



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

# **Hunting Phish: An Exploration of the Human-Based Detection and Management of Phishing Attacks**

*Kholoud Althobaiti*



Doctor of Philosophy  
Institute for Language, Cognition and Computation  
School of Informatics  
The University of Edinburgh  
2021



# Abstract

Phishing communications aimed at deceiving people pose a severe threat for organisations, necessitating the need to focus on preventing potential victims from falling for phishing, as well as the formulation of policies and the development of solutions to enable quick responses to ongoing attacks. With the above in mind, this thesis aims to explore phishing features in human-facing interventions as well as the organisational response to phishing attacks.

I started with an exploration of the phishing features related to URLs since they are one of the most robust features of phishing communication. To this end, I conducted a structured review of URL-based phishing features that appear in publications targeting human-facing and automated anti-phishing approaches to obtain a more comprehensive feature list and create a cross-community foundation for future research. I find that research on automation has utilised most of the features, but features were minimally explored in the human-facing anti-phishing research. Features that are rarely used in human-facing phishing work are still be utilised by experts, suggesting that average users could potentially use them too if they were presented in a usable way. Thus, I designed a usable URL feature report that aims to make experts' information sources accessible to non-experts to help general users judge URLs accurately. This report was designed iteratively with experts and average users before being evaluated in an online study. I show that the report supports users in accurately judging URLs' safety.

In order to explore the organisational response to phishing attacks, I conducted a case study to investigate the processes of handling phishing reports, teams' interactions to improve defences, and the hindrance to a fast and effective response. The observed work patterns are a distributed cognitive process requiring multiple distinct teams with narrow system access and specialised knowledge. Sudden large campaigns can overwhelm the Help Desk with reports, significantly impacting staff's workflow and hindering the effective application of mitigations and the potential for learning.

The results from the several studies conducted throughout this thesis highlight the need for users' awareness; such awareness would aid them in avoiding clicking phishing URLs and would also help organisations to manage the impact. Indeed, the majority of the existing research on phishing is directed towards the goal of improving proactive measures rather than reactive measures; however, it is necessary to focus on strengthening every element in the phishing life cycle. My work shows that there are still many opportunities to add tool-based support into the process, both at the end-user level and in support of organisational IT staff.

# Acknowledgements

This thesis is dedicated to my children. Loving you made me stronger and more independent.

I am grateful for the great deal of support that I have received throughout my PhD journey. In this regard, I would first like to thank my principal supervisor, Dr Kami Vaniea, who has given me ample time, effort, and opportunity to acquire the knowledge needed to complete this PhD. Your insightful feedback has pushed me to sharpen my thinking and has brought my work to a higher level. I also want to thank my second supervisor, Dr Maria Wolters, for her valuable guidance throughout my studies. Further, I want to acknowledge my friends in the TULiPS lab, who always provide stimulating discussions and feedback on my methodologies and writings. I would particularly like to single out my co-authors for their support with the sections of my thesis that have been published. A special thank you to my best friend Hend who was always willing to assist with the consistency of the prototypes and slides designs.

Many thanks go to the IS teams, who helped me with the data collection. A special thank you to the Help Desk manager for facilitating all the meetings and resources for the data collection.

I want to praise and thank Allah, the almighty, who has helped me accomplish this thesis. In addition, I would like to thank my husband, Fahad, for supporting me throughout my studies and providing happy distractions that allowed me to rest my mind outside of my research. You are always there for me, and without you, I could not have completed this thesis successfully. I also want to thank my parents for their wise counsel, emotional support, and daily prayers. To my brothers and sisters: sharing my achievements with you feels particularly special. You always support me in my achievements and wish me the best. Again, a special thanks to my children Rama, Turkey, and Riyad: you are always my inspiration to succeed and the reason why I keep going when I feel like giving up. Finally, I would like to thank my friends for their sympathetic ear without judging. With you, I have shared thoughts, laughter, pride, and frustration.

# Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

*(Kholoud Althobaiti)*



# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Overview of the Thesis Research . . . . .	5
1.1.1	Human-based phishing detection . . . . .	5
1.1.2	Organisational handling of phishing . . . . .	7
1.2	Research Contributions . . . . .	8
1.3	Thesis Outline . . . . .	8
<b>2</b>	<b>Background and Related Work</b>	<b>11</b>
2.1	What is Phishing? . . . . .	11
2.2	Phishing Ecosystem . . . . .	12
2.2.1	Planning and executing the attack . . . . .	12
2.2.2	Proactive defences: Where an attack might fail . . . . .	14
2.2.3	When defences fail . . . . .	17
2.2.4	Reactive measures: Opportunities to mitigate the attack . . . . .	18
2.3	Conclusion . . . . .	19
<b>3</b>	<b>A Review of Human- and Computer-Facing URL Phishing Features</b>	<b>21</b>
3.1	Introduction . . . . .	21
3.2	Uniform Resource Locator (URL) . . . . .	23
3.3	Methodology . . . . .	24
3.3.1	Procedure . . . . .	24
3.3.2	Inclusion/exclusion criteria . . . . .	25
3.3.3	Limitations . . . . .	25
3.4	Phishing Features . . . . .	27
3.4.1	URL lexical features . . . . .	27
3.4.2	Host features . . . . .	33
3.4.3	Rank-based features . . . . .	35



3.4.4	Search engine features . . . . .	36
3.4.5	Redirection-based features . . . . .	37
3.4.6	Certificate-based features . . . . .	38
3.4.7	Black/white list features . . . . .	39
3.5	Discussion . . . . .	40
3.5.1	Shifting effectiveness of features . . . . .	40
3.5.2	Balancing ‘safe’ and ‘phish’ data sets . . . . .	41
3.5.3	Host-obscuring tactics . . . . .	41
3.5.4	Exploring human-facing features . . . . .	41
3.6	Conclusion . . . . .	42
<b>4</b>	<b>Making URL Phishing Features Human Comprehensible</b>	<b>43</b>
4.1	Introduction . . . . .	43
4.2	Deciding if a URL Goes Where the User Thinks it Goes . . . . .	46
4.2.1	Mouse over the link and look at the URL . . . . .	46
4.2.2	What is the ‘correct’ URL anyway? . . . . .	47
4.2.3	Redirects and short URLs . . . . .	48
4.3	Design Goals . . . . .	49
4.4	Designing the Initial Report . . . . .	50
4.4.1	Notice and reminder . . . . .	50
4.4.2	Facts . . . . .	51
4.4.3	Tricks . . . . .	52
4.4.4	Severity colours . . . . .	53
4.5	Iterating with Focus Groups . . . . .	53
4.5.1	Participants . . . . .	55
4.5.2	Procedure . . . . .	55
4.5.3	Outcomes . . . . .	56
4.5.4	Expert interview . . . . .	64
4.6	Features Validation . . . . .	65
4.7	Online Study . . . . .	66
4.7.1	Questionnaire instrument . . . . .	66
4.7.2	Survey results . . . . .	70
4.8	Limitations . . . . .	73
4.9	Discussion . . . . .	74
4.9.1	User empowerment . . . . .	75

4.9.2	Training . . . . .	75
4.9.3	Use cases . . . . .	76
4.9.4	Focused attention . . . . .	76
4.9.5	Deployment potential . . . . .	77
4.10	Conclusion . . . . .	77
<b>5</b>	<b>A Case Study of Phishing Incident Response in a University</b>	<b>79</b>
5.1	Introduction . . . . .	79
5.2	Related Work . . . . .	81
5.2.1	Phishing as a security incident . . . . .	81
5.2.2	IT incident management . . . . .	84
5.2.3	Distributed cognition . . . . .	85
5.3	Participating organisation . . . . .	86
5.4	Methodology . . . . .	87
5.4.1	Introductions and setup . . . . .	87
5.4.2	Observing the Help Desk . . . . .	87
5.4.3	Interviewing university teams . . . . .	89
5.4.4	Periodic review of findings . . . . .	90
5.4.5	Interview data analysis . . . . .	90
5.5	Results . . . . .	91
5.5.1	Phishing campaigns– managing the load . . . . .	91
5.5.2	Converting escalated calls into protections . . . . .	98
5.5.3	Compromised accounts – a serious reputation and workload problem . . . . .	102
5.5.4	Information flow patterns are not in fixed orders or directions .	106
5.5.5	Providing feedback and guidance to end users . . . . .	108
5.6	Discussion . . . . .	111
5.6.1	Phishing management is a distributed cognition process . . . .	111
5.6.2	The University wants reporting, but can it handle all the reports?	112
5.6.3	Ticketing tools support distributed cognition but have room for improvement . . . . .	114
5.6.4	Best practice, is it helping? . . . . .	115
5.7	Limitation . . . . .	116
5.8	Conclusions . . . . .	117

<b>6 Discussion and Conclusion</b>	<b>119</b>
6.1 Discussion . . . . .	120
6.2 Future Work . . . . .	123
6.3 Limitations . . . . .	125
6.4 Conclusion . . . . .	125
<b>A Survey on the URL Report</b>	<b>127</b>
<b>Bibliography</b>	<b>177</b>

# Chapter 1

## Introduction

Protecting people against malicious communications – also known as phishing – is a challenging problem from the perspective of both user support and institutional support. Phishing works by targeting users' perceptions and tricking them into providing sensitive information to an entity they believe is safe but is actually malicious. A classic example is providing personal bank login information to an attacker, but equally problematic is providing the login credentials of the organisation for which a person works. Such attacks can be devastating, as they are often the first stage of a multi-step attack. The vast majority of serious attacks, such as ransomware, start with a successful phishing attack before moving on to more serious types of attack [1]. Hence, it is in the interest of both individuals and the organisations that employ them to identify and avoid these attacks. Unfortunately, doing so is challenging for many reasons. For one, the attacker's explicit goal is to make the attack difficult to distinguish from a real communication. If one type of attack fails, the attacker will adapt their approach. There are also issues of scale: only one person in an organisation needs to fall for phishing to open the door to more serious attacks. To handle these problems, organisations use a combination of automated protections, user-level training, user-level support and an active information technology support approach.

Organisations put resources into protecting against phishing partially because phishing is one of the most common and disruptive attacks in the UK [1]. Financial losses resulting from phishing can also be rather expensive, exceeding \$29 million in the US alone in 2017 [2] and growing to \$1.7 billion in 2019 [3]. Financial losses resulting from phishing can also be highly expensive, exceeding \$29 million in the US alone in 2017 [2] and growing to \$1.7 billion in 2019 [3]. In addition to financial losses, phishing also entails risk to organisational reputation. Attackers sometimes

use a compromised organisational account to attack someone at another organisation – for example, by sending a payment invoice between two contacts who often exchange such invoices [4]. Having a phishing attempt be sent from a legitimate account may indicate to clients that the organisation does not take security seriously. Phishing can also involve more than the mere collection of sensitive information. Some phishing involves malware, where even clicking on a link or opening an attachment can lead to a successful attack [5]. Emotet was an example of such malware: it initially spread by inducing users to click on links that then downloaded a malicious executable. Whilst Emotet started as a ‘simple’ phishing attack, it was able to use privilege escalation attacks to spread ransomware and harvest sensitive information [5], resulting in an average remediation cost of \$1 million per incident in the US [6].

Phishing is also specifically designed to be difficult for both humans and tools to accurately identify [7]. Regarding tools, attackers can craft their messages to resemble normal communications by using similar word frequencies, phrases and structures. They can also avoid directly naming commonly targeted organisations, such as PayPal, and instead use harder-to-detect images or language that might be clear to a human but are challenging for a machine to identify, such as ‘your recent purchase’. Similarly, phishers attempt to use words and phrasings that target humans’ methods of assigning meaning to the communications they receive – for example, using the phrase ‘IT Account Services’ without naming the organisation and letting the victim fill in whatever IT account they believe is meant. In other words, humans naturally apply information from their own contexts when they are reading an email of any type. Such contextual information may include who they expect to email them, with which financial organisations they have accounts and what a ‘normal’ email communication looks like for different communication partners. Although a skilled attacker can use some of this context to trick users, it is also one of the best defences users have, as most phishing contains some element that does not match the actual context. For example, if an email claims to be from PayPal but the ‘from’ address uses a different domain, the email does not match the user’s contextual expectations. Unfortunately, however, attackers can manipulate a surprising number of elements in an email, including the ‘from’ address, which may be either spoofed or a compromised legitimate account. As a result, users must rely on a combination of background knowledge about what ‘normal’ communication from the entity in question looks like and a small number of highly accurate features. One example of a highly accurate feature is the link (URL) on which the attacker is attempting to induce the user to click, as this link must route

to a site under the attacker's control and therefore cannot be a link to the legitimate site. However, these highly accurate features are harder for end users to understand and compare with their context. Without support, most users cannot accurately determine whether a link will lead to the website they expect [8,9]. The key lesson here is that phishing is complex and its accurate detection requires a combination of abilities, some of which are unique to humans (e.g. applying context) and some of which are much easier for machines (e.g. parsing key facts from URLs).

While ideally all end users would identify and ignore all phishing attacks, organisations also generally recognise that even skilled and well-trained users make errors and click on or interact with email elements that they should not. Organisations therefore can take action to prevent or limit the damage caused by users interacting with phishing. The most basic action is to provide training and education around phishing which improves the ability of employees to detect and avoid it. Organisations can therefore take action to prevent or limit the damage caused by users' interactions with phishing attempts. The most basic action is to provide training and education on phishing, which improves employees' abilities to detect and avoid it. Organisations can also use their own network and security controls to protect users who do interact with phishing attempts. For example, if an organisation knows that a phishing email has been sent to many employees directing them to click on a link containing `paypal-accounts-freestuff.evil.com`, the organisation can block all attempts to connect to that domain using its firewalls. In this way, organisations can use information they know about an attack to support and protect employees, rather than relying exclusively on employees to detect and avoid attacks on their own.

Organisational phishing management incorporates both proactive and reactive defence elements, which feed into each other within the phishing ecosystem. From an organisational viewpoint, a phishing ecosystem has several players: a phisher, target users (who may become victims in a successful attack) and the organisation's information service (IS) teams (who aid in rectifying instances of phishing). As seen in Figure 1.1, when an attacker sends a phishing email, the first layer of defence is the proactive protections that attempt to prevent it from reaching users. These proactive measures include activities such as setting up firewalls [10], setting up phishing and spam detectors for incoming emails [11], training end users to identify phishing attempts [12] and using email and browser warnings for people who click on malicious links [13, 14]. However, automatic proactive solutions are not fool-proof [15], and even with such protections in place, a small percentage of phishing attempts can get

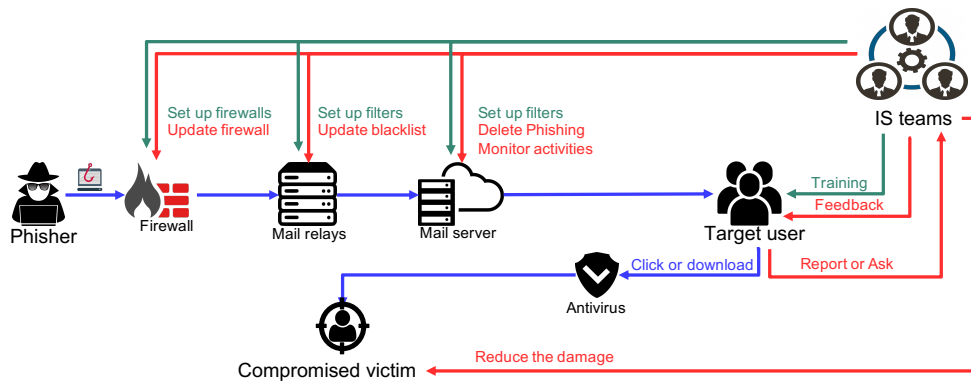


Figure 1.1: The proactive and reactive measures in the phishing ecosystem where the *green* arrows depict the proactive measures, and the *blue* arrows depict the path a phishing email takes after a phisher launches an attack as it passes through the proactive measures. Then the *red* arrows indicate the processes IS staff use to respond to a reported phishing attack.

through, leading to victims [16, 17]. Since automated approaches cannot catch everything, organisations also rely on employees to identify and report phishing. Training itself is a proactive measure, since it happens in advance of the phishing attempt, but in order to be effective it needs to be paired with reactive approaches. A user may identify potential phishing based on proactive training, but once they have done so, they may need feedback regarding whether what they have identified is indeed a phishing attempt [18, 19]. This requires the organisation to have resources in place to provide such feedback, as well as to reactively respond to any identified attacks [4]. Other proactive measures, such as setting up automated detectors, also require reactive approaches, such as monitoring alarms, to determine whether the identified abnormal activity is indeed malicious and take action accordingly [20–22]. Once attacks are identified, organisations are then in the position of being able to mitigate phishing impacts and reduce the damage experienced as quickly as possible by updating and improving proactive measures and removing the malicious content or activities (e.g. updating blacklists, removing phishing emails from inboxes). Such a multi-layered mix of proactive and reactive defences is generally considered best practice and is recommended by the UK’s National Cyber Security Centre [4].

## 1.1 Overview of the Thesis Research

User support is vital in the phishing ecosystem, wherein users do not have the absolute ability to identify phishing. In some cases (e.g. URLs), users lack the ability to easily understand the destination and are therefore unable to leverage their own contextual knowledge to best protect themselves – and indeed, even if they are able to understand the content and protect themselves, they may still require assistance with more complex communications [23]. This can be problematic, as some users do not contact the relevant IS personnel for advice due to their expectations of a delayed response [24]. Further, organisational personnel have access to their organisation’s email systems so they can use reported emails to stop further messages from a specific attack and protect other users who might also be targeted. In both cases, users will require support from IT in order to protect themselves.

In this work, I examine two primary elements of the aforementioned ecosystem: :

- Fast user-facing support (to aid users in accurately and independently evaluating URL safety).
- Understanding how phishing reports are handled at a university.

These two key topics were selected because they cover two different ways in which users can protect themselves: individually, by using tools to detect phishing themselves, and collaboratively, by relying on support staff to protect them.

### 1.1.1 Human-based phishing detection

I focused on URLs in my research because they are the most common element of phishing attacks and are reliable indicators of such attacks [25]. In the current user training research, it remains unclear which phishing features humans need in order to identify phishing attacks; moreover, the majority of users fail to recognise phishing URLs if obfuscation techniques are not covered in trainings [26]. Whilst research concerning how people parse and use URLs is limited, URL features have been extensively studied in the domain of automated phishing detection [27].

#### Exploring the phishing features

In this part of the research, I sought to identify a comprehensive set of effective phishing features based on the automated detection domains and then examine how those



features overlapped with the elements explored in research on how people detect phishing using URLs. To identify this overlap, I conducted a structured literature review. However, given the extensive scope of research on URL features, I reviewed literature on URL features in automated detection research (such as machine learning) as well as literature on human studies involving URLs until I reached a point of saturation and found no new URL features. By combining these two literature sets, I was able to answer the following two research questions (RQs):

RQ 1. What are the phishing URL features used in existing research?

