area of focus for software developers. They should prioritise giveaway messages that contain the names of celebrities.

User Awareness: Recent findings have shown that current anti-phishing training materials focus on email (Mossano *et al.*, 2020), neglecting other vectors such as MIM apps. The problem of MIM phishing does not call for developing new anti-phishing training materials but the need to update existing training materials to inform users of the threat and its nature. As seen in (ActionFraud, 2018; iRadio, 2018), low awareness of MIM app users contributes to their susceptibility. In Table IV, we propose a list of facts and advice which we believe can be adopted as a foundation for basic training on phishing in MIM apps. The lessons are similar to the ones for Quishing (Sharevski et al., 2022) and [1].

Table IV. Tips and advice for mitigating mobile instant messaging phishing

5.3 Limitations of this study

This study has some limitations. First, the sample phishing messages were taken from public websites via the Google search engine. As such, we may not have captured all the phishing emails circulated within the selected period. For instance, phishing messages not reported or posted online were not captured. However, as phishing in MIM apps is a new trend, there are no verified databases of such scams, leaving us with the current approach as our only option. It is hoped that, like emails, a dedicated MIM phishing corpus will be created by researchers to address this problem. We also acknowledge the relatively small sample size used in this study, which we attribute to the emerging nature of MIM phishing. However, our sample falls within the range used in the literature (Jones et al., 2020). Finally, deductive content analysis, although promising, is believed to sometimes impact the ability of researchers to recognise other contextual features that may relate to the phenomenon being investigated (Hsieh and Shannon, 2005). For example, it is possible that other features of MIM phishing were not identified because the researchers relied on existing categories from the literature.

6 Conclusion and future work

This study provides insights into the nature of phishing in MIM apps by investigating how phishers create deceptive URLs and apply Cialdini's persuasion principles during MIM phishing. The results suggest widespread use of these principles during MIM phishing and the adoption of known phisher URL construction techniques of combining domain spoofing, HTTPS and complex URLs to create deceptive URLs. The results suggest that source credibility in MIM phishing is enforced by impersonating well-known organisations and getting unsuspecting users to forward messages to their contacts. Using two dimensions (i.e., the organisation impersonated and the sender) to enforce source credibility is a novel aspect of phishing in MIM apps. This study also revealed that phishers craft messages with terms that portray the concept of shared risk to assure targets that those sending the message have acted on it.

In future work, we intend to collect more sample phishing messages, perform a more in-depth analysis of these attacks, and create a taxonomy of phishing attacks in MIM phishing. Furthermore, in the next phase of this research, we will investigate the effectiveness of the URL construction techniques and persuasion principles found in the study by experimenting with users to determine their susceptibility.

Notes

1. https://www.kaspersky.com/about/press-releases/2021_phishing-in-messenger-apps-whats-new (04 January 2022).
2. https://faq.whatsapp.com/393169153028916/?cms_platform=android&helpref=platform_switcher (Accessed: 27 May 2023).
3. https://help.snapchat.com/hc/en-us/articles/7012345182356-How-Snapchat-Uses-Google-Safe-Browsing (Accessed: 27 May 2023).
4. https://developer.android.com/privacy-and-security/safetynet/safebrowsing (Accessed: 04 November 2023).

References

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).

Ahmad, R. and 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. and 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.

Akbar, N. (2014) 'Analysing persuasion principles in phishing emails'. University of Twente. Available at: http://essay.utwente.nl/66177/ (Accessed 04 November 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), https://doi.org/10.1177/21582440211031879

Bijmans, H. *et al.* (2021) 'Catching phishers by their bait: Investigating the dutch phishing landscape through phishing kit detection', in *30th USENIX Security Symposium (USENIX Security 21)*, pp. 3757–3774.

Butavicius, M. *et al.* (2015) 'Breaching the human firewall: Social engineering in phishing and spear-phishing emails', *in*: *ACIS 2015 Proceedings*. 98. https://aisel.aisnet.org/acis2015/98

Cialdini, R. B. (2006) 'Influence: the psychology of persuasion, revised edition', *New York: William Morrow*.

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).

Felt, A. P. *et al.* (2017) 'Measuring {HTTPS} adoption on the web', in *26th USENIX security symposium (USENIX security 17)*, pp. 1323–1338.

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?', *in: Australasian Conference on Information Systems, Perth, 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', in Tryfonas, T., Askoxylakis, I. (eds) *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. Proceedings 3*. Springer, pp. 36–47. https://doi.org/10.1007/978-3-319-20376-8_4

Ferreira, A. and Jakobsson, M. (2016) 'Persuasion in scams', in: Jakobsson, M. (eds) *Understanding Social Engineering Based Scams*. Springer, New York, NY. https://doi.org/10.1007/978-1-4939-6457-4_4 .

Garera, S. *et al.* (2007) 'A framework for detection and measurement of phishing attacks', in *Proceedings of the 2007 ACM workshop on Recurring malcode*, pp. 1–8. https://doi.org/10.1145/1314389.1314391

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.

Gragg, D. (2003) 'A multi-level defense against social engineering', *SANS Reading Room*, 13.

Hadnagy, C. (2018) *Social Engineering: The Science of Human Hacking*. Second Edi. Canada: Jhon Wiley & Sons Inc.

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.