RQ 2. Are the features used in the automated phishing detection research also explored in user training and support research?

This review, detailed in Chapter 3, allowed me to identify a large number of URL features that have been shown to be effective in automated detection research. I also concluded that only a small number of these features have previously been used in human-factors work aimed at helping users.

### **Designing a phishing report**

Using the results of the review, I explored the features I uncovered via a real-time support report, detailed in Chapter 4, that provided available information about URLs to aid users in reading them and, ultimately, in accurately judging their safety. However, the list of features was too large to fit in a report paper. Thus, to provide a wide variety of necessary features to users and help them understand the meaning of such features, I first answered the following questions:

RQ 3. What information about URLs best supports humans in detecting phishing URLs?

RQ 4. What are practical approaches for presenting information about URLs to Internet users?

I examined the initial design of the report with the minimum set of reliable phishing features (Section 4.4) that users can use to accurately recognise phishing URLs before iteratively refining the report with the aid of eight focus groups (Section 4.5).

### **Evaluating the effectiveness of support**

Whilst the goal of the focus groups was to evaluate the report and refine it until a version was reached that satisfied participants, more investigation is required to suf-

ficiently measure the broader effectiveness of the report. The questions I have for evaluating the report are as follows:

- RQ 5. What are the range of values likely to appear under the report's actual usage?
- RQ 6. Can participants use the proposed URL report to understand where a URL will lead and differentiate between legitimate and malicious URLs?
- RQ 7. Can participants understand the meaning of each phishing feature presented in the report?

The first question resulted in a test of the visual appearance of the designed URL features report on 2,617 phishing URLs and 2,023 safe URLs (Section 4.6), whilst the other questions focused on the subjects' abilities to use the report to judge URL safety and their comprehension of the presented features (Section 4.7). The results indicate promising effectiveness of the proposed report.

### **1.1.2 Organisational handling of phishing**

This thesis began with an exclusive focus on individuals, since they are the targeted victims and the first clickers of phishing URLs. However, when individuals contact the IS team in charge of handling phishing (either because they detected a phishing communication or because they need their help), the information within such emails can be used by the IS staff to update the organisation's defence systems and protect other users who are susceptible to phishing attacks. Whilst this is a broad picture of phishing incident handling and is the expected practice, little is known about the actual practice, workflow and challenges such teams face. Hence, I conducted a case study in collaboration with an academic organisation to answer the following RQs:

- RQ 8. How does an academic organisation respond to reported phishing emails?
- RQ 9. What barriers does an academic organisation face when responding phishing attacks?

To answer the questions, I collaborated with the academic organisation's IS department by observing the help desk team, who deal with user reports of phishing incidents, and interviewing the teams involved in phishing-related issues, including the help desk. My findings indicate that the workflow of phishing handling is a distributed cognitive process requiring multiple distinct teams with narrow system access and tactical knowledge. In addition, phishing reports may overwhelm the involved teams and hinder effective responses to phishing. This suggests a need for future research on how to

utilise automation to better handle phishing reports.

## 1.2 Research Contributions

This thesis consists of two projects, both of which are related to the phishing ecosystem.

In the first work, in Section 1.1.1, I surveyed the literature related to URL phishing and identified a subset of reliable and human-friendly features to design a report that supports users in reading URLs and accurately judging their safety. My contributions here are as follows: 1) identifying the URL features used in phishing research; 2) determining the gap in the URL phishing features used in automated and user training and support research; and 3) designing and evaluating a URL features report intended to support average users in judging URL safety.

In the second work, I focused on understanding the reactive measures that are triggered after the reporting of phishing emails within the context of a wider view of the phishing ecosystem. With a focus on phishing reports, my work focused on investigating the practices and challenges organisations face when dealing with phishing incidents. My contributions are as follows: 1) providing a detailed picture of the reactive measures for phishing attacks; and 2) identifying the challenges academic organisations face within the phishing ecosystem.

This thesis's findings suggest that overcoming phishing is not easy and that better ways of solving this issue still need to be established. Although the current literature concerning URL features has paved a clear path for developing support interventions, training users to read URLs is complicated. Similarly, responding to phishing incidents has been overlooked in phishing-related research. Hence, there is a clear need to step back and examine the phishing ecosystem and understand what gaps need to be addressed for phishing defence.

## 1.3 Thesis Outline

The remainder of this thesis is organised as follows. Chapter 2 reviews the existing literature on the features of phishing URLs. Chapter 3 provides a foundation for human-level support by reviewing common URL phishing features in the automated and human phishing detection literature. Following this, Chapter 4 builds on the findings in the previous chapter by discussing the design and evaluation of a human-facing inter-

vention. Chapter 5 introduces the findings of a case study concerning organisational handling of phishing incidents. This is followed by a discussion of the implications of the findings and the overall limitations of the thesis in Chapter 6. Chapter 6.4 presents a conclusion for the thesis as a whole.



# Chapter 2

## Background and Related Work

To make a positive impact on the fight against phishing and solidly situate my thesis in the literature, this section provides an overview of phishing attacks and a wide view of the phishing ecosystem from an organisational perspective.

### 2.1 What is Phishing?

Phishing is a play on the word ‘fishing’: phishers throw bait to ‘fish’ for a victim’s assets [28]. The internet community initially recognised phishing as a threat during the 1990s [28]. Since then, phishers’ spoofing strategies have constantly evolved to match trends [29], such as adjusting attacks to match current events [30]. For example, during the COVID-19 pandemic in 2020, attackers spoofed government and healthcare authorities like the World Health Organization (WHO) [31]. In 2018, people in the corresponding region received many emails related to the European Union’s General Data Protection Regulation (GDPR) compliance, some of which were phishing emails spoofing businesses in several industries and targeting a wide range of sectors, from the general population to high-profile individuals [32].

‘Phishing is a scalable act of deception whereby impersonation is used to obtain information from a target’ [33, p. 8]. Its scale varies from large attacks targeting hundreds of victims to tailored attacks targeting a small group of people. Furthermore, because it heavily depends on social engineering tactics that exploit human factors, it is the most preferred and common type of cybercrime attack [34].

## 2.2 Phishing Ecosystem

This section covers the preparation and execution of phishing attacks, as well as what attackers gain when they achieve a successful attack (e.g. stealing credentials, damaging assets). I discuss how individuals and organisations, as important elements of the phishing ecosystem, are motivated to apply phishing protections, with a focus on the literature on phishing emails and URLs. Furthermore, the phishing ecosystem is discussed, with particular emphasis on the gaps in the literature in this field.

### 2.2.1 Planning and executing the attack

To conduct a successful phishing attack, phishers plan, set up and distribute the attack.

#### Planning

When phishers plan their attack, they consider several factors, including the target identities to be attacked, communication medium and technique [35]. With regard to target identity, they may enact a large-scale attack on specific organisations or random individuals as well as a small-scale attack on a group of people about whom they search for essential details to help them better plan the attack tactics and tools [36, 37]. For example, a large-scale phishing attack (campaign) needs looser tactics, whilst targeting high-ranking individuals (‘spear phishing’) requires a more customised approach [37].

The communication medium is the means by which the attacker initiates the attack [38, 39]). These media may be websites, instant messengers (IMs), mobile apps, social networks (SNs), short messaging service (SMS) or multimedia messaging service (MMS) messages or voice media, though the most often targeted medium is emails [30, 36, 38, 39].

To execute their plan, attackers decide on certain technical and behavioural attacking techniques [38, 40]. Example technical techniques include spoofed URLs – that is, manipulated malicious URLs to look like trusted ones (e.g. by adding the expected identity to the wrong position in the URL, such as the subdomain [8]) – and the use of shortened links, which makes it difficult for users to predict which website they are visiting [41]. Another example is spoofed emails, which manipulate the sender address to impersonate a legitimate user [42]. Behavioural techniques are exemplified by social engineering: the tactic of persuading or manipulating individuals to comply with the attacker’s request (e.g. to provide sensitive information [43]) through

the use of the principles of authority (tendency to interact with important senders, e.g. the government), commitment (tendency to comply with requests that are in line with the recipient's principles, e.g. helping poor people), liking (tendency to interact with senders the recipient likes), perceptual contrast (tendency to perceive the least threatening email as more trustworthy), scarcity (tendency to comply with urgent requests) and social proof (tendency to perform actions solely because others are doing so, such as visiting a top-rated shop). Each of these may result in higher click rates [44–46], as they may exploit curiosity, fear and empathy [29, 45, 47]. The most popular technique, however, is spoofed URLs, as they can be used in several communication media (commonly emails [25, 38]) and may be complemented by social engineering that better suits the target and the attack medium [38]. These strategies involve sending an email that utilises contextual information concerning the target individuals and asks them to click on a provided link to lure them into visiting malicious websites [38, 48, 49].

### **Set-up**

Preparation of phishing material depends on the targeted communication media. For example, when attackers use the email channel, they prepare the body of the email text and the malicious element, such as URLs or attachments [38]. Notably, phishing attacks do not require specific expertise, as phishers can either develop a phishing attack from scratch or use an existing resource [38, 39]. For example, phishing kits can remove obstacles for criminals to execute their attacks, as they contain pre-designed webpages for popular organisations or scripts to collect victims' credentials, working as a back-end component for the phishing attack [35, 37, 50] or, indeed, other online resources that are found on the dark web. Phishing websites, as an example, are available on the market, with many options to fit each phisher's needs, whether they are a novice or are already tech savvy; the average cost of a pre-designed web page is \$24, and the cheapest template costs about \$2 [37].

### **Distributing the attack**

In this stage, the actual attack takes place. If the medium is email, the attacker uses the email addresses they collected in relation to their target to send a phishing message tailored to deceive them [51]. Regardless of whether they are focusing on a small group of people or a larger-scale target, they do not want their emails to be caught by the recipient's spam filter. One way to avoid this is to send emails from a legitimate



sender email address or to impersonate a high-ranking employee in the organisation, which requires either a successful fraud attack against the impersonated person (most common [52]) or high-level technical skills [37]. A simpler option is to send the email via a public mail service such as Gmail and manipulate the sender email address to make it appear legitimate [37]. Lastly, phishers often distribute phishing emails to a mailing list to which the victims are subscribed [37]. Spoofing of email addresses makes the email address unreliable indicators of an email's trustworthiness. This was proven during the 2016 US election, when security experts agreed that the phishing email sent to John Podesta was actually from Google [53].

## 2.2.2 Proactive defences: Where an attack might fail

Various solutions can help combat phishing attacks, but none of them is a 'silver bullet'. Rather, organisations generally apply multiple solutions for more than one method to prevent phishing attacks, an approach referred to as 'multilevel defence'. Attackers must destroy or circumvent the outer layers (e.g. filters) before reaching the inner layers (i.e. end users) [30, 39, 51]. Here, I discuss organisational-level defences, as they also involve individuals.

### Invisible automatic defences

Automatic preventive measures are designed with the goal of reducing the number of successful attacks by passively filtering incoming emails [54–58]). They scan all incoming emails and URLs and remove any that appear on a list of known malicious communications (i.e. blacklists) [59] or that result in a score above a certain threshold (computed using a combination of approaches). Many organisations either use well-known blacklists or manage their own internal blacklists. Blacklisting causes financial damage for attackers, since it reduces traffic by up to 95% [60]: they either block phishing messages or take down phishing landing pages and phishing emails [61].

Whilst effective in detecting a significant amount of malicious content [15], blacklists only include previously seen communications. Thus, when it comes to identifying new threats, machine learning approaches are often used to automatically categorise communications as either safe or fraudulent [15, 27, 39, 62]. Some works have suggested splitting automatic detection into two layers to produce real-time prevention [63], whilst others have considered authenticating senders to block attempted forged emails [64].

From a usability perspective, automatically removing phishing communications is optimal as long as it is highly accurate and does not delete important communications [40,65]. Fortunately, automatic phishing detection works well, generally producing true positives and demonstrating minimal false positive rates [15,59]. However, the nature of phishing is such that, if any attack gets through, it could potentially be very expensive for an organisation [2, 66]. Thus, users are warned about phishing emails and URLs and are expected to adhere to these warnings, as they are an important component of phishing prevention. However, users do not always adhere to warnings due to the lack of comprehensibility of the warning text [67], the trust in and familiarity with the website [67] or overconfidence in users' own knowledge.

### **Proactive training**

Humans are the last point of defence for organisations, as detecting phishing emails requires humans' awareness of the context in which they have received the phishing message – for example, from whom they expect to receive a message and which websites they expect to visit [61]. However, people are often either unaware of what to look for in a phishing message [68] or follow weak or inaccurate heuristics to differentiate between phishing and legitimate content. Examples of weak heuristics include looking at the From address and the the presence of grammar and spelling errors [69, 70], the existence of a recognisable logo [71, 72] and the use of users' names in the message content [70, 73, 74], which cannot help with intentional visual deception by mimicking the visual design of a legitimate email [69]. Examples of heuristics that result from misconceptions about security indicators include looking for the lock symbol in the browser address bar, which users generally believe indicates the website is safe [69, 75], leading them to judge a website's safety based on the existence of the lock. However, the only actual safety indicator is the presence of an extended validated certificate, indicating that a website owner has verified their ownership of the website and its legitimacy [75]. In addition, studies have shown that users trust URLs when a brand name is embedded somewhere in the URL [8, 74] or if the URL is simple and short [75]. Finally, people fall for phishing attempts more often when they are familiar with the websites at hand or are overconfident in their knowledge about phishing. The studies conducted in [67, 76, 77] indicate that website reputation and user familiarity with a website play important roles in whether users ignore a warning, which indicates the importance of enhancing users' comprehension.

In contrast, experts usually identify phishing emails by hovering over links and

looking at the sender's email address along with other technical information of the email. They have typically learned to look at these features based on training materials [61]. Training average users to identify phishing messages is a common approach that is often combined with automatic detection [12, 71, 78, 79]. On average, organisations spend about \$290,000 every year on training [80], which is often either done upfront [12, 78] or embedded in daily work [81].

Upfront training explains concepts in a dedicated training session. Its effectiveness depends on the user's ability to recall and apply these learned lessons in later situations; this can be challenging because people tend to forget unused information [60, 82], thus necessitating periodic trainings [30]. Moreover, even if users can remember the information learned in trainings, attackers also continuously adjust their tactics over time, invalidating some of the learned information [26, 83].

Embedded training is designed to be integrated into users' daily routines, such as receiving training if they fall for phishing attacks. Whilst overall effective [84], embedded training is also costly, as it requires a human administrator to take the time to craft simulated phishing communications [40] that must be realistic and up to date [85, 86].

Even if training can improve users' ability to detect phishing websites, URLs can be manipulated in ways that are not easily discernible to the human eye. No amount of training will solve these issues because humans' physical limitations lead to falling for phishing [61, 69, 78]. For example, replacing an English character with an identical-looking non-ASCII character (e.g. the letter 'B' with the Greek letter 'Β') or replacing a character or two with similar characters (e.g. the letters 'rn' with the letter 'm') makes it more difficult for users to notice such manipulation at a quick glance, with about 40% falling for this tactic even after training [87]. Another example is redirection, wherein users are redirected to another website or webpage after they click on a link, which they would not know without clicking on the URL (for more details on this point, see Section 4.2). Consequently, whilst phishing education offers skills that people need in order to improve their phishing detection abilities, they are unlikely to be able to accurately identify suspicious URLs without the assistance of tools that are currently not easy to find or use [83, 88].

### **Phishing detection support**

In phishing detection support, a computer assists the user by providing extra information. These support systems can take several forms, including browser warnings,

chatbots and toolbars. Park et. al. have suggested that this collaborative approach is best capable of achieving the desired results through its utilisation of the complementary strengths of a human and an agent [7].

The browser plugin Netcraft [89], for example, warns users about blacklisted web pages once they visit them. Otherwise, it clearly displays the website's country, site rank and hostname, amongst other facts, to help users evaluate fraudulent URLs. Similarly, SpoofStick presents the domain name in the browser toolbar to highlight cases wherein a legitimate-looking domain name is in the wrong position [90]. Yang et al. designed security warnings based on website traffic rankings [91]. The Faheem chatbot provides basic facts about any given URL, including the existence of any misspellings, non-ASCII characters and redirection [92]. Users can also ask the bot to elaborate on any term and receive a more detailed explanation. TORPEDO, a Thunderbird add-on, presents and highlights the domain of a URL linked in hypertext in an email. The add-on disables the link for three seconds to encourage users to stop and think about the URL's safety [70].

The above security indicators take a similar approach to the report proposed in Chapter 4. Users are presented with information about the URL prior to visiting it under the assumption that, with support, they will be able to identify unexpected aspects of the link. This differs from existing solutions in that it focuses on how to express potentially complex URL and web-hosting concepts to users in an easy-to-comprehend way that fits their need [93]. Existing solutions focus on providing either support for more technical users (who may already have a strong lexicon of internet terms such as 'host', 'domain' and 'hosting provider') or basic support that does not add much to upfront trainings. My work aims to bring this type of information to a broader audience.

### **2.2.3 When defences fail**

Phishers will not achieve their goals if victims do not interact with the phishing attempt in the way that the phisher expects. Interactions can vary from clicking on links to downloading attachments to replying to the phishing email with sensitive information [38]. Organisations can put in place means of protection against phishing, but attackers need only one person to ignore the warning and give away their credentials; if this is achieved, the attacker can use the victim's response to perform many other cybercrimes [60] and then remove the evidence of the attack so they can repeat it later

without being caught [15].

The potential later cybercrimes may involve performing spear phishing by utilising the lateral phishing discussed earlier, in which the attacker utilises the privileges of an insider [94–96] by using the victim’s (usually email [95]) credentials to initialise more dangerous and sophisticated phishing attacks [39,52,95], which can bypass email filters [95, 96]. This type of attack can harvest more clicks, as knowing the sender contributes to higher trust in email links [47].

Another breaking-in scenario involves the installation of malware after the victim clicks on a malicious link or attachment. Malware can be used for a wide range of further attacks, including data theft, ransomware, remote-access Trojans, malicious network botnets or DDoS attacks [15, 30, 36, 49].

Other follow-ups on phishing attacks can include sending spam [95, 96] or finding sensitive information in past emails [96]. Sensitive information may be used to influence political decisions; for example, in 2016, the phisher mentioned in has exposed John Podesta’s emails to the public during the 2016 US elections [53, 97].

Although phishers can acquire cash using social engineering tactics (e.g. asking users to send them money, stealing financial credentials), they may further sell the stolen data to the market to obtain additional cash [37, 38].

#### **2.2.4 Reactive measures: Opportunities to mitigate the attack**

Phishers are continuously attempting to find new ways to bypass the proactive protections discussed in Section 2.2.2. Therefore, organisations also need to have reactive processes in place that monitor for new attacks. Once a phishing email reaches users’ inboxes, some users will open the email and click on the links (11%) or attempt to give away data (4%) [98]. If the proactive solutions are effective, the user will be blocked from visiting the page by a firewall [11, 99] or their web browser [100, 101]. If user training has been successful, some will identify the phishing attack and follow the advice to report it [102–104], typically to an IT help desk or some other type of security operations centre.

Once a report is received, it becomes a phishing incident that requires a specific incident handling process [105]. Here, internal teams will use the reported phishing emails to apply immediate mitigation [20, 106], such as adjusting and applying filters to email inboxes to remove the offending phishing email and prevent any further interaction from users. The teams may also use log files to determine which, if any, users have

already clicked on the malicious links and force a password reset for those users. They may also use the email content to update proactive protections (e.g. updating the mail relay filter to identify and prevent similar future emails from breaching the defences again). Timely responses greatly aid organisations in reacting to and mitigating attacks, thus reducing the number of potential victims who engage with the phishing communication and therefore reducing or avoiding organisational damage [20, 107, 108]. Although these reactive measures are applied in well-established organisations, little is known about the day-to-day process of phishing handling, the details of the associated workflow and the challenges organisations face when responding to phishing incidents. Hence, it is necessary that we learn more about organisations' reactive measures, as difficulties responding to phishing attacks give phishers the opportunity to attack more individuals and cause more damage to organisations [95, 96].

## **2.3 Conclusion**

Looking at the phishing ecosystem indicates that it is difficult to teach users about all the aspects of phishers' tactics and that there is not enough literature about the current state of phishing tactics that users can use to identify phishing emails. In addition, little is known about the reactive measures used by organisations, such as the practices they follow and the challenges they encounter when handling phishing. Thus, I conducted a more in-depth study to understand the phishing-handling workflow and fill the gaps in the phishing ecosystem literature.



# Chapter 3

## A Review of Human- and Computer-Facing URL Phishing Features<sup>1</sup>

### 3.1 Introduction

Phishing is not only expensive [109], it is also hard for both humans and computers to detect accurately [7, 110]. After all, the goal of a phisher is to first get their message to users by bypassing automated detection systems, and then deceive users into interacting with the message. However, while phishers can manipulate many aspects of their communications, there are a few aspects that are very challenging for them to fully hide, such as the destination of URLs. In this paper, we review phishing research and catalogue URL-based anti-phishing features aimed at both humans and automated systems. Our aim is to create a foundation for future research to improve the state of human-facing support.

Ideally, all phishing detection, URL or otherwise, would be done automatically without human involvement but there are two major challenges to doing so. First, while automatic detection is impressively accurate with classification rates as high as 99% [111] and the preferred first line of defence for most users [112], the remaining 1% is highly problematic and potentially very damaging [113]. Second, humans are

---

<sup>1</sup>This chapter was published in the European Workshop on Usable Security (USEC) in June 2019 [83], and it was a collaboration with Ghaidaa Rummani and my supervisor, Kami Vaniea. I was the lead author of the paper and I have the second author as a second coder in order to ensure an objective and transparent analysis and reporting of findings. I wrote the first draft and collaboratively edited it with the co-authors.



needed to report and annotate new phishing attacks so that automated systems can in turn be updated to detect the latest threats. Effectively, humans label phishing, which is then used to train automated systems, which in turn causes phishers to change tactics [26, 114], leading to undetected phishing, which is then reported by humans, starting the whole cycle over again.

Automatic phishing detection of URLs comes down to deciding if a URL's destination, is 'bad' or not. For a human, 'bad' can be defined as any website other than the one they intend on visiting. But computers lack users' understanding of context, so they must instead define 'bad' based on pre-labelled lists (black /white lists), heuristics (rule-based) [59], and building machine learning classifiers using labelled examples [27]. This difference means that humans and computers likely find different features more or less useful when making phishing judgements. Park et al. conducted a lab study to compare the abilities of machines and humans to detect phishing emails [7]. They found that humans are as good as machines in labelling legitimate emails. For phishing emails, some emails were easy to spot for humans but not machines while others were easier for machines to detect than humans. They concluded that a collaboration between machines and humans is needed to reach an optimal solution to combating phishing.

Since humans are not naturally skilled at detecting phishing, *education* and *support* are used to help them accurately detect it. Education approaches attempt to train users to look at specific features of the URL or communication. Some trainings also include guidance on how to use phishing features to differentiate between a safe and malicious page. Educating users takes time, and providing updates to that education is also very expensive, so theoretically there is a natural bias towards teaching features that are easier for humans to understand and that are stable across time [26]. However, some URLs are impossible for people to read even if they have high awareness. Punycode (RFC 3492) URLs, for example, allow Unicode characters to be encoded using ASCII such that there is no human-visible difference between the real URL and the malicious one even though the computer would see a difference [88]. Detecting such problems requires *support* systems where the computer extracts and highlights feature data to support the human in making a decision. Two recent examples from research are TORPEDO [70] and Faheem [92]. Both of which provide the user with just-in-time information (features) with the goal of supporting users' decisions.

In this paper, we explore the literature to answer two questions: (1) What phishing URL features are used in existing research? (2) Are the features used in the automated

detection research also explored in human-facing research?

To our knowledge, no prior literature review has considered URL-based phishing features in reference to both humans and computers. Several works have compared automated and human training approaches [15, 40, 112, 115, 116]. Two general surveys looked at automated web phishing detection [117, 118]. Phishing features used in machine learning solutions were previously reviewed in [27] (2017) and [119] (2015). Reviews of feature usage in web content [60, 120] and DNS [121] also exist.

We review phishing literature from three libraries, compile a list of phishing features, and then group those features into categories. We find that there are a very large number of features and that all feature types have been tried in the automated detection literature. However, several categorisations of features have minimal exploration in the human-facing work. Examples include host features (i.e. DNS) and page popularity (i.e. PageRank). We also find that the domain of the URL is heavily used in human-facing work, but minimally used (beyond blacklists) in automated work.

## 3.2 Uniform Resource Locator (URL)

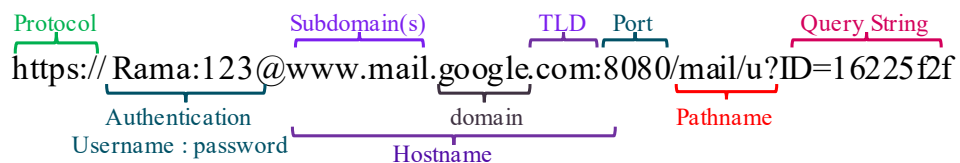


Figure 3.1: Example URL along with its structure.

As shown in Figure 3.1, a URL is made up of a protocol, authentication, hostname, port, pathname and query. These are in turn made up of smaller components. Hostname, for example, is made up of the subdomains, domain, and top level domain (TLD). Only the protocol, and TLD are strictly necessary to create a working URL, though in practice the domain is also required. Generally to resolve a URL, the browser uses the Domain Name System (DNS) to locate the hostname’s IP Address, then contacts the server using the protocol and port, it also provides the path, query, and authentication so the server can ‘locate’ the requested resource.

The phisher wants users to load a page under the phisher’s control. To do so, the phisher must, for a domain they control, accurately state protocol, domain, and TLD. This domain/TLD could be similar to an organisation they wish to impersonate, but

cannot be identical. Subdomains are controlled by the domain owner, so a phisher can create arbitrary subdomains for their domain with virtually no oversight. The other elements (authentication, port, path, and query) can be ignored by the malicious server, and therefore can contain any syntactically valid information the phisher wants. The limits placed on the URL, mean that the domain/TLD is the most accurate in terms of where the URL will lead, but a phisher can select the domain to be confusing or put valid-looking URL elements into the other elements to confuse humans readers [75].

### 3.3 Methodology

Our research goal is to create a representative list of URL phishing features that have been tested with machines and/or with humans by reviewing past research papers.

#### 3.3.1 Procedure

Literature was collected first using Google Scholar (October 2018), and then ACM Digital Library and the IEEE Xplore Digital Library (December 2018).

We searched Google Scholar with the keyword ‘phishing URL’ to look for phishing features. One researcher started with the publications rated most relevant and reviewed the title and abstract to make sure the paper matched the inclusion criteria (Section 3.3.2). They then reviewed the content, marking any sections that discussed phishing features. Identified features were then included in a spreadsheet used to track reviewed papers. They kept reviewing till new papers were no longer adding meaningfully different phishing features. For example, the features ‘count of subdomains’ and ‘count of dots in hostname’ effectively measure the same thing and were not considered meaningfully different. After completing the procedure for the first database, the researcher reviewed the second, and then third databases. A second researcher then went through each annotated paper and verified that features had been accurately identified in the paper and the spreadsheet.

Next, we grouped the features into categories. Where possible, categories were formed and named using common conventions from reviewed papers. ‘Lexical Features’, for example, appears in several papers and always refers to features extracted from URL text. For each category shown in Table 3.1, we also summarize key aspects of features such as how they are used (automatic, human) and limitations to their use (time, storage, and dependencies).

### 3.3.2 Inclusion/exclusion criteria

We included papers whose primary focus was the determination of whether a URL would lead to a phishing page without requiring the loading of the full page content. We focused on features that detected phishing rather than other attacks, such as XSS. We excluded papers which were not focused on URLs, such as those only looking at email content without also analysing URLs. Poster papers, extended abstracts, theses, and technical reports were also excluded due to lack or limited peer-review. We also excluded content-based features that required the full page to be loaded as our focus is pre-load. Due to the limited number of papers related human-facing features, we included any paper that studied people's susceptibility to phishing and otherwise met the above criteria.

### 3.3.3 Limitations

Stopping reviewing at saturation limits the scope of the work and likely results in some features not being included. We decided to stop at saturation anyway because: 1) there are a very large number of URL phishing-related papers (~26,400 Google Scholar results), 2) papers tend to have high overlap in features used, and 3) the more effective features tend to appear in multiple papers. However, the result is still a bias toward well cited papers that closely match our search keywords.

Table 3.1: Summary of Identified Features

Feature Category	Feature Subcategory	Most popular feature	Use of the features			Criteria		
			Automated	Human education	Human support	Time	Storage	Dependency
Lexical	Domain	Domain	Low	High	High	Low	Low	No
	Other URL components	Authentication	High	Mid	Low	Low	Low	No
	Special Characters	Number of dots	High	Low	Low	Low	Low	No
	Length	Length of URL	High	NA	NA	Low	Low	No
	Numeric Representation	Raw IP address	High	High	Mid	Low	Low	No
	Tokens & Keywords	Phishing keywords	High	Low	NA	Mid	Mid	No
	Deviated domains	Similarity with PhishTank	High	High	High	Mid	Mid	No
Host	Embedded URL		Low	NA	Low	Low	Low	Maybe
	Whois	Domain age	Mid	NA	Low	Mid	Low	Yes
	DNS	No records	Mid	NA	NA	Mid	Low	Yes
Rank	Connection	Connection speed	Mid	NA	NA	Mid	Low	Yes
	Domain Popularity	Alexa Rank	High	NA	Low	Mid	Low	Yes
	PageRank	Google PageRank	High	NA	NA	Mid	Low	Yes
Redirection		No. of Redirections	Mid	NA	Low	Mid	Mid	No
	Encryption	Is it HTTPS?	High	Mid	Low	Low	Low	No
Certificate	Certificate values	Is EV?	Low	NA	Low	Low	Low	Maybe
		Query the Full URL	Mid	High	Low	Mid	Low	Yes
Search Engines		PhishTank	High	NA	Mid	Low	Low	Yes
	Black/White lists	Blacklisting the IP	Mid	NA	Low	Mid	High	Yes

## 3.4 Phishing Features

Our review identified 94 papers, 58 of which were from Google Scholar with an overlap of 24 papers in IEEE or ACM. We categorise identified features into: lexical features, host features, rank features, redirection features, certificate features, search engine features, and black/white list features. In this section, we discuss the primary features for each category and their use in automated, and human-facing detection methods in detail.

### 3.4.1 URL lexical features

Parsing the URL string itself and using the resulting components as features is very popular and reliable. Lexical features are attractive because they require low processing time, low amounts of data storage, and can be processed without having to call out other services [122], which is also a nice privacy feature. As a result, they have high real-time efficiency [123]. Since URLs are unique to a site, they are also impossible to fully spoof, so while it is possible to create a similar-looking URL (i.e. `pavpaypal.com` or `evil.com/paypal`) it is not possible to use the correct URL domain (i.e. `paypal.com`) in a phishing URL without first compromising the domain.

Although the use of URL lexical features alone has been shown to result in high accuracy ( $\sim 97\%$ ) [124, 125], phishers have learned how to make predicting a URL destination difficult by carefully manipulating the URL to evade detection [125–127]. Therefore, combining these features with others, such as the host based features (Section 3.4.2), provides higher accuracy [128, 129].

#### The URL domain

The domain is a prevalent feature in anti-phishing, likely because while phishers can register new domains, they are not generally able to attack a user visiting a legitimate domain. Extracting the domain is also easy, requiring only simple URL parsing. But using it alone to classify URLs is difficult because context, such as where the user wants to go, is missing. Instead it is combined with other features such as comparing it to the page title or meta-data [130, 131].

Human education papers commonly teach users how to parse out the domain as a way of enabling them to compare the domain to the one they expect to be visiting [12, 78, 132, 133].

For the papers using a human-support approach, such as those discussing Spoof-Stick and Netcraft, these tools used the hostname to help users correctly identify the sites they visit. In other tools, the domain part was pointed out as the destination [92] or highlighted once the mouse hovered over the link [70]. However, human-support only works if users are aware of what the correct domain is. Domains that do not directly line up with a recognised brand name, such as `www.nytimes.com`, can still confuse users even if they can parse the domain out correctly [92].

### Other URL components

The URL standard defines multiple components [134], and while the hostname is the most commonly used feature, the other components are also commonly used [135–138]).

The authentication components, identified by the presence of '@', appears right after the protocol, making it an easy place to put a brand name and fool users (i.e. `http://bank.com@evil.com`). Authentication components are rare in legitimate URLs, so nearly all modern automatic filters use it as a feature [111, 120, 122, 123, 130, 139–152].

Some automated detection papers use the existence of non-standard port numbers as a feature, where standard port numbers are either a common port number of the associated number with the protocol [135, 136, 144, 145, 149, 151, 153–158] because phishers use different port number to escape the detection [154]. However, port numbers are generally rare in URLs (0% in legitimate vs. 0.01% in phishing) [136].

'Non-standard' TLDs are also used as features but there is no consensus on the definition. Specific country-code TLDs (ccTLDs), such as '.cn' and '.ru', are used as features [156] while others focus on whether the TLD is a ccTLDs or a generic TLD such as '.net' and '.com' is used [158–160] and it is found to be a strong feature for classification (2017) [160]. In [149], ccTLDs are compared to host locations to see if the owner is located in the same country. Although ccTLDs are cheaper to obtain and sometimes used in phishing URLs, '.org' was found to be the most popular TLD for phishing websites in 2014 [114].

Other works apply weights to TLDs based on their training set [161, 162]. Weighting the features results in TLDs like '.info' and '.kr' being in the top phishing features while '.gov' and '.edu' being in the top legitimate features [162]. A set of 5 TLDs including '.com', '.net', and '.org' are also used [163].

Human-facing approaches have tried teaching users about URL structure compo-

nents, such as TLD and authentication, to enable them to differentiate between the hostname and other URL components [12, 78, 92, 164]. Faheem [92] warns users about non-standard port numbers.

### Special characters

The presence of special characters such as ‘/’, ‘=’, and ‘\_’ is an aspect that has been used in many papers [120, 129, 135, 136, 142, 146, 151, 160, 163, 165–167] along with the frequency of their appearance [145]. Based on analysis of PhishTank URLs, 77% of phishing hostnames contain special characters [165].

Hyphens are one of the most commonly used features and the existence of a hyphen symbol in the domain is a phishing feature in automated and human-support methods [12, 92, 120, 141, 142, 145, 146, 148–150, 160, 168]. Hyphens appear in legitimate URLs as well (2% in legitimate vs. 9% in phishing [135]), so they cannot be used as an indicator in isolation [123, 165]. Phishing websites tend to use the hyphen commonly to separate the brand name from the suffix (TLD) or prefix (i.e. `www-paypal.com`) [169], signifying the existence of a hyphen and suffix/prefix in the domain is a compelling phishing indicator. Other researchers included the number of hyphens as a feature [129, 151–153, 160, 165, 170]. The maximum number of hyphens in legitimate URLs hostnames is one while in phishing it is two or more [165]. Interestingly, the feature was one of the insignificant features in their classifier performance [160].

Dots and slashes are special characters that delineate components. Hence, the number of dots is linked to the number of subdomains and is a strong commonly used indication of phishing [120, 122, 123, 130, 135, 136, 140, 142, 146–150, 163, 165, 171–173]. in the URL [59, 111, 129, 135, 143, 161, 162, 166, 174] no info about the location [141, 160]— Analysis of phishing and legitimate URLs in [135, 136] found the number of dots in legitimate URL hosts ranges from 1 to 5, where 5 is rarely found in legitimate URLs, while in phishing URLs it ranges from 0 to 30. Some papers mark URLs with 3 or more dots as suspicious [152]. Therefore, the more dots, the more suspicious the URL [150, 152]. The number of slashes is also a phishing feature [122, 146, 149] with a threshold of five in some research [130, 165]. Having a hostname with no dots, consisting of only a single TLD, was also used as a feature [165].

The hyphen is the only special character used in human education [12, 164, 168] and human-support [92]. However, since hyphens only indicate phishing if there are too many of them, and ‘too many’ is not well defined [136], these features may not be



a good match for future human-facing research.

## Length

Attackers tend to use long complex URLs as another way of hiding the true destination. Length-type (character count) features are commonly used to detect phishing URLs. Though, shortened and simple URLs can also be misclassified based on it [175].

One common feature is the length of the full URL [59, 129, 135–137, 140, 147–149, 151, 152, 157–162, 165–167, 167, 171–174, 176, 177]. The URL length is one of the features that contributed best to the classifier performance of [171, 173]. Taking dataset bias into consideration, phishing URLs are typically longer in publicly available blacklists than non-phishing URLs which are usually Alexa top sites [177].

Length of other URL components is also used as a feature, such as the length of the hostname [135, 136, 141, 142, 161–163, 165, 171–173]– on average 20 characters in legitimate URLs [165], subdomain [149, 151], domain [129, 166, 167, 171, 173, 177], path [135, 136, 141, 160, 167, 170], or query [176]. The hostname’s length (max 240 in phishing vs. 70 characters in legitimate) was the most useful as compared to the full URL and path’s lengths [135]. Other features include average and longest domain and path token length, domain and path token count [129, 158, 160, 161, 161, 178], length of max-length in domain name [142, 145, 163, 167, 178].

Human-facing methods in our dataset did not consider the URL length as a feature; nevertheless, participants in [133] assumed the longer the URL the less secure it was. Alsharnouby et al. also found that people without training tend to classify URLs based on their perceived simplicity [75].