Hu, Z. *et al.* (2019) 'Adoption intention of fintech services for bank users: An empirical examination with an extended technology acceptance model', *Symmetry*, 11(3), p. 340. https://doi.org/10.3390/sym11030340

iRadio (2018) *Tyto Park Scam*. Available at: https://m.facebook.com/thisisiradio/posts/1927459280648472/?comment_id=1927565900637810 (Accessed: 30 May 2023).

Jones, K. S. *et al.* (2020) 'How social engineers use persuasion principles during vishing attacks', *Information & Computer Security*. Vol. 29 No. 2, pp. 314-331. https://doi.org/10.1108/ICS-07-2020-0113 doi: 10.1108/ICS-07-2020-0113.

Kim, D. and Kim, J. H. (2013) 'Understanding persuasive elements in phishing e-mails: A categorical content and semantic network analysis', *Online Information Review*. Vol. 37 No. 6, pp. 835-850. https://doi.org/10.1108/OIR-03-2012-0037

Lin, E. *et al.* (2011) 'Does domain highlighting help people identify phishing sites?', in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 2075–2084, doi: 10.1145/1978942.1979244.

McHugh, M. L. (2012) 'Interrater reliability: the kappa statistic', *Biochemia medica*, 22(3), pp. 276–282.

Mossano, M. *et al.* (2020) 'Analysis of publicly available anti-phishing webpages: contradicting information, lack of concrete advice and very narrow attack vector', in *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, pp. 130–139. DOI: 10.1109/EuroSPW51379.2020.00026

NCSC (2018) *Phishing attacks: defending your organisation*. Available at: https://www.ncsc.gov.uk/guidance/phishing (Accessed: 25 January 2021).

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).

O'Hara, K. *et al.* (2014) 'Everyday dwelling with WhatsApp', in *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. pp. 1131–1143 doi: 10.1145/2531602.2531679.

Roberts, R. *et al.* (2019) 'You are who you appear to be: A longitudinal study of domain impersonation in tls certificates', in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pp. 2489–2504. https://doi.org/10.1145/3319535.3363188

Sheng, S. *et al.* (2009) 'An empirical analysis of phishing blacklists', in *6th Conference on Email and Anti-Spam, CEAS 2009*.

Sharevski, F. et al. (2022) 'Phishing with malicious QR codes', in 2*022 European Symposium on Usable Security*, pp. 160–171.: https://doi.org/10.1145/3549015.3554172

Smuts, N. (2019) 'What drives cryptocurrency prices? An investigation of Google trends and telegram sentiment', *ACM SIGMETRICS Performance Evaluation Review*, 46(3), pp. 131–134. https://doi.org/10.1145/3308897.3308955

Stajano, F. and Wilson, P. (2011) 'Understanding scam victims: Seven principles for systems security', *Communications of the ACM*. doi: 10.1145/1897852.1897872.

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. https://doi.org/10.1002/spy2.165

Uebelacker, S. and Quiel, S. (2014) 'The social engineering personality framework', in *Proceedings - 4th Workshop on Socio-Technical Aspects in Security and Trust, STAST 2014 - Co-located with 27th IEEE Computer Security Foundations Symposium, CSF 2014 in the Vienna Summer of Logic 2014*. doi: 10.1109/STAST.2014.12.

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.

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.

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

Zielinska, O. A. *et al.* (2016) 'A temporal analysis of persuasion principles in phishing emails', in *Proceedings of the human factors and ergonomics society annual meeting*. SAGE Publications Sage CA: Los Angeles, CA, pp. 765–769. https://doi.org/10.1177/15419312136011

Table I: Most frequently used words in MIM phishing samples

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

Table II: Categories of companies impersonated and their frequency

Company Category	Frequency
Retail	15
Fintech	13
Food and beverages	5
Government	5
Software	5
Transportation and logistics	5
Hotels, restaurants and leisure	3
Household products	2
Interactive media and Internet entertainment	2
Automobiles	1
Footwear and apparel	1

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

URL obfuscations sets	Frequency
Similar domains and complex URLs	8
Complex URLs and HTTPS	7
Similar Domains, Letter substitution and homoglyph	6
Similar Domains and HTTPS	5
Tiny URLs and HTTPS	5
Similar Domains Complex URLs and HTTPS	4
Similar Domains and homoglyph	3
Similar DomainsLetter substitution	1

Table IV: Tips and advice for mitigating mobile instant messaging phishing

• Scammers often use messenger apps such as WhatsApp, Telegram and Viber to share malicious messages. These messages can be shared by your contacts who have fallen for such scams. They can also be shared in messenger apps groups by members of those groups.
• Scammers often ask users to share malicious links with other users as part of an approach to abuse the trust users have in their contacts. Always verify the legitimacy of a message before responding by contacting the relevant authorities or organisation.
• Suspicious messages in messenger apps can appear legitimate. However, you should always check the links as they will most likely have an incorrect spelling or redirect to a different website.
• Shared links in messenger apps can contain a link preview. The link preview shows the logo of the organisation, the domain name and a short page description. Scammers can manipulate the link preview, thus never take it as a sign of legitimacy.
• Remember, scammers can forward suspicious messages to you from hacked accounts., so remain cautious.
• If you receive a suspicious message from an unknown number, delete it, report and block the sender.
• If you receive a suspicious message from a known number, delete it and contact the user to inform them.