## Numeric host representation

Legitimate URLs primarily use the registered hostname of the website, while phishers sometimes use different representations of the hostname to hide the destination to make the URL text difficult to understand [179]. Examples include IP addresses (i.e. `http://216.58.204.46`) [77, 111, 120, 122, 123, 129, 130, 135–137, 140, 141, 143, 145, 147–152, 155, 158, 161, 162, 165, 169, 176, 179, 180], dotless IP address, also called ‘decimal address’ (i.e. `http://3627733550/`) [146], encoded IP address hex value: (i.e. `http://0xd83acc2e`) [120, 129, 135, 136, 179], or even encoded a hostname or part of it as Unicode (i.e. `http://%63%6E%2E%63%6F%6D`) [120, 142, 143, 147, 165]. In [165], 65.16% of phishing URLs contained Unicode. IP addresses are the most common

feature in automated detection and the only numeric feature used in human-facing detection [12, 70, 78, 92, 168].

### **Tokens and keywords**

Many automated detection papers tried tokenizing the URL and treating it as either a bag of words [124–126, 158, 161, 162, 172, 175, 181, 182], an N-gram [13, 127], a combination of tokens and bi-grams [174], or character frequency [141, 153]. The bag of words approach is effective, but the models are unstable over time and require frequent updating [129, 158].

Common keywords are also looked for in phishing URLs, such as ‘secure’, ‘account’, or ‘confirm’ [111, 122, 123, 129, 135, 136, 145–147, 149, 152, 161, 162, 165, 176, 179, 179, 183]. Similarly, human education has tried teaching users not to click on URLs with security-related keywords [12, 168]. However, keywords are unstable over time because attackers adapt and change words [123, 177]. Sananse et al. argue these features appear in both legitimate and phishing websites [157].

Number and average of terms are used as features [160, 165, 171, 173], with  $>4$  terms in the host indicating phishing [165].

Path extension such as ‘.txt’ is a feature as attackers can add scripts to benign websites making ‘.js’ pages more dangerous [158, 159, 184].

Specific out-of-place URL components can also be a feature. For example, the presence of two HTTP or HTTPS in the URL [122, 152], presence of TLD in the domain, subdomain or path position, such as ‘cnn.com.malicious.org’ [122, 123, 143, 145, 146, 149, 151, 152, 179, 182], or a prefix (i.e. `www-chase.com`) [59, 151]. Out-of-place brand names can also be features, such as in the subdomain or path [111, 142, 145, 168, 178, 179, 185]. The NoPhish education game [180] added the brand name in the subdomain to help users understand phisher tactics. Similarly, in [168], users learn not to click on long hostnames if they contain part of well-known brand name. Providing correct parsing of URLs can also help users learn to read them [92].

Although phishing education research teaches users not to click on URLs if the domain has unknown terms (unrelated words) [164, 180], it is challenging to identify arbitrary words automatically [174]. Some papers attempt to detect random strings, using methods such as comparing URL tokens to proper or common nouns [111], or by calculating the string’s entropy [151, 158]. Nonetheless, URLs are not necessarily constructed from proper nouns, as is the case with Amazon ‘`www.amazon.com`’ as opposed to the New York Times ‘`www.nytimes.com`’. In addition, Internationalized Domain

Names (IDNA) cannot be detected with these features. Digits in the URL hostname may indicate randomness and are common in the host of phishing URLs (30%) compared to trustworthy URLs (3%) [135, 136]. The number or continuity of digits is also looked for in the host [59, 144, 153, 163, 177] and other URL components [160]. Or even, the continuity of characters such as letters, digits and symbols and the number of each [158].

### **Deviated domains**

Construction of domain names to mimic legitimate ones is another lexical trick. An approach is to replace the TLD with a different TLD [185]. UTF8 encoding can also be used to produce identical-looking characters from different languages and alphabets, such as replacing the English ‘o’ with the German ‘ö’ [35] or using confusing character combinations such as ‘rn’ for ‘m’ [146].

Prior work computed the similarity of domains and pre-computed whitelisted domains [149], the target domains provided by PhishTank [140, 171, 173, 186], Vulnerable Sites List [166], top Alexa domains [126], and dictionary words [163]. In addition, Garera et al. looked for the brand name in the domain concatenating with other characters [179]. [171, 173] looked at if the starting/landing domain appears in, or is a substring of, the title in full or part. Verma and Dyer [145] also used features like Euclidean Distance to find deviated domains.

Training approaches teach users to check for spelling mistakes letter by letter [164, 168, 180]; however, users ability is not adequate to identify visually deceptive domain names [12, 75, 78, 132]. For Unicode, human detection games excluded this feature due to the limitations of humans ability to recognise the subtle differences [164, 180]. Human-support solutions, such as Faheem [92], show that providing the user with assistance by automatically looking for similar popular domains and unexpected Unicode characters can be quite effective.

### **Embedded URLs**

The query string can also contain a request for the destination site to forward the user on to another site. The occurrence of ‘//’ in the query string is used as a feature in automated detection, however, it was not a top 5 feature [152]. Number of domains combined with TLD is also a phishing feature [123, 147]. No human education covered embedded URLs but human-support (TORPEDO) does give redirection information on

mouse over [70].

### 3.4.2 Host features

Querying community managed data sets, such as DNS, or reading HTTP headers, can provide many features; such as domain registration date, where it is hosted and who owns it.

Host-based features increase the overall accuracy [182] as comparisons between lexical, host, and rank features found that host features contribute the most to classification performance [161, 174]. However, connecting to DNS or Whois also requires on average a 1.6 second delay [129] which is time expensive. Phishers can also avoid presenting accurate host features by using link shortners, web hosting services [183], or using compromised accounts so that registrations appear associated with the compromised account owner and not the phisher [187]. Some of these avoidances can be themselves used as features, for example, identifying if the host information is hosting provider or a link shortening service [183].

#### Whois features

Whois is a query protocol that provides 48 features relating to websites [157]. Phishing websites, if they have details at all [130], generally have recent Whois registration dates, near future expiration dates, or recent update date [59, 123, 129, 137, 143, 144, 147–149, 156, 160, 162, 166, 167]. While these three dates are commonly used in research, the domain age is the most commonly used, with a range of definitions for ‘recent’ – 2 [187], 3 [188], or 6 months [152, 189]. Fang et al, [189] found that approximately 95% of the phishing URLs in their dataset were less than six months old and Hao et al. found 55% of domains appeared the day after they were registered [190].

Other record information includes geolocation-based features, such as the time-zone, netspeed [156], physical location of the country/city [114], or the IP geolocation [162, 167, 172]. Also, the existence of the domain in Whois [144, 152], the alignment between the URL domain and the domain registered in Whois [143, 152, 157], the registrars or registrants [129, 142, 166, 167, 172, 182]. Finally, the domain match between Registrar URL and Registrar Whois Server [157]. Some registrars also operate as hosting providers, some of which regularly scan their hosted domains for malicious pages (e.g. GoDaddy [157]) and remove them, while others do not. One feature is the historical reputation of the hosting provider associated with the URL [188].

The Whois based features used in human-support systems are: the server location by Netcraft [191], site country origin and length of registration by CallingID [192].

### **DNS features**

Human-friendly hostnames are converted into IP addresses using DNS, so one common tactic of anti-phishing groups is to remove records of known phishing sites. A missing DNS record is a strong phishing feature [137, 148, 149, 152, 156, 169, 188, 193]. It maintains information that is used as features, such as associated IP addresses (host, mail exchange and name server), Autonomous System number, domain name, sender policy framework, associated BGP and country code [128, 129, 149, 156, 160–162, 174, 178]. IP address segments [154], time to live (TTL) [146, 149, 162, 167], the number of resolved IP addresses [161, 178], number of name server and the number of IPs name servers associated with [178]. Additionally, the ratio of malicious Autonomous System numbers for the resolved IP address and Name servers associated with the resolved IPs [178]. DNS record information is used to ensure that sites are not hosted in a portion of the Internet that is considered disreputable or known for hosting phishing websites [128, 129, 162, 172, 174, 190]. Temporary phishing websites also tend to not have PTR record values [149, 161].

Although DNS provides helpful information, DNS fluxing is used to hide the attacker's identity in an ever-changing network. To avoid Fluxing, Veni et al. look up the domain name of a URL and repeat the DNS lookup after TTL [178]. Notably, while using DNS server records is expensive and may face performance and resource strains [191], requesting the website after TTL period is prohibitive for real time detection. No DNS features appeared in the human-facing papers.

### **Connection features**

Although we exclude full page download features, features regarding the connection to the website can contain useful information about the server, such as the http headers, without requiring a page download. Fields in the HTTP response headers contain information such as HTTP status [147, 155], content-type [155], content-length [155, 178] – negative in some phishing websites, and cookies [130, 155] as some phishing websites store cookies on foreign servers.

Connection speeds [162, 172, 178, 182] are faster on reputable websites, and also domain lookup tends to be quicker as popular websites tend to have a local DNS server.

Human-facing papers did not contain connection features.

### 3.4.3 Rank-based features

Because they are not real websites, phishing sites tend to have a lower visitor count, and are not commonly linked to by other sites, resulting in low popularity and a low PageRank.

#### Domain popularity

Domain popularity is used in several automated detection systems [171, 194]. For example, Alexa's rank is a common feature [111, 138, 140, 142, 146, 149, 151, 161, 171, 195]. Alexa produces this value based on the relative popularity of URLs throughout the previous three months [138], the threshold for legitimate URLs is 150k [157, 169], or 300k [138]. Alexa also provides rank reputation [138, 149]. However, a side-effect of this ranking system is that it is domain based; therefore, URLs from services such as link shorteners, and web hosting websites can still achieve a high popularity [144], shielding malicious agents using these services. Another metric used is webtraffic [137] or the popularity ranking of Netcraft [162].

Looking at human-support systems such as Netcraft and CallingID, we see that they provide site rank based on the hostname popularity [192]. In our opinion, providing the hostname popularity may mitigate the problem of free web hosting domains provided the web host gives each website its own subdomain thereby creating unique hostnames for each site. Yang et al. also designed a warning to tell users about abnormally low website traffic ranks [194]. They argued that including the concept in the warning design reduces the click-through rate in automated detection systems.

#### Page popularity

Roughly, PageRank is a weighted count of how many other pages link to this web page, where the weights are the other pages' PageRanks. Intuitively, it measures how many other well-known pages link to this one.

Google's PageRank is successfully used as a feature in automated detection [111, 123, 137, 138, 149, 157, 161, 179, 179] where URLs that are ranked less than 5 are classified as phishing in [157]. Veni et al. combined the PageRank results from AltaVista, AllTheWeb, Google, Yahoo, and Ask to get a more accurate PageRank [178]. Garera et al. used a number of PageRank features such as, the URL and hostname

PageRank, the page presence in the index of a crawler dataset index [179], while the page presence in Google index was used by [137, 162, 179].

The PageRank is a robust feature since Google updates it frequently; however, it still produces false positives, and is recommended to be used in conjunction with other features [138, 161]. Another problem with the PageRank is link-farming where attackers manipulate the rank by increasing the number of websites that link to the URL [161, 178]. To evade farming problems, Veni et al. [178] added more features such as the number of the different links that link to the page, and whether it is linked to other malicious URLs.

#### **3.4.4 Search engine features**

Search engines optimise for finding the website that the user most wants or expects, given only a small set of keywords. This behaviour makes search engines an excellent proxy for what web site a user might expect to see given some contextual keywords and allows automated systems to identify what website a phishing attack is trying to mimic, because a search for the website title will often bring up the real site [138, 149, 150]. Search engines can also be used to spell check a domain and see if it is a short edit distance from a real one.

Automated detection systems query search engines using the full URL, domain name, page title with domain name, or domain name with TLD [130, 135, 136, 138, 140, 146, 191]. Varshney et al. compared the search results of queries containing domain name, page title, description, and domain name with the page title [191] and found the page title with the domain name gave the highest accuracy. Both page title and domain name can be fetched without needing to load the entire page. A website is considered to not be phishing if it appears in the top 30 [123] or top 10 [130, 146] results in a search for its own URL. In 2014, Basnet et al. compared the results of searching Google, Yahoo and Bing for either the URL or the domain [135, 136]. They found Google to have the highest accuracy; however, they still used all three in their phishing classifiers. While the complexity of querying a search engine is lower than querying the DNS [191], querying three of them in real time may be too time intensive.

Google spelling suggestions are used to detect the similarity between potential phishing domains and popular domains [138, 149, 150]. Another usage of search engines is finding intra-URL relatedness features, the relatedness between domain with TLD and the rest of the words in the URL, are used in [140, 195] using Google Trends

and Yahoo Clues. The only limitation in this feature is the limited number of returned results.

Search engine features are not always consistent and can be different based on the location of the searcher [157]. This issue is particularly problematic for automated systems where the server and the user may not be in the same country.

Human-facing systems tend to advise users that when they are uncertain about a destination, they should search for it and select one of the top results as a way of finding the ‘correct’ website [12,70,92]. Using this advice, users are able to accurately classify URLs [75].

### 3.4.5 Redirection-based features

Redirection can be identified using the HTTP status code, page meta-refresh tag, or JavaScript. The latter occurs with shortened links [114], where URLs with a small number of characters redirect to longer URLs. These are common in Twitter phishing URLs [196].

Automated systems can also follow redirection links providing two features: the initial URL, and the landing URL [197,198]. However, phishers will sometimes cloak URL redirects, by checking for features of the user’s system first, and then deciding which landing page to send them to. Therefore, other features must also be used. For example, the number of different domains and IP addresses in the chain [199]. The number of redirections also indicates phishing URLs [144,171,173,199] and has been used as a phishing feature [148,199]. URLs with  $<2$  redirections were found to be legitimate,  $2 - 4$  were suspicious, and  $>4$  were considered phishing [148]. The similarity between the hostname in the URLs in the redirect chains is also a feature since legitimate URLs typically redirect to same-domain URLs [171,173].

For shortened links, Gupta et al. analysed blacklisted Bitly shortened links [197]. They found several features such as the time between the shortening and the domain creation, or the time between the shortening and using the link. The numbers of redirections has been also shown as a feature in nested shortened URLs, where 80% of the phishing tweets have at least one redirection [196].

Humans are unable to identify redirection prior to clicking without machine support, and even after clicking redirection can happen so fast that a user cannot see it happening. Some human-support systems detect redirection for users and show them the final destination before clicking [70,92].



### 3.4.6 Certificate-based features

SSL/TLS is a protocol commonly used for encrypting web traffic. An initial step of the protocol is for the client's browser to fetch the public key certificate from the server. The certificate is used to validate the public key, which in turn is then used to setup an encrypted connection.

#### Encryption

Early automation work found that legitimate URLs usually supported encryption while phishing URLs generally did not [59, 130, 144, 148, 149, 152, 165, 170, 171, 173, 176, 177]. Since obtaining a valid certificate cost money, it made some sense that legitimate sites would be more likely to have them. However, after the introduction of LetsEncrypt, which provides free certificates to websites, support for encryption is no longer a significant phishing feature as both legitimate and phishing sites now have valid certificates [35, 59].

Similarly, prior advice to end-users was to 'look for the lock icon' which signalled encryption. This is still good security advice [92, 164], and can impact user decision making [70, 75, 77], but it no longer helps detect phishing.

#### Certificates values

Values found in the certificate fields, beyond setting up the encryption, are also used as features. Torroledo et al. used ~40 TLS features to classify URLs, such as the validation level, issuer location, or if it is paid or free [200]. The certificate start and end date are used in [155, 200]. Trusted certificate authorities, such as Comodo, Symantec, GoDaddy, GlobalSign and DigiCert [155, 188] are used to judge the certificate trustworthiness.

Public key certificates can be verified at one of three levels which range from a simple check that the domain is controlled by the certificate requester (domain-validate) [155], to the Extended Validation (EV) certificate which requires the issuer to perform extensive checks of the identity of the organisation the certificate is being given to [200]. EV certificates are effective at proving identity, but they also expensive to obtain, and many sites do not have them [35, 133].

Most modern browsers show EV certificate information to end-users by providing the validated organisation's name in a green box next to the domain. In a lab study, 19% of participants referred to the certificate when deciding safely [75]. However, in

our paper set, all certificate-based features shown to humans required that the page be loaded first except for Netcraft which provide information on request [133].

### 3.4.7 Black/white list features

When a URL is labelled as phishing, it is typically added to a publicly visible blacklist so that other anti-phishing tools can quickly block it [185].

#### Simple list

Lists are used in automated detection as a strong feature due to their low false positive rate – almost 0% for newly observed phishing URLs [114]. Blacklists are also very efficient. It may take humans a while to label the URLs as phishing, but once labelled, computers can easily compare URLs against common blacklists [182, 185], such as PhishTank [157, 184], Google Safe Browsing [146, 184, 197, 201, 202], VirusTotal [201, 202] or Anubis [201], which can be accessed via API or even downloaded locally.

#### Proactive list

Unfortunately, while blacklists are very accurate, they are insufficient to detect all phishing websites [132], likely because blacklists only contain previously seen URLs. Prior work utilised the lists to proactively discover unreported phishing URLs and trustworthy ones [185, 193].

Several approaches have proposed features derived from blacklists, such as marking a URL as malicious if its domain matches malicious domains [135, 201, 203]. Or even if its IP address is on a blacklist [135, 162, 172, 189]. However, blocking a IP addresses runs the risk of inadvertently blocking good sites if the phishing site is using a free hosting service, so features were proposed that compared the URL's domain to a pre-computed list of web hosting services; therefore, Prakash et al. blacklisted an IP address depending on the number of phishing URLs associated with it [185].

Attackers often reuse phishing kits, therefore, a URL could have similar pathname (directory) to previously blacklisted URLs [35, 185, 186, 204]. They also reuse their redirection servers [199], therefore, another feature is to expand shortened URLs before adding to the list [135, 199] and also include URLs in the redirection chain, including HTTP, meta and JavaScript redirection [185, 193], in the blacklists.

Proactive detection of phishing URLs is computationally expensive since it requires storage space in the majority of the features and more time to compare the

downloaded list.

Creating whitelists of validated websites is more complicated. Automated systems typically use high-profile popular sites such as Alexa's top sites [179] or a customised whitelist of domains often targeted in phishing attacks [179]. However, maintaining a comprehensive list of validated URLs is intractable [132]. An alternative solution is to add URLs to a local whitelist after users visit it [194, 205].

Blacklists are heavily used in human-support systems with most modern browsers actively blocking URLs that appear on popular blacklists. Plugins like Netcraft, as a first step of defence, also block reported and verified phishing websites. Similar to CallingID, it also shows the users risk scores on a coloured scale along with the other phishing indicators discussed previously [192]. Human-support systems also leverage user's awareness of phishing indicators and allow them to report phishing URLs for labelling and potential inclusion on blacklists. Cloudmark Anti-Fraud relies solely on users' reports and verification to block phishing URLs [192].

## 3.5 Discussion

The above described research provides ample opportunities to improve human-facing approaches, particularly human-support systems which have the technical ability to leverage many of the features used by automated approaches and provide that data to humans in a meaningful way. Below we discuss some of the more thought-provoking issues.

### 3.5.1 Shifting effectiveness of features

While many automated detection papers discussed the effectiveness (weights) of their features, comparing results is challenging. Effectiveness was evaluated differently across papers, such as statistically by comparing the feature prevalence in benign and phishing URLs [135] or by comparing the classifier accuracy between different groups of features [161]. Feature effectiveness also changes based on the data set and the domain [206]. Link shorteners, for example, are common on social media, but less so in email communication [199]. So, different features work better in different situations.

Effectiveness also changed over time as phishing tactics themselves changed. For example, lexical features like the number of subdomains are subject to both the current website design trends and phisher behaviour. The introduction of free signed certifi-

cates by LetsEncrypt also impacted the lock icon (encryption) phishing feature used by many people.

### **3.5.2 Balancing ‘safe’ and ‘phish’ data sets**

A serious methodology problem we kept running into was the unbalanced selection of data sets to represent safe and phishing URLs, also noted by [177] who points to [173] as being one of the only papers to use balanced data sets.

Papers commonly use repositories like PhishTank or Google Safe Browsing to get phishing URLs. These provide a realistic view of what users are seeing. PhishTank, in particular, provides the full raw URL as it was reported. Finding safe URLs that are representative of a ‘normal’ URL is more challenging. Many papers use repositories such as Alexa’s top sites or Open Directory Project (DMOZ). The problem with these data sets is that they are taken from directory listings and therefore may not represent the composition of a typical safe URL. For example, query strings are rarely included in directory listings, but are very common in say Amazon.com URLs for products. Essentially, the safe and phish datasets tend to be drawn from different distributions which brings the true effectiveness of features, such as length, into question because the source of the difference is unclear.

### **3.5.3 Host-obscuring tactics**

Many features are dependent on the ability to accurately extract the URL’s hostname, such as lexical features and host-based features. Phishers can obscure the hostname by using link shortening services or redirection which hide the final destination URL. These tactics impact the ability of both humans and computers since people can only see the initial URL and computers must take the extra step of resolving the URL to get the final destination before trying to make any predictions. Some research exists on how to detect and resolve these URLs technically, but minimal research looks at how people think about redirects, or how to meaningfully present the information about them.

### **3.5.4 Exploring human-facing features**

Human-education approaches try to help the user learn to extract phishing features on their own and interpret them. But humans take time to learn and teaching them new

things is challenging, so it is vital that human-education approaches require minimal prior knowledge and be robust with a few false positives. For example, long URLs and hyphens are bad human-facing features because they are both found in legitimate and phishing URLs which may confuse users. Conversely, domains are a good choice because how the user interacts with them does not change quickly and they can leverage their knowledge.

We recommend that future work explore the use of automation features in human-support. While many features cannot be understood by humans unaided, there is great potential for human-support solutions to use them.

### **3.6 Conclusion**

With the aim of providing a foundation for future research into human-facing phishing support, we reviewed features used in the automated, human-education and human-support phishing detection systems. In total we reviewed 94 papers and grouped the resulting features into 7 categories including: lexical, host, rank, redirection, certificate, search engines, and black\white lists. We found that all feature categories were used by automated phishing detection, but human-facing approaches have only evaluated some of them.

# Chapter 4

## I Don't Need an Expert! Making URL Phishing Features Human Comprehensible<sup>1</sup>

### 4.1 Introduction

Determining if a URL in a communication is malicious phishing designed to trick users or not is something that even security experts struggle to do without the aid of tools and additional information. As observed in Chapter 3, looking at the URL text is a good first step but fully reading a URL and determining its actual destination is surprisingly complex and often requires the help of third-party services that provide information like how long ago the domain was registered or if the URL redirects anywhere. Yet, despite these complexities, URLs remain one of the stronger indicators of malicious communication [25, 207], particularly if a communication claims to come from an organisation but the URLs lead to other destinations. For example, an email claiming to be from PayPal but containing links to `PayPal-com-security-website.org` is quite likely a malicious email. While the example is simple, it brings up several key issues with detecting malicious links. First off, it requires that the person making the judgement knows PayPal's correct URL and is also able to compare it to the one in the email. The 'correct' URL for a website is not necessarily obvious; for example,

---

<sup>1</sup>This chapter was published in the In Conference on Human Factors In Computing Systems (CHI) in May 2021 [9], and it was a collaboration with Nicole Meng and Kami Vaniea. I was the lead author of the paper, as I conducted the focus groups, the features analysis and users study, whilst the second author helped with polishing the final design of the prototype. I wrote the first draft and collaboratively edited it with the co-authors.

which of the following is the correct website for the New York Times newspaper: `nyt.com` or `newyorktimes.com`? The answer is that both URLs redirect to the real URL `www.newyorktimes.com`. Comparing URLs is also not necessarily easy. End users often confuse elements of a URL, such as the domain and subdomain [8] making comparing URLs error-prone. Experts handle these complexities using a range of tools and information sources that help them make decisions, but end users are often only provided with training on lexical reading [12, 70, 79] and possibly a tool checks if the URL has been confirmed as malicious. In this work, we aim to change this situation by making the types of approaches used by experts more accessible to end-users.

Obviously, users are not the best first option to detect phishing. Automated phishing filters are far less expensive and also relatively accurate [85]. They can quickly compare a URL to lists of known phishing or break the URL into features used to classify it as phishing or not. Most organisations already use automated approaches to protect users on their networks with great success. However, this usage means that any phishing communications a user sees has likely already been through an automated filter and therefore has already been scanned against common computer-friendly features. Assisting users in making these judgements on their own is necessary because automated approaches are not yet 100% accurate [7] and experts are also typically not available to consult on every potential phishing communication in a timely manner. Phishing communication also often uses tactics to pressure the user into responding quickly, such as threatening to shut down their account, charge them money, upset their boss, or lose out on a limited time offer [72, 208]. These time pressures make it emotionally hard for users to report the email and then wait for an official response, leading them to use their own or their peers' judgement [24]. The effects are readily apparent in public phishing reports. Phishing is regularly listed as one of the top causes of data breaches (93%) [66] and the most frequent Internet crimes complaint to the FBI [3]. Financial losses from phishing can also be expensive, exceeding \$29 million in 2017 [2] and \$1.7 billion in 2019 [3]. Tools supporting users in accurately making such decisions on their own are therefore needed. So when providing users with support, we need to think beyond simply telling them if something is phishing or not and instead focus on helping them leverage their contextual knowledge of the situation in conjunction with available data to reach a decision.

When judging the safety of a URL, experts generally have more experience and data sources to draw from, but at the end, they look for discrepancies in the data and their expectations [61]. They can collect the data using tools like Whois (ICANN's

domain lookup) to learn about the registered domain owner or understand the implications of links up-time and popularity. However, using such tools requires an impressive amount of both access to information and knowledge about how to interpret them. Each URL case also requires a slightly different set of information sources and knowledge, causing training users to make such judgements on their own overly burdensome.

Our goal is to support end users so that they can engage in some of the informed reasoning experts currently use when they want to decide on a URL's safety. More specifically, we want to take existing information sources along with knowledge about how to interpret those sources and use them to help end-user decision making. To do so, we started with a grid-based report structure inspired by the Privacy Nutrition Labels work by Kelly et al. [209]. The grid presents the user with information about the URL, drawn from the previous study (Chapter 3), and is annotated with explanations aimed at helping users interpret the information. We then iterated on its design with the assistance of 8 focus groups consisting of end users, security experts, and design experts to simplify the interface and improve the explanation of features. After we created a stable design, we analysed what it would look like on 4640 URLs from two phishing datasets and two safe URL datasets. The goal was to determine if there was any redundant information on the report and also if the features we chose do generally align with the safety state of the URLs. Finally, we ran a user study with 153 Prolific users to determine if they could correctly understand the report contents and make accurate safety judgements.

We found that focus group participants saw how such a report could be useful in cases where they were unsure about a communication. The later focus groups also found the report design useful and informative. Final versions of the report featured only showing relevant information and colours to help users know where to focus. In our analysis of how the report would look with real URLs, we found that for most URLs, the report only needed to show about 7 of the 23 possible information rows, greatly limiting the user's reading burden. The colours also tended to align with the URL being phishing or not. The largest cause of inaccurate red colour being the URL not appearing high in Google searches for the URL, meaning that Google does not associate the URL with the terms it contains. The online participants and focus group participants exhibited similar interpretations of the various report elements, suggesting that our richer focus group data was a good representation of what Prolific users also thought.



## 4.2 Deciding if a URL Goes Where the User Thinks it Goes

Phishing communication often works by convincing the user that the message they received is from a legitimate group they want to interact with. Examples include their bank, their IT department, their email provider, their gaming account, or a lottery site they just won money on. Attackers are also often interested in account credentials, so it is useful for them to mimic existing services to trick users into entering their credentials or providing other sensitive data that the user might normally only give to a trusted party. One side effect of this approach is that the user has strong expectations of where they think they are going when they click on any links. The other side effect is that, except in very rare situations, the attacker does not have access to the company's real URL and instead must set up a fake one, so the URL the user clicks on is owned by someone other than the group the user thinks they are interacting with.

In our review of URL phishing features used by humans and by automated systems (Chapter 3), we observed that the domain part of the URL is the most used feature in human-based detection because they can compare it against their expectations. It is less useful for computers because the computer has to guess if the URL matches the content of the communication. The problem is that while theoretically combining the domain with contextual knowledge should make phishing easy to detect by humans, in practice, people struggle to accurately parse URLs [8], making comparison extra challenging.

### 4.2.1 Mouse over the link and look at the URL

One common piece of advice users are given is to mouse over links in communications and look at where they go [210–212]. This is good advice, especially in cases where the URL is very different from expectations such as an email, supposedly from PayPal, containing a `moonstone235432.net` link. But the advice gets harder to follow if the attacker uses any of a wide range of tricks [26, 123, 179, 182].

A URL is made up of many elements which impact its destination and can be easy for end users to confuse. For phishing, the most important element to look at is the hostname, particularly the domain [12, 78, 79, 81, 132, 133, 213]. This part of the URL controls what server will be contacted to fetch the page, essentially, who controls the page. In order to divert the user to a page they control, the attacker must specify

their own domain and use tricks to make it look legitimate. We provide a summary of the tricks here and refer you to Chapter 3 and other reviews [26, 60, 179] for a more comprehensive overview.

The simplest and oldest approach is using a complicated-looking domain name as the raw IP address, hex or decimals characters instead of the real one [141, 179]. A slightly more advanced approach is to pick a domain that looks visibly similar to the real one but is actually different [123, 179, 182, 214]. Even skilled security experts have difficulties with this kind of deception [53, 69]. For example, in so-called homoglyph attacks English characters are substituted with identical-looking UTF8-encoded characters from different alphabets such as `páypal.com` and `paypal.com` [26, 88, 215]. Another example of a look-alike attack is misspelling (typosquatting). A classic example is substituting characters like ‘vv’ for ‘w’ or capital ‘I’ for lowercase ‘L’ which look identical with a sans-serif font [88, 216]. These two types of look-alike attacks, while dangerous, are very popular and hard to detect by current industry anti-phishing tools [217, 218].

Another trick is to leverage users’ inability to differentiate between URL components [26]. For example, Albakry et al. [8] found that users cannot differentiate between a company name in the subdomain vs the domain of a URL. Similarly, Reynolds et al. [88] found that users struggle to correctly parse URLs, but have high self-confidence in their ability to interpret URLs resulting in a dangerous combination that helps attackers. A common trick involves putting a brand name into an incorrect position, such as in the subdomain (e.g. `amazon.evil.com`), path (e.g. `evil.com/amazon`), search string (e.g. `evil.com?amazon`), or even username (e.g. `amazon@evil.com`). A similar trick is to swap out the top-level domain (TLD) such as `amazon.evil` instead of `amazon.com` [219] or put a fake TLD into a subdomain (`amazon.com.evil.com`).

### 4.2.2 What is the ‘correct’ URL anyway?

The ability to compare a URL in a communication with the ‘correct’ domain for that organisation is a major factor in determining if a URL is legitimate or not. Unfortunately, it is surprisingly hard to determine which domains are associated with a given organisation.

Large companies will typically use a domain name that matches their brand name; e.g. CNN uses `cnn.com` and Chase Bank uses `chase.com`. But some organisations select domains that are not necessarily obvious; for example, Fifth Third Bank has a domain

of `53.com` which relates to its brand but is not obviously the correct URL. Often, companies also have multiple domains associated with their brands such as Microsoft that owns `microsoft.com`, `live.com`, and `xbox.com` some of which are not obviously Microsoft domains from their text. There are several good reasons why a company might have multiple domains such as having multiple product lines or registering ‘defensive domains’ to protect their customers from both typing errors and attackers.

Organisations can also have similar names to other organisations, which makes it challenging for users to know what domain name to compare against. For example, many banks share the name ‘Citizen’s Bank’ resulting in a confusing array of both bank names and URLs. Citizen’s Bank (`citizens-bank.org`) and Citizen’s Bank (`citizenbank.bank`) are two different banks which are not to be confused with Charter One Bank (`citizensbankonline.com`) which is owned by another Citizen’s Bank (`citizensbank.com`). The point here is that it is not trivial to just look at a domain name and associate it with a particular organisation.

Finally, websites that host others’ content can make the situation even more confusing. For example, `windows.net` is owned by Microsoft but is a content hosting site. That is, people pay Microsoft for webspace and then can create websites like `evil.windows.net` which are then hosted from a Microsoft-owned domain [220]. The result is a phishing site that is linked to a real Microsoft domain but where the content of the page is actually controlled by attackers.

### 4.2.3 Redirects and short URLs

While the domain shown in a clicked URL is often the same as the final destination URL, that is not always true. Organisations commonly do minor redirects such as adding ‘www’. Some may also redirect to their preferred brand such as `nyt.com` redirecting to `www.nytimes.com`. More challenging is URLs that obscure the real URL completely, making the URL’s destination impossible to predict without assistance [41, 196]. For example, URL shortening services (e.g. `bit.ly`) [221], QR codes [184], and URL-rewriting by email servers (e.g. `safelinks.protection.outlook.com`). Thankfully, users do seem to be aware that they cannot predict the destination of shortened URLs [8].

## 4.3 Design Goals

There are three large problems that need to be solved: 1) human judgement is needed to determine if a URL is safe because the human has contextual knowledge that is not available to the computer, 2) URLs are made up of a large number of components that are hard to parse correctly and contain information like certificates and redirects that require computer assistance to read, and finally, 3) there are many disparate data repositories that contain data pertinent to URL trustworthiness, e.g. DNS records of registration dates and phishing feed lists of known malicious URLs, which have a wide range of interfaces and locations making them non-trivial to use.

Therefore, as mentioned in Section 4.2, to judge a URL correctly, a large range of URL features is required as predicting the destination of URLs is non-trivial. So, to best assist users in this task, we drew inspiration from the privacy policy nutrition label work by Kelly et al. [209] where a large number of privacy policy elements were put into a food nutrition label like format. We thought that a similar approach might show important URL features to users in a consistent format that might allow them to learn over time. Thus, our goal is to develop a ‘URL nutrition label’, including framing URL information in a way that assists users in leveraging their contextual knowledge and expectations to judge if a given URL belongs to the organisation they expect. We call our design a URL feature report. Our report aims to address the following key design goals:

- A. Comprehensive.** The report should include enough information to help users make an informed decision about the safety of almost all URLs, including the ones in Section 4.2. To avoid overloading users, the interface should also present only necessary information [222, 223].
- B. Support knowledge acquisition.** Each phishing indicator needs to have an explanation that helps non-experts understand the information as well as support higher-level reasoning about it [65].
- C. Promote confidence.** Users need to have confidence in their final decision in order for the report to have its intended impact. Therefore, the report should support users in confidently making decisions on their own rather than blindly trusting recommendations. We aim to support users’ confidence by providing conceptual and procedural knowledge (know-how) when explaining the phishing indicators [224, 225].

- D. Inspire Trust.** The report should inspire users to trust it by regularly providing accurate information and explaining its recommendations in a way that a user can verify themselves. Building trust with users when they need help will also improve their acceptance to taking the help [67, 226].
- E. Support comparisons.** The report should allow users to compare the aspects of the report to their own understanding and, potentially, against reports of other URLs. Supporting comparisons makes it easier for people to use the report for their tasks. The consistent positioning of information also allows them to learn the location of data for faster future access [209]. For example, a user may bank with Skrill and sees on the report that the domain is registered to an address on the Isle of Man, unsure if that is correct or not, they also ask for a report on Skrill's main website to see if it is also registered to the Isle of Man.

## 4.4 Designing the Initial Report

Reading a URL and making an accurate judgement requires accessing a wide variety of URL facts as well as understanding what those facts mean. These facts are consistent between URLs. As part of our design goals, we focus on selecting features that will help people most in making informed decisions about URL safety and how to present them to users. For an initial list, we started with the findings from (Section 3.4). We then narrowed the list down to features that had been shown to be robust and had the potential to be human-friendly. We also excluded features that were highly technical and could not be combined with contextual knowledge to make informed decisions, for example, the DNS-based features. To present the features in a comparable, well-arranged format as required by our design goals A and E, we split our initial design into four sections (Figure 4.1).

### 4.4.1 Notice and reminder

At the top of the report, we show the URL that was asked about for reference. For URLs that redirect, we display both the requested URL and the one that would be redirected to if they clicked the link. We also check the URL against known malicious URLs and clearly state if it is already known to be safe or malicious. PhishTank and Google Safe Browsing both provide lists of reported malicious URLs approved by security communities, which could be used to automatically alert a user [100, 114].

⚠ Think carefully before opening the link since it is not reported as dangerous or safe. Use the information provided below to decide for yourself if the link matches your expectations.

You asked about:  
**https://www.bestchange.ru/exchangers/mkt=en&id=234**

**Facts:**

Domain: This URL is hosted here.	bestchange.ru
⊗ <b>Top search result</b> We searched Google for your URL, the top result is not a match.	http://uniespro.blogspot.com/
⚠ <b>Website popularity ranking</b> How often people go to this website. Well-known organizations should have a rank less than 300 thousand.	More than 1 million - <b>Not popular</b>
⚠ <b>Google PageRank</b> How often this page is linked to by other well known pages	0 out of 10 <b>Low</b>
⚠ <b>Encryption</b> The website supports https so that no one else can read or modify the page.	Basic level encryption <b>Communication will use common encryption approaches but no verification of the owner has been done</b>
⚠ <b>Unverified Owner</b> Organizations can pay to have their ownership verified.	Basic level verification. <b>This organization owns the domain, but no further verification was paid for.</b>
⊗ <b>Website age</b> When the domain was first registered	16-08-2019 <b>Less than a month</b>

**Tricks:**

⚠ <b>domain is similar to a popular organization</b> It is phishing attempt if you expect to go to this domain instead of the domain in the top.	bestchange → bestchange
---	-------------------------

**Understanding this report:**

- ⊗ Known issue
- ⚠ Warning sign
- No issue

**Report Summary**

https://www.bestchange.ru/exchangers/mkt=en&id=234

⚠ We cannot guarantee the safety of danger of this link.

Used Manipulation tricks <b>1</b>	Search Result <b>No Match</b>	Domain Age <b>1 month</b>	Domain Popularity <b>Low</b>
--------------------------------------	----------------------------------	------------------------------	---------------------------------

Color Code: ⊗ Known Issue ⚠ Possible Issue ○ No Issue

**Manipulation Tricks**

Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

Similar to a popular domain: "bestchange.ru" is similar to popular domain "bestchange.ru".	https://www.bestchange.ru/...
---	-------------------------------

**URL Facts**

Facts about the URL to help you compare between what you know with what this URL have.


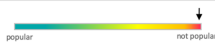
<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	bestchange.ru
<b>Location</b> The physical address where the domain owner claims they live or conduct business.	GDPR masked
<b>Domain Age</b> The date when the domain was first registered.	16-08-2019 <b>1 month</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>does not match</b> your URL: http://uniespro.blogspot.com/

Figure 4.1: Our initial design of the report which was shown to the first focus group (left) and the final design of the report (right).

On the other hand, the Extended Validation Certificate (EV certificate) is used to mark a URL as safe because it indicates that a site's ownership has been verified by a certificate authority, which is sufficient evidence [227]. For other URLs, neither blacklisted nor with a verified owner, the safety is unknown. Thus, we avoid false positives and inspire trust in the report's safety information (Goal D).

#### 4.4.2 Facts

In this section, we provide more details about the website's URL features to help users decide whether this domain indeed belongs to the expected institution or not. Each fact is presented with a fact name in bold on the left, followed by a short description and the value on the right (Goal B). This section has the most consistent structure; however, we only show relevant features to achieve goal A of being comprehensive without overloading the users. Red text is used to both highlight potential issues as well as provide guidance as to what the problem might be. For example, in the initial design, a Google PageRank of 0 is low, suggesting that the page is probably not `apple.com`.

The first and foremost indicative feature is the domain itself. If the user is able to detect that the domain is not what they expected, they are likely to succeed in avoiding

the attack. We adopted the common advice to search for the company's name in Google and look at the top few results. This works well because most modern search engines use popularity as an ordering metric [12].

Two more revealing components are the relative popularity of a website, which we determine using Alexa's most popular domains [75,91,171,179], and the PageRank of a web page [75,179]. These two popularity scales both imply how popular a website is, but we present both to users because they do not always agree, most commonly in web hosting situations. If the domain is a web host, the Alexa popularity is the same for all pages and subdomains under that domain, whereas the PageRank may differ between pages under one domain. Finally, we use the domain age from Whois records in our report since users can efficiently compare it to the expected duration of the organisation's online presence.

Encryption is another hint for safety, indicating if the connection with the server will be encrypted or not (https vs http). HTTPS adds encryption, so users' information remains protected from unauthorised access in transit. Unfortunately, encryption is not a highly reliable indication of phishing websites [35], especially since the introduction of LetsEncrypt [228] which gives free encryption certificates to anyone. It is, however, a useful security aspect.

### 4.4.3 Tricks

There are many ways to manipulate a URL to look legitimate. In this section, we aim at identifying and pointing out these malicious tricks to users. Since the existence of tricks is very indicative of phishing, we check the URL for about 16 different tricks, many of them lexical. For example, we examine the URL for the existence of misspelling by comparing the domain with top targeted domains on PhishTank and Alexa's top 10,000 domains [217]. Each identified trick is then shown to the user as a row under the tricks section along with both an explanation of the trick and evidence (Goal C). To limit the length of the report, only identified tricks are shown, and if a URL has no tricks, we simply state that no tricks were found. Some URLs, such as `https://apple.com`, have no tricks at all, while others, such as the one shown in Figure 4.1, have one or more.

In addition to misspellings, we also look at mixed language use, i.e. the existence of characters from conflicting alphabets. While no longer popular [26], the existence of IP address, hex or decimal characters often indicate phishing URLs. We also reverse

IP addresses to the human-readable domain when possible.

Other tricks used by attackers to mislead users are also specified in our report such as the number of subdomains, ‘@’ in the hostname, and out of position ‘http’, ‘https’, TLD, and ‘www’. We also use the PhishTank’s top targeted brands to identify if a targeted brand name is in the subdomain [144]. Additionally, we consider redirections, including multiple chained redirections. We determine the number of external domains [171], the number of shortened URLs [197] and the blacklisted URLs in that chain [199] and flag them as suspicious if they exceed a threshold. The full list of tricks is shown in Table 4.1.

#### 4.4.4 Severity colours

We use a traffic light system with symbols to draw attention to important information. A coloured circle on the far left indicates how problematic the value is, ranging from green (no issue) to red (known issue), whereas no symbol is provided for neutral information such as the domain. Red indicators are restricted for reliable features with few false positives, such as an IP address in the hostname [12, 70, 92]. At the bottom of the report is a legend explaining the symbols and colours.

The colour thresholds for each feature are adopted from automated detection research with more restriction for red colour as presented in Table 4.1. The first block in the table includes context-related elements such as the domain name, which are mostly presented without colour indicators. The second block shows facts that use different colours while the last block lists the tricks that are displayed in red or yellow when applicable.

We aim to support user’s decision confidence through the use of colour (Goal C). If a user sees many severe (red) colour indicators in a potentially suspicious URL, then it should reinforce their confidence in their decision; conversely, many green rows may help them be confident of a URL’s authenticity.

## 4.5 Iterating with Focus Groups

We conducted a set of eight focus group sessions with Human-Computer Interaction (HCI) experts, security experts, and students from a UK university with a non-technology major. In the first focus group, we took a co-design approach but learned that users rationally just want the system to tell them if it is safe or not, which cannot



Table 4.1: The threshold for the features used in the URL report. ‘-’ means that the row will not be shown in that situation. Features were also added (\*) and removed (\*\*) from the report due to design iteration changes.

	<i>Feature</i>	<i>Red</i>	<i>Yellow</i>	<i>Green</i>
Facts (mostly neutral)	Domain	-	-	-
	Category *	Malicious	Web-host	-
	Registrar Location *	-	-	-
Facts	Domain Popularity	-	< 300K	< 150k
	PageRank	-	0-3	4-10
	Domain Age	< 3 M	< 6 M	≥ 6 M
	In Search Engine	No match	Partial match	Match
	Encryption **	-	Unencrypted	-
Tricks	No. of External Domains	> 4	2-4	-
	No. of Short URLs in Chain	-	> 1	-
	Blacklisted in Chain	>0	-	-
	IP Address	1	-	-
	Non-standard Port	-	1	-
	No. of Subdomains	> 4	3-4	-
	Credential in Host	1	-	-
	Has Unicode ‘%’	1	-	-
	Hex Code in Host	1	-	-
	Non-ASCII	Mixed lang.	Non-ASCII	-
	Out-of-position TLD	A token	-	-
	Out-of-position Protocol	A token	-	-
	Out-of-position ‘www’	A token	A sub-token	-
	Top Targeted in subdomain	A token	A sub-token	-
	Similarity to Top Targeted	-	1	-
	Similarity to Alexa Top 10k	-	1	-

be accurately done. So for later groups, we focused on discussion and feedback. After each focus group, we used the feedback to iterate on the design.

Consequently, each group saw a slightly different version of the interface, starting with the left version in Figure 4.1 for G1 and ending with the right image after G8. As the FG sessions progressed, we saw fewer new suggestions and more discussion of the content and phishing itself, with G7 providing only minimal improvements, suggesting that we were reaching saturation or at least had created a reasonably understandable design. The study complied with our university’s ethics procedure.

### 4.5.1 Participants

Our first three focus groups consisted of experts in HCI (G1) and security (G2, G3) (Table 4.2). The purpose of these groups was to provide expert-level advice and to ensure that our design both matches strong HCI standards and is accurate in terms of security. G1 and G2 were recruited from our University community. G3 was recruited from a local security workshop and contained security experts from industry. All three groups were unpaid and participated primarily out of interest in the project topic.

We recruited non-expert participants from The University of Edinburgh using various email lists, including students from art, psychology, and physics while computer science and informatics were excluded. We chose this group because students are known for falling for these types of attacks meaning that they represent the type of people our report should support. They also rely heavily on the Internet for their studies [229] making them vulnerable to malicious links [107]. They were compensated £10 for 90 minutes.

Table 4.2: Focus groups including their participants' expertise and group size.

<i>Group</i>	<i>Type</i>	<i>Size</i>	<i>Gender</i>
G1	HCI	3	2F, 1M
G2	Security	2	1F, 1M
G3	Security	4	4M
G4	Non-technical	4	4F
G5	Non-technical	4	3F, 1M
G6	Non-technical	5	4F, 1M
G7	Non-technical	5	5F
G8	Non-technical	5	3F, 2M

### 4.5.2 Procedure

We first provided a consent form and collected demographics via a paper survey. In the expert focus groups (G1-G3), we gave a 10-minute presentation on phishing, our motivation for the project, and common URL manipulation tricks. The presentation was provided to ensure all the experts are aware of the context, which allowed us to best leverage their expertise. For average users, we excluded the URL manipulation tricks part of the presentation because we wanted their normal reaction to the reports

without prior knowledge of the tricks. As a warm-up for all groups, we asked participants to share a recent experience with phishing communications, including how they discovered that it was phishing. Doing so helped the participants better conceptualise what ‘phishing’ meant while also providing a set of concrete examples which were often referenced in later discussions.

After the initial discussion, we handed out two sheets of paper: an email containing a URL and the report about the URL. The email was provided so that participants would have the contextual information necessary to use the report. We used real non-malicious emails previously sent to the researchers as a start-point and replaced some of the existing URLs with malicious ones. Participants were told to imagine that they had received the email but were worried about it, so they entered the URL into an online report generator and got the provided report. They were first asked to use the report to decide on their own if the message was real or phishing. Meanwhile, they were encouraged to mark elements of the interface that they found helpful or confusing with provided coloured pens. Participants also had access to a range of co-design style materials, including blank paper, stickers, coloured pens, sticky notes, and scissors. After everyone finished, the researcher moderated a discussion about the report. This process was repeated with another 2-4 email and report combinations, depending on time.

### **4.5.3 Outcomes**

#### **Overall impressions**

Participants generally liked the report, both content and design, and found themselves well supported in making a decision. Initially, they wished for a clear statement whether the URL is safe or not. After we explained that most URLs cannot be definitively classified that way, they tended to understand, but the concept did not come naturally. A G6 member, for example, started with the strong view that safe/unsafe presentation would be best, but after being presented with a URL from a real phishing email sent to most of the University population, he immediately identified it as phishing and recalled that his own anti-phishing tools had failed to identify it at the time. Our report showed him that the URL lead to an organisation located in South Africa, which is not an expected location for a Microsoft URL.

Users had mixed opinions about the interface and its long-term usability. Early groups found the interface overwhelming and very long, but the perception improved

through iteration on the content and presentation. The last few groups found the interface appealing and were even interested in using it either in their daily lives or as a tool when uncertain about a phishing message. They described the report as a useful tool to make a confident decision about the safety of a URL. A member of G5 for example explained its usefulness as ‘Alongside with intuition, there is relevant support and information for me to make decision on whether to trust the website subsequently’. Similarly, a member of G7 explained: ‘I think for most people this would provide enough information to make informed decisions with a high level of confidence. Very interesting’.

They had varied feelings about the trustworthiness of auto-detection tools in general. A G6 participant stated: ‘I trust the machine a lot, but I will trust myself more. This interface will help me to educate myself’. This attitude is not only in line with the goal to support a user’s decision but also typical of phishing training which teaches users to not completely rely on the severity level of the indicators but also encourages them to consider their expectations. Similarly, a participant of G7 said: ‘It would work for helping classify URLs to safe and not safe. It is important to educate users and not just trust the software of taking decisions’. Participants were also able to learn from the report itself as a participant in G7 described: ‘I learned to prioritise the results’.

### **Visual appearance and interaction**

*Symbols and colours.* Over the course of the focus groups, we adjusted the use and prominence of symbols and colours according to feedback. The first group, G1, found the coloured symbols in the left column too small to read and were concerned that they would not be sufficiently obvious to readers and have unclear meaning. We therefore added descriptions such as ‘known issue’ on a solid background colour to make the meaning clear (Figure 4.2). In the final design, we removed these aids altogether and added a legend just below the summary so that it would be visible when needed.

With the new design, G2 was concerned that the colours might be inappropriate for colourblind users and G4 mentioned that different cultures might interpret red and green differently. In Chinese culture, for example, red is considered a happy colour. To handle both issues, we converted to a colourblind friendly pallet and water-marked severity symbols to clarify the meaning [230]. The final report was tested on the iOS grayscale display which produces a colourless version of the report and allows to evaluate how the choice of colours would be for a colourblind person. Later focus groups had mixed opinions about the water-marked severity indicators. Some agreed that the

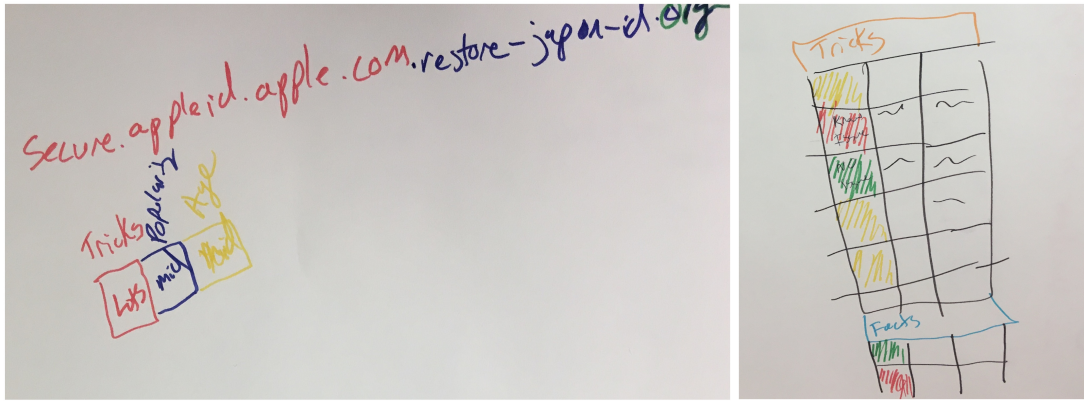


Figure 4.2: Both images were created by a participant in G1. The left image shows a URL with the domain, subdomain, and additional URL components highlighted. Below is the proposed summary with *Tricks*, *Popularity*, and *Age*. In the right image, the suggested new report structure with a list of tricks followed by URL facts.

symbols enhanced the meaning while others found them distracting. Therefore, we removed them from the final version and kept only the symbols in the legend.

*Facts Order.* Focus groups had several suggestions about how to adjust the presentation order of the rows. Members of G1 suggested we weight the presented features and present the most reliable features at the top. G1 also suggested that ordering the facts by colour indicator, beginning with the most severe (red) at the top. Another suggestion was ordering them based on each feature priority. Given that we wanted to present features in a consistent order (Goal E), we decided against these approaches. In addition, the relative value of facts depends on the information only the user knows. For example, popularity is a valuable feature if its value is unexpected, such as if the user thinks the URL is an Apple domain, but the popularity is low.

Another suggestion by G6 was to remove non-critical (green) facts so as to not overwhelm the user as we hide green rows for tricks. However, a G5 participant felt that green facts were easy to ignore if not needed. We decided against hiding green facts because doing so would make it harder to compare reports. It might also incorrectly make all URL reports look overly red and negative, leading to users incorrectly rejecting good URLs.

## Report content

Additionally to the report interface, we iterated over the report components and wording. After each group, we incorporated suggestions to make the report better accessible and understandable.

*Domain and Hostname Highlighting.* At the top of the report, we highlight the URL domain to provide the domain information together with summarising facts. Our initial design did not include domain highlighting since we list the domain as the first fact. However, after moving the manipulation tricks further to the top and adding a summary, the domain is not obvious enough. Therefore, we added domain highlighting at the top similar to how industry tools use it. Using highlighting similar to web browsers also keeps it familiar for non-technical users as a member of G1 suggested that the concepts of domains and subdomains are too technical and lay users are unlikely to understand them. So the highlight in the domain row will help the user learn about the domain aspects.

*Report Summary.* As mentioned before, participants of almost all groups suggested that we provide a clear binary answer of whether the URL is safe or not.

When G1 understood that a binary answer is not possible for all URLs, they instead suggested an overall score or severity bar for URLs that could be easily used for judgement. Similarly, a participant from G3 wanted some sort of classification, such as maliciousness percentages. We felt that a single overall score would mislead users and not encourage them to read and learn from the presented information. Instead, we tried to use a combination of clearly visible colours and added a summary highlighting key issues to the top of the report to help users get the requested high-level sense of safety.

Security groups G2 and G3 saw two versions of the report, one with and one without a summary. Both thought the summary was a good idea and debated about which topics should be included. They liked that the summary told them which features to focus on first. Since both expert groups considered it beneficial, we added a summary section at the top of the report for the remaining groups and continued to iterate on its presentation.

Several iterations later, we settled on four summary boxes: used manipulation tricks, search result, domain age, and domain popularity. *Manipulation tricks* was chosen because their presence is a strong indicator for phishing [27] and the meaning

was clear to most focus groups. The *search results* box indicates whether the URL appears in Google top results when searching for it. Both *domain age* and *popularity* are common features that made sense to users and were generally well understood in focus groups. Groups G2 and later considered the summary to be quite useful. Initially, they felt that such a summary definitely required the rest of the report for explanation. However, after reading the report, they quickly understood the meaning and had no difficulty using it when reading future reports.

*Tricks.* G1 found that the tricks section is very useful, especially the clarification of what is wrong, but the facts section did not adequately explain the meaning of the information. A G1 participant said: ‘If I show the tricks to my gran, she will say yeah cool, but if show her the facts she wouldn’t know what is going on’. The tricks are indeed a stronger indication of a malicious link, potentially eliminating the need to look further [231]. Thus, as suggested by G1 in Figure 4.2, they should appear at the top. Especially since for later groups, key facts such as age already appeared in the report summary.

One of the comments from G6 recommended removing the tricks we found in a verified (safe) URLs because it will distract the users; however, we decided that displaying tricks even for safe URLs will help users to learn to judge the features’ importance for making informed judgements. For example, non-ASCII characters can occur even in safe URLs after the evolution of Internationalised Domain Names (IDNA); thus, based on the context, users can use the feature to judge if a URL is safe or not. This decision supports our design goal B.

*Location and Category.* We added a location field to better support users in identifying any inconsistency between the domain location and the expected location. Usually, the stated physical location of malicious domain registrars differs from legitimate ones [27]. However, understanding the meaning of the location was challenging for focus group members, with some users interpreting location based on the trustworthiness of the country. For example, in G4, one of the participants stated: ‘Apple in Japan, so what? Japan is not questionable’. She was confused and thought that the location referred to the server location rather than the location of the organisation. We thus adjusted the description to clarify that it was the location of the domain owner.

A security expert from G3 commented that one of his common approaches to detecting phishing websites is to look up the URL’s category on FortiGuard which cate-

gorises URLs into groups such as shopping or governmental organisations. This feature can be used to check whether a suspicious page has a similar category as the expected one [232]. Additionally, FortiGuard categorises the full hostname of provided URLs in case a domain includes different subdomains such as WordPress.

*Web Hosting.* As mentioned previously, some popular domains host content for others. As a result, it is possible for the domain to be popular and registered a long time ago, but the specific page or subdomain is malicious. Discrepancy between domain popularity and PageRank should highlight this situation to users, but focus group participants found the discrepancy confusing rather than helpful. So mid-way through the focus groups we started experimenting with wordings suggested by the participants to directly explain the issue. We tried several approaches, including dividing the facts into page and domain facts or creating a large warning on the top of facts. In the final design, to determine which domains automatically offer web hosting services, we used the FortiGuard website categorisation service [233]. Then, we hide the domain-only facts (location, age, and popularity) and in their place we state that: ‘This domain hosts multiple sites, some are good and some may be problematic. Usually only small companies and personal websites are hosted by other domains.’ In general, the focus group participants liked this warning and felt that it was very important and useful for their decision making. G7 felt that they lacked direction on where to look after seeing it. They understood that there might be a problem but they were unsure how to distinguish between safe and malicious hosted sites. Conversely, a G6 participant commented how the warning helped him to be confident visiting personal pages since he expected them to be hosted on other sites.

*Domain popularity and PageRank.* The domain popularity is drawn from Alexa and is an indicator of how often people visit the domain, with the most visited domain being ranked 1. The PageRank roughly indicates how often other pages link to this page [234]. Here, the most linked-to pages have a rank of 10. The two measures are naturally easy to confuse as they both deal with popularity. They also have inverted scales, with 1 being good for domain popularity and bad for PageRank. In the initial design, we tried to explain the difference in words. However, G1 suggested a visual range for the numbers instead to clearly indicate which value is problematic. To further emphasise the scale, we added colours to the range. G2 and G3 saw coloured bars with raw values below and showed no difficulties reading them. G5, however, commented



that the numbers looked like they were written in error due to the opposite directions of the popularity scale numbers. After iterating several approaches on later groups, we settled on removing the numbers entirely and simply showing ‘popular’ and ‘not popular’ at the ends of the ranges.

Differentiating between the domain popularity and PageRank was challenging for all groups except the security ones (G2, G3). Domain popularity made the most sense, likely because it is roughly based on the number of people visiting the site, which is easy to explain and understand. The concept of PageRank, however, was much harder to grasp for participants even when verbally explained. Also, the ‘domain’ versus ‘page’ difference was subtle, leading to difficulties articulating why a page might have different popularity from the domain.

Eliminating one or the other was also not an option as our security groups explicitly mentioned how useful it was to include both since they are fundamentally different measures. For example, sites like WordPress have high domain popularity even if a hosted page has a low page rank. So it is possible for a very popular site to be hosting an unpopular malicious page. To reduce confusion, we iterated on the wording to improve section explanations. G8, in particular, was shown several wording options and provided extensive feedback on how to express the concepts more clearly. However, even in the final design, the difference between domain popularity and PageRank is still hard to grasp quickly.

*Encryption.* Initially, we thought that encryption would be useful information in the report. In the encryption component, we stated if the connection was encrypted or not. But we were also concerned that a user would equate encryption with owner validity, so we added, ‘This URL is encrypted, but we couldn’t verify the owner’. G1 understood this concept and explained: ‘If this is an Apple URL, you would expect to have a verified owner’.

However, we found that showing encryption information may mislead users. For example, a G2 participant marked a legitimate URL as phishing because she did not think a reputable company would use a HTTP connection. When we tried incorporating ownership information as well to provide a more complete view of encryption, participants just became more confused. A participant from G4 asked, ‘why is it a good sign if you cannot verify the owner of the organisation?’ after seeing that the connection was encrypted (green) but the website owner could not be verified due to the information not being in the SSL/TLS certificate. Showing information that could

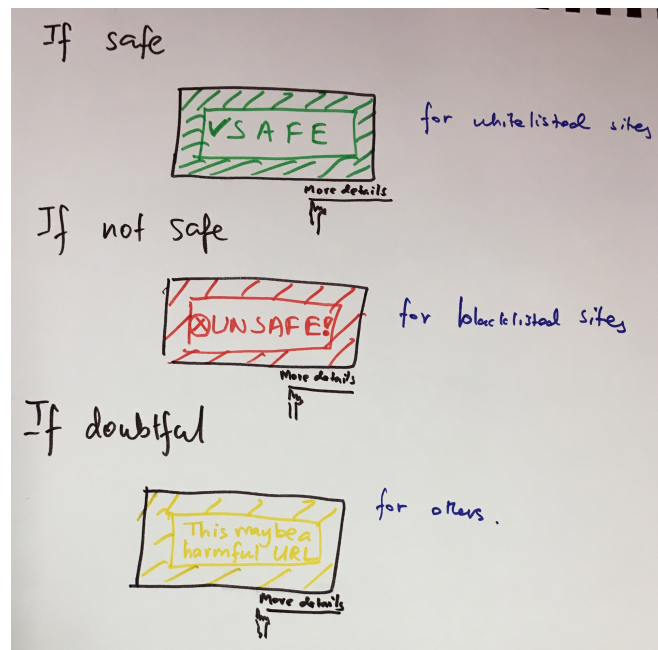


Figure 4.3: A G6 participant began designing a report showing 'safe', 'unsafe', 'doubtful' and adding an option for 'more details'.

mislead the user violates our design goals D of inspiring trust and C of confidence by showing correct information that will lead to the right decision. Therefore, in line with our design goal A to avoid overload, we removed encryption entirely from the report. The only exception is Extended Validation (EV) Certificates. These are TLS/SSL certificates that have gone through an extensive ownership verification process. When we encounter an EV certificate, we put a green check in the summary and a statement like: 'MoneyGram International Inc. verified its ownership of this domain.' The focus groups liked this feature and found it helpful. A member of G6 stated, 'that's all I need'.

### Ideas for workflow integration

Participants were enthusiastic about having easy access to the report and explained how they would integrate it into their work flows. A G2 participant suggested developing a browser plugin or email client, which shows an option to 'Look up this URL' when a user right-clicks on a URL and would show the report in a new tab. They explained that a plugin would be best for them as they thought people are 'too lazy' to visit a website. A G5 participant suggested providing the summary for each link on a page automatically.

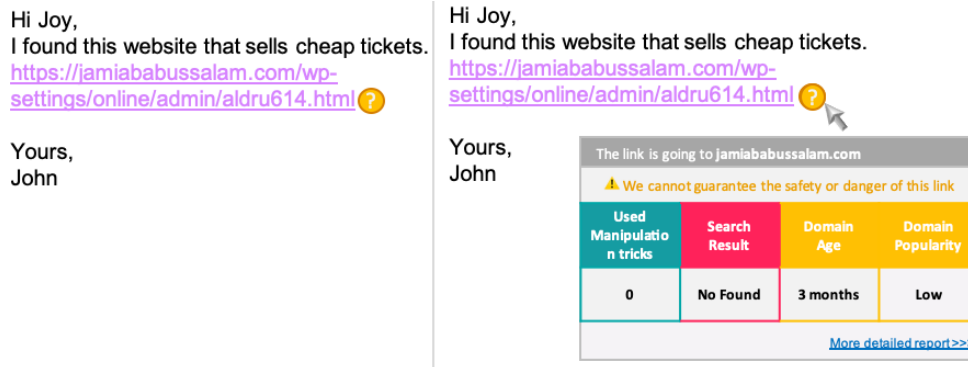


Figure 4.4: An example of how the report could be integrated into a user's workflow as suggested by G3 and G6. The image shows a small circle flagging the URL and a version of the report summary when hovering over the symbol with the option of opening the full report.

Both G3 and G6 suggested a tiered design where initially only limited data like a score or the summary is shown with an option to view the full report (Figure 4.4); possibly limiting the type and detail of data depending on if the user is novice or advanced, e.g. an IT helpdesk employee. In G6, a participant suggested that they would like to see a high-level safety flag before seeing any reports (Figure 4.3): 'Give me three flags (green-safe, red-blacklisted, yellow-unsure), then give me the ability to drill down the summary and if I want more details, give me a link to the web page'. Adding the high level estimate will not 'overwhelm [them] with the details at a starting point' when using everyday. This suggestion contributed to evaluating only the 'report summary' in later sections to validate its effectiveness.

#### 4.5.4 Expert interview

We showed an early version of the report to two experts who are designing anti-phishing training for the University and responding to requests regarding phishing from the front-line helpdesk. Overall, they were very positive about the report design. Although they felt that the report is too complex for average users, they thought that it might be of great use to the help desk staff who has to handle reports of potential phishing URLs. We also asked them if there was any part of the report they felt was unnecessary. They commented that the encryption component, which we later removed, is not needed to judge if a URL was phishing and will likely confuse users. Otherwise, they felt that all other parts of the report were necessary to make an informed decision.

## 4.6 Features Validation

While the features selected for the report are known to be strong phishing features, we still wanted to test the visual appearance of the report on known phishing and safe URLs. We analysed 6877 URLs from four data sets, two phishing (PhishTank [235], OpenPhish [236]), and two safe (DMOZ [237], ParaCrawl [238]), to explore what the report could realistically look like for users. The sets of safe URLs contained 2615 URLs, of which 592 (23%) were excluded due to ‘4xx’ and ‘5xx’ response codes. For phishing URLs, we collected a total of 4262 URLs, of which 1645 (39%) were excluded due to unsuccessful response codes. Phishing URLs from OpenPhish and PhishTank were processed every two hours to extract features while the pages were still live.

To reduce the load on readers, reports only include data relevant to the particular URL. For example, tricks do not appear in every URL, so we hide them by default and only show them if they are present such as the presence of non-ASCII characters. So while there are 23 possible rows, only 6.7 rows were shown on average (*Min* = 4, *Max* = 10), with phishing (*Mean* = 6.8) and safe (*Mean* = 6.5) having similar row counts.

Tricks were rare for safe URLs, with only 77 (3.8%) showing one trick and the remainder having no tricks. Phishing URLs more commonly had tricks with 868 (33.2%) containing between 1–3 tricks. The primary cause of tricks for safe URLs was the similarity between the domain and one of the top 10,000 popular domains (30 URLs), a phenomenon already seen in previous research [214]. Thus, we only show a yellow indicator for this feature to avoid false positives and hand the task of comparing and deciding whether this is expected in their context to the user.

Looking at the red colour in the reports, 30.2% of safe URLs and 88.0% of phishing URLs had at least one red row. Of the safe URLs with a red row, 99% did not appear in Google’s top 10 search results, causing the red. Search results may therefore vary a good bit, making it a difficult feature to interpret. However, it is still a good indicator of the illegitimacy of a page, which is why we decided to keep it. We also found that 23.4% of phishing URL reports had only green rows. Further examination showed that only three of them were compromised websites, while the rest were URLs that redirected to safe URLs; thus, the features referred to safe URLs, which were indeed safe. This redirection tactic is used by attackers to serve advertisements and then send the user to the expected safe website [217]. Therefore, it is important for the report to urge users to visit the shown final link instead of the original one. We also considered

using the colour frequency in each report to predict whether a URL is safe or not. Applying linear regression, we found that the frequency of each colour in the report significantly predicts whether a URL is safe or not ( $R^2 = 0.44$ ,  $F(4635) = 1238$ ,  $p < .001$ ) with Red ( $\beta = -0.48$ ,  $p < .001$ ), Yellow ( $\beta = -0.17$ ,  $p < .001$ ), and Green ( $\beta = 0.11$ ,  $p < .001$ ).

Finally, we measured the features' redundancy to ensure we are not showing unnecessary features. We computed pair-wise correlations between features using their severity colour as presented to users as the feature's value and found no correlation between any of them.

## 4.7 Online Study

Focus groups are an excellent way to get rich feedback but a poor way to get a truly wide range of participants. To address that gap, we decided to use an online survey to test the clarity of report content as well as its ability to support users in making accurate safety judgements about a URL. We used a between-subjects experimental design where each participant saw one of: the full report, just the summary, and just showing the URL with domain highlighting. The full survey is attached in Appendix A.

### 4.7.1 Questionnaire instrument

For all conditions, the survey started with informed consent. Participants were then asked how familiar they were with 13 website terms and 6 companies, followed by study instructions to not visit any of the links and only read them. A question then tested whether they had read the instructions and terminated the study if they failed it twice. To test their existing URL-reading skills, participants were then given three URLs and asked to choose which company those URLs lead to. The first URL has Google in the pathname, the second has Facebook in the subdomain part, and the third is for New York Times, which uses an abbreviation of the brand name.

Participants were then shown 6 URLs. For each URL, they were told to imagine that they wanted to visit a particular company, given a brief description of that company, for example, 'eBay, an auction and consumer to consumer sales website', and then asked if the URL 'leads to a page owned by the above company or is it a malicious URL'. In the domain highlighting condition, the URL was domain highlighted in the question. For the other conditions a report was provided, and participants were

encouraged to use it when answering. Participants were then asked how confident they were in their decision, followed by a question about what most influenced their decision. After answering questions about all 6 URLs, all groups were asked a set of comprehension questions to make sure they did, or could, understand the content of the report or the report summary. The comprehension questions were multiple choice that ask the participant what they thought the different parts of the report meant. Full and control were asked about the full report elements, and the report summary group was asked about the report summary elements. The answer options were drawn from common misunderstandings observed in the focus groups. The survey ended by collecting background information. We used a phishing susceptibility scale from Wright & Marett [239] with 5 sub-scales to test participants' computer self-efficacy, web experience, trust, risk beliefs, and suspicion of humanity. Also, we included basic demographics questions on age, gender, and highest degree obtained.

### **Study conditions**

In the following, we describe the conditions and the questions that differed between them.

*Domain highlighting.* In this condition, we showed the full URL with the domain highlighted and asked participants if the URL leads to the given company name. Existing research already shows that users cannot read URLs unaided [8,75,88,92]. Domain highlighting has already been adopted by several browsers, e.g. Safari, making it a state-of-the-art approach that has been shown to help users decide on URL safety [90]. Thus, we chose domain highlighting as the control condition. To determine what most influenced participants' safety judgements, they were asked to select up to three of: the domain, the protocol (https), the URL path and query strings, their prior knowledge, and their familiarity of the company's URL.

*Full report.* This condition is the longest. Before showing the 6 URLs to participants, we first showed them a fictitious report with obviously fake values and asked 10 questions in a random order about the features, i.e., 'How old is this website?'. Doing so gave participants some basic practice with the report and allowed us to test for any serious misunderstandings about how to use it.

To determine what most influenced the participants' safety judgements, they were shown a list of the different report elements along with 'my own prior experience

reading URLs' and asked to select up to three that most influenced their decision.

Finally, we asked a set of 7-point Likert assertion questions to measure the report usefulness and satisfaction, loosely based on the SUS, such as "I can learn a lot about phishing using this report" and others drawn from focus group participants' opinions about the report. We ended with an optional free text comment section.

*Report summary.* Many participants in our focus groups suggested showing the report summary when a user hovers over a link. The idea has merit, so we evaluate it here as a middle option between domain highlighting and showing the full report. Participants in this condition saw only the summary part of the report, with no option to see the full report.

To determine what influenced the participants' safety judgements most, they were shown a list of answers, including the summary report boxes, elements of the URL, their own prior experience, and the colours. After answering questions for all 6 URLs, they were asked about the meaning of each of the summary report elements, with multiple choice answer options derived from common focus group misconceptions and an 'other' option. Finally, they were asked the same 7-point Likert questions about the report usability as the Full Report condition.

Table 4.3: Summary of the URLs used in the online study to judge URLs.

<i>URL</i>	<i>Hardness</i>	<i>Popularity</i>	<i>Safety</i>	<i>Group</i>		<i>% of participants accurately judged safety</i>		
				<i>G1</i>	<i>G2</i>	<i>Highlight</i>	<i>Full report</i>	<i>Summary</i>
<a href="https://resolutioncenter.ebay.com/policies/?id=123">https://resolutioncenter.ebay.com/policies/?id=123</a>	Parse and match	Popular	Safe	X	X	81	100	81
<a href="https://ftmurl.com/www.ebay.co.uk/item=3032759652">https://ftmurl.com/www.ebay.co.uk/item=3032759652</a>			Phish	X		96	96	88
<a href="https://www.bestchange.ru/exchangers/mkt=en&amp;id=234">https://www.bestchange.ru/exchangers/mkt=en&amp;id=234</a>		Unpopular	Safe	X		44	88	83
<a href="https://www.bestchange.ru/exchangers/mkt=en&amp;id=234">https://www.bestchange.ru/exchangers/mkt=en&amp;id=234</a>			Phish	X		92	100	92
<a href="https://email.microsoftonline.com/login/?mkt=en-GB">https://email.microsoftonline.com/login/?mkt=en-GB</a>	Domain knowing	Popular	Safe	X		64	73	50
<a href="https://www.365onmicrosoft.com/login/?langua=en-GB">https://www.365onmicrosoft.com/login/?langua=en-GB</a>			Phish	X		73	100	96
<a href="https://international.bitrex.com/account/?id=2423">https://international.bitrex.com/account/?id=2423</a>		Unpopular	Safe	X		73	85	65
<a href="https://international.bitrex.com/account/?id=2423">https://international.bitrex.com/account/?id=2423</a>			Phish	X		56	96	92
<a href="https://fb.me/messages/t/788720331154519">https://fb.me/messages/t/788720331154519</a>	Misleading flags	Popular	Safe	X		15	92	85
<a href="https://l.facebook.com/l.php?u=http%3A%2F%2F67.23.238.165">https://l.facebook.com/l.php?u=http%3A%2F%2F67.23.238.165</a>			Phish	X		60	100	83
<a href="https://www.tripod.lycos.com/pricing/?plan=free-ad">https://www.tripod.lycos.com/pricing/?plan=free-ad</a>		Unpopular	Safe	X		56	92	96
<a href="https://webmasterq.tripod.com/pricing?plan=free-ad">https://webmasterq.tripod.com/pricing?plan=free-ad</a>			Phish	X		65	77	81



We categorised URLs into three reading difficulties levels: (1) Parse and Match: any URL knowledgeable person can find the domain and compare it to brand name, (2) Domain Knowledge: a URL knowledgeable person has to know which organisation a domain belongs to before judging the URL, and (3) Misleading Flags: URLs have information that may mislead participants to misjudge them. For each category, we have two organisations, one popular and one not popular based on the top targeted domains on PhishTank, and for each organisation, we have a phishing and a safe URL (see Table 4.3). With 6 organisations, we ended up with 12 URLs in total. For each condition, participants were divided into two groups, with every group being shown one link of each organisation at random, six URLs in total.

The presented URLs are real-life URLs with the phishing taken from our analysed data set in Section 4.6. We made minor manipulations to control some variables. They have an approximately similar length, https protocol, as well as path and query strings. To reduce a bias in the selected URLs, we ensured that the colour indicators were in line with real observed colour combinations from the data set. As we had abnormal false positive search results, we included one safe URL with a red search result. For participants' safety, we selected phishing URLs that were no longer active. Additionally, in case they clicked on the links, we added a hyperlink that leads to a page belonging to the research group about the danger of clicking on these links.

## 4.7.2 Survey results

### Participants

We recruited participants from Prolific for a 30-minute study on phishing. The time estimate was based on a short pilot study. We limited participants to those with approval rates above 90% and Native English speakers to avoid language issues. We then excluded those who did not answer the attention check questions accurately.

We had a total of 153 participants (*domain highlighting* = 51, *report summary* = 50, *full report* = 52), 63.4% were female. Participants had an average age of 31.89 years ( $\sigma = 9.9$ ). Compensation was £3.5. The average time required to complete the survey was 18.26 minutes. For the prior URL reading skill, on average 1.6 of the questions were answered correctly with only 14 answering all questions correctly (9%) and (15%) not answering any URL correctly.

### Accuracy of safety judgement

We found that participants in general were able to accurately judge URLs' safety. The average accuracy was highest for the full report (5.5/6, SD = .28), with the report summary also doing well (4.96/6, SD = .38) and the domain-highlight doing the worst (3.88/6, SD = .48). The false positive (FPR) and false negative (FNR) rates are also encouraging with all participants more likely to incorrectly mark safe URLs as phishing: *full report* (FPR = .12, FNR = .05), *report-summary* (FPR = .23, FNR = .11), and *domain-highlight* (FPR = .44, FNR = .26).

We used an ANOVA followed by applying Cohen's  $F$  for the effect size to test if the three conditions (domain-highlight, full report, and report summary) impacted the accuracy of participants' judgements and we found a statistically significant impact of the condition on the judgement accuracy ( $\alpha = .01$ ,  $p < .001$ ,  $r = 0.29$ ). We then computed follow-up t-tests and found a significant difference between all three pairs of conditions: domain-highlight and full report ( $p < 0.0001$ ,  $d = 0.7$ ), domain-highlight and report-summary ( $p < 0.0001$ ,  $d = 0.4$ ), and report-summary and full report ( $p < 0.001$ ,  $d = 0.3$ ).

We separately tested if any other variables impacted accuracy using ANOVA as well. These variables are the time spent on the question, the condition, and the level of difficulty of each URL, the actual safety (malicious/trustworthy), participants' confidence in their answer, their familiarity with the company, the company the URL leads to, their prior knowledge of URL reading, and their phishing susceptibility factors. We found that the accuracy of users' judgements is significantly impacted by the condition, the URL safety, and the URL hardness level ( $\alpha = .01$ ,  $p < .001$ ), with a large effect size for the condition ( $r = 0.31$ ) and small for the other two (0.16 and 0.13). The remaining variables had no significant impact on the judgement accuracy.

We tested URLs associated with 6 organisations as shown in Table 4.3 where each organisation had a phishing and safe URL associated with it, resulting in 12 URLs tested. For all URLs, the full report has higher accuracy than the domain highlighting. The summary report is slightly mixed, mostly sitting between the domain highlighting and the full report, but occasionally showing more accuracy than the full report. The four 'parse and match' URLs are theoretically the easiest to determine from only reading the URL string, which is mostly born out with the high accuracy for even the domain highlighting condition. The exception, `bestchange.ru`, was incorrectly marked as phishing by the majority of participants in the domain-highlighting condition. For

the ‘domain knowing’ URLs, a user has to know the correct domain of the organisation to be able to accurately judge safety if unaided. Here the `email.microsoftonline.com` URL was the most challenging for all conditions. The `bìttrêx.com` URL was also challenging for the domain highlighting group, possibly because they were unsure if the non-ASCII character (Vietnamese) should be there or not. In the misleading URLs, Tripod was a confusing case where the safe URL positions the brand name in the sub-domain while the phishing URL includes the brand name in the domain but actually is a hosting service for other websites. Similarly, the `fb.me` legitimate short URL confused many of the domain-highlight participants.

### **Comprehension of the report elements.**

Full report participants were able to provide a correct answer for 7.73 out of 10 report comprehension questions on average. The most common error was in regards to the location feature where 57.7% (30/52) of participants indicated that the location means the physical location of the server they were contacting rather than the self-reported location of the organisation that registered the domain.

PageRank continued to be a source of confusion with 34.62% (18/52) of participants providing an incorrect answer. They commonly confused PageRank with domain popularity indicating that the value meant how popular the site was rather than the individual page. One option we are considering for future work is to hide PageRank when it is in alignment with the popularity (both high or both low) and only show it when it is different with direct explanations of how the miss-alignment could be problematic.

All questions in the section had an option to indicate that the description was confusing. The web hosting element confused participants most with 12% indicating that the description is unclear and 73% (38/52) of them answering it correctly. The result suggests that the new wording is mostly working though there is some room for improvement.

In the summary group, participants were able to provide a correct answer for an average of 4.69 out of 6 questions. The most common error concerned the web hosting feature with 57.7% (38/52) answering correctly.

### **Report usefulness and satisfaction**

We asked participants to pick the report elements that they found most helpful after deciding about each URL to get a sense of if they were relying on a small set of fea-

tures or using the whole report. Participants chose different information for different URLs. ‘Domain age’ was the most influential feature for 4/12 (eBay and Bittrex Safe URLs and eBay and Microsoft Phish URLs), ‘Manipulation tricks’ for 3/12 (Facebook, Bittrex, and BestChange phishing URLs), ‘Domain popularity’ 4/12 (BestChange, Microsoft, Tripod and Facebook safe URLs), and ‘Search result’ for 1/12 (Tripod Phishing URL). For the safe Microsoft URL, we displayed a warning that ‘microsoftonline’ is similar to ‘Microsoft’ and it does not match Google’s top search results, however, it was still not the most helpful feature. The fact that participants were looking at different elements for different URLs shows that they were making use of the full report and balancing and weighing features, instead of just sticking to the one aspect that made the most sense to them. Participants indicated that they used prior knowledge but it was not in the top three features for any of the URLs.

The self-reported answers for satisfaction indicate that the full report ( $Mean = 5.78$ ,  $Median = 6$ ,  $SD = 0.72$ ) was more preferable than the summary-report ( $Mean = 5.36$ ,  $Median = 5.57$ ,  $SD = 1.13$ ). For the full report, participants found that the survey taught them about phishing, the report would help them recognise phishing URLs, they understood the report content, and they did not need to learn new skills to use it. However, in the summary group, users felt they needed to learn a lot of things before using the report.

## 4.8 Limitations

Our report aims to support users in deciding if potential phishing URLs are or are not safe to click on. Therefore, our work is limited to the types of information available to a user in advance of loading the page itself and does not include solutions that look at the safety of the resulting page such as identifying compromised code or layouts that are visually similar to frequently targeted sites.

We endeavoured to put together focus groups looking at HCI, Security, and non-technical students to get a range of opinions and experience. We also conducted multiple focus groups to offset some of their known issues, such as participants getting distracted by irrelevant topics, or being influenced by a dominant peers’ opinions. We also ensured that a moderator was present to keep the groups focused and on topic. Finally, we also used an online survey to further verify our focus group findings on a large scale.

However, Prolific, similar to other online micro work sites, is known to have users

who are more computer literate than the average internet user, they also tend to be more privacy aware [240] which may impact their knowledge of URL reading. Though, recent studies of online workers suggest that online workers, including Prolific workers, still struggle with predicting where a URL will go [8, 88]. This type of user is also the type of person that less skilled people may go to for assistance [24], so supporting them well is likely to have a broader positive impact.

The phishing feature list we used is also not exhaustive. There are a wide variety of features used to detect phishing URLs, and many of them appear in only one or two papers. To mitigate the issues, we made use of existing reviews of the range and accuracy of features. However, features that have been mostly tested on automated systems are not necessarily the best features possible for people. In this work we have started with known robust features and narrowed in on those that best support people, but it is possible that other features exist that support people better but did not show up in our review.

## 4.9 Discussion

URLs are known to be complicated for users to read unaided making it challenging for them to accurately judge the safety of a URL, even when they are aware of context like what website they expect to visit. Our report design is intended to support users in making informed decisions about URLs that they are concerned about. While many users have access to some form of expert advice, either through their employer's help desk or through the help mechanisms of the targeted company, that expert advice takes effort to engage with and the response will likely be slow if it comes at all. Our report is intended to help users help themselves when they encounter a URL they consider suspicious by allowing them to make use of many of the same data sources used by experts.

Our focus group participants were able to use the report by utilising their own contextual knowledge and their expectations of the organisation the URL represents. For every URL, they picked a different feature that influenced their answers, giving them the flexibility to decide what phishing feature is the right clue for each case. Our online study showed similar findings with both the full report and summary report conditions able to more accurately identify phishing URLs than users who only had domain highlighting to help them. Online participants also made use of many elements of the report to make their decisions, supporting the view of our expert focus group

participants that a large number of features is needed to accurately judge URL safety.

### 4.9.1 User empowerment

In security work it is easy to take a paternalistic approach with end users where the security expert knows best, gives users minimally explained rules to follow, then gets upset or blames them when those rules are not followed as expected. Part of the goal of this work is to shift that interaction to one where the users are given more knowledge about the specific situation as well as more control while also being asked to remember less facts and rules. Ideally the report structure will also support user confidence and fast feedback where if they think a URL is potentially unsafe, they can quickly gain more information about it. While tools that support users in making decisions about safety do exist, there are few of them and they are mainly aimed at expert users who already know terms like ‘domain’, most of the other tools instead focus on providing users with binary decisions with minimal to no reasoning provided [241]. While the binary advice is what users would like, having too many false positives also leads to users no longer trusting tools [76], so such tools have to be careful about which URLs they mark as unsafe.

Empowered users are also important because the context each person works in is different making it nearly impossible to provide one set of comprehensive rules that work well for everyone. The behaviours of phishers also adapt and change over time [26]. Asking users to keep all this information in their head is impossible [226], but with the support of a tool, users can always be basing their decision on up-to-date information and guidance. Over time they may also start learning the tricks and indicators that are most useful for the type of content they see.

### 4.9.2 Training

While the primary purpose of the report is decision support, it also has a potential for education. Existing education approaches tend to focus on training the user either through a dedicated up-front training [224, 242] or through smaller training embedded in existing work practices. In both cases, a security professional decides what is most important for the user to learn and bases their training around that. The timing of the training is also outside the user’s control, either dictated by the organisation or appearing in their normal work unasked for. The design of our report is intended to fill a gap where instead of telling users what and when they should learn about phishing,

we instead wait till they have a specific case that they would like to learn more about. Similar to earlier work by Kumaraguru et al. [243], this type of training is timed at a *teachable moment*. But where earlier work has focused on the moment after the user fell for a phishing attack, our report instead focuses on a time point when the user is curious and seeking advice.

### 4.9.3 Use cases

The report is meant to support a user who has already identified a potentially fraudulent URL, but is uncertain about it and either does not have access to experts or does not have time to wait for expert feedback. Our focus groups, however, also suggested other uses cases, such as using it to help explain a safe/unsafe decision to someone else. People often reach out to peers when uncertain about potentially malicious communications [24, 244], having a report like ours would enable people to provide not only a recommendation but also a reason behind that recommendation. Similarly, participants thought that the report might be useful for help desk workers who may not be security experts, but are regularly asked about potentially fraudulent communications. Such a report might improve their ability to respond to requests accurately as well as provide useful feedback to users. Finally, they suggested adding this type of report to automated systems to better persuade users to adhere to the warning [67]. Several of these use cases would be interesting to study in future work.

### 4.9.4 Focused attention

Initially we set out to create a short and simple report that users could easily use at a glance to understand the safety of URLs. One hard lesson we have learned over the course of this project is that URL safety is a complex topic. Many of the basic concepts needed to understand the evidence require explanation for an end user to be able to understand them and use the knowledge correctly. Because the URLs being looked at are expected to have already gone through a phishing filter, what the user needs the most help with is comparing their own contextual knowledge, which was not available to the automatic filter, to the information about the URL. This type of comparison is necessary at this stage in the process and requires focused attention from the user to accomplish the task. Our report is designed to support users in this process by clearly explaining the different elements, supporting comparisons, and enabling users to more efficiently use the report on subsequent accesses. However, it is important to

recognise that the report is best used in cases where the user is proactively seeking more knowledge about a URL rather than pushing the report into their workflow unasked for.

#### **4.9.5 Deployment potential**

While the main contribution is the report design itself, we also consider the practicalities of potentially implementing and deploying it. In order to test the report content with real data (Section 4.6), we programmatically sourced the feature data and computed what would be displayed in the report. Our code automatically queries several third party APIs, such as PhishTank and Google Safe Browsing, to retrieve features not possible to extract from the URL itself. Using such APIs in a deployed system would be practical since they are continuously updated and, thus, require minimal to no ongoing maintenance for the report accuracy. Other report elements like the tricks and the explanations may require more expert involvement to maintain over time, but the effort of doing so is not large. To further explore the potential of deployment, a masters student build a prototype of the report as a Chrome browser plugin as a thesis project [245]. Their system allowed a user to request a variation of our report for any URL they saw inside the browser. As the prototype was a proof-of-concept, it is not suitable for real-world testing. But it did demonstrate the feasibility of integrating such a report into common user tools, like browsers.

### **4.10 Conclusion**

We have presented the design for a new URL feature report which assists users in deciding whether a URL is malicious or not. The reports are intended for users who are trying to judge a URL's safety as part of a primary task. To refine the report's design, we conducted 8 focus groups with experts in HCI, experts in security, and average users. Finally, we conducted a survey to measure the readability and effectiveness of the report. We found that participants could generally read the reports, understand phishing features, and use them to successfully decide if a URL is malicious or safe. However, some participants still had difficulty understanding the more complex concepts such as PageRank and location.





# Chapter 5

## A Case Study of Phishing Incident Response in an Educational Organisation<sup>1</sup>

### 5.1 Introduction

Keeping organisations secure requires effective procedures to handle reports of fraudulent emails *phishing*. Phishing attacks are often used to gain access to accounts and other information that is then used in more damaging attacks [107]. Protecting employees from such attacks is a key component of most large organisations' security plans, often including training employees on how to identify and report phishing as well as putting in place internal procedures to quickly respond to phishing reports.

Phishing is by far the most common and disruptive type of attack for UK organisations [1,246] which can be partially seen in the amount of effort they put into managing it. Worldwide, 93% of organisations measure what phishing is costing them including downtime, monetary losses, reputational damage, and time spent by their IT teams on remediation [247]. Organisations also spent considerable resource training end users to identify phishing attacks with 98% allocating more than 30 minutes to train each user and 61% training users at least every month [247]. The effort is showing results,

---

<sup>1</sup>This chapter was submitted to the ACM Conference on Computer-Supported Cooperative Work (CSCW) in 2021. One of the sections about the phishing life cycle was moved to chapter 2 in order to fit into the thesis structure. This work was a collaboration with Adam Jenkins and Kami Vaniea, with myself as the lead author of the paper. I conducted all the observations and interviews. Adam Jenkins was the second coder and ensured an objective and transparent of the data analysis and reporting. He additionally contributed to the discussions concerning how to best interpret the results. I wrote the first draft and collaboratively edited it with the co-authors.

with 63% of highly disruptive breaches being reported by staff as opposed to found through automated monitoring [246]. The impact of on-the-job phishing training can even be seen in general user surveys where users report learning protection practices at work or learning from others who had training at work [248].

How organisations make use of phishing reports to update their defences is not currently well studied. Having a procedure in place to receive and act on phishing reports is an accepted industry best practice recommended by many authorities [107, 249, 250]. But very little is known about the practicalities and workflows staff use to accomplish these goals, or how effective the various approaches are. What we do know is that effective phishing attacks can cause damage within minutes of making it to an inbox [107], so quickly responding to phishing reports is essential as it allows for rapid response and mitigation, which in turn limits the damage.

In this case study we focus on answering how a large University in the UK handles phishing reporting and mitigation. Universities are an interesting organisation to study for several reasons. First, the education sector has the highest phishing clicking rate even compared to sectors such as finance and healthcare [107]. So Universities are a prime target for attack and those attacks are currently successful. Second, Universities can be quite large, but their support staff are not funded or organised the same way a large financial institution's might be. Third, the yearly turn over of students makes typical approaches such as training more challenging to do. Finally, Universities have valuable resources to protect. The most obvious examples are the personal details of its students and staff, the content of ongoing research projects, and various intellectual properties [109]. Universities also expected to protect access to contracted services such as the JSTOR digital library with whom they have contracts promising to limit access to current staff and students.

Our investigation of the University began with shadowing staff working at the central Help Desk where phishing reports come in. We complemented these observations through contextual interviews with teams across the University that handle phishing-related issues.

We found that Help Desk staff become inundated by reports when large attacks occur and must balance their workload by prioritising security risks against potential impacts. For example, compromised staff or student accounts are potentially of greater threat than a single errant phishing email. Awareness of phishing incidents primarily come from end users' reports, however are not limited to this single source with alternative internal and external sources, requiring more sophisticated coordina-

tion and well-choreographed hand-offs between teams. This collaboration is seen as a distributed cognition across the teams which is essential for an effective response, enabling quick updates to automatic protections. However, lapses in current practice prevent teams from fully reflecting on incidents though practice flexibility also results in catching unexpected issues. End users who report phishing are typically given generic feedback which may not match their exact query. We also see that mitigation attempts can be hampered by the mitigating team having incomplete information, such as only looking at a single phishing example from a campaign, and therefore, missing subtle variations used by the attackers.

This case study is intended to form the basis of further studies into phishing report management processes. We believe that further research in this area will guide organisations to better inform their phishing incident plans. This case study highlights several challenges in handling phishing reports and the problems stakeholders face when managing phishing attacks. We also recommend that future research focus on augmenting phishing reporting systems using automation to help minimise staff time requirements while also making full use of every phishing report.

## **5.2 Related Work**

To situate our research, we consider relevant prior research that explores security incidents management and research that specifically focused on identifying, assessing, and responding to phishing incidents. We supplement this with previous work on general IT management and best practices.

### **5.2.1 Phishing as a security incident**

The cybersecurity landscape can change rapidly as new vulnerabilities and mitigations arise, resulting in distinct working practices from other general IT incidents [251]. Security incident management, therefore, involves reporting, assessment, response, and learning from incidents to evolve existing security practices [105]. Phishing incidents are not unique in this regard, as response procedures are developed and refined through reflective evaluation and feedback to adapt related procedures and best practice [16,99,252,253] including the development of practices that consider the range of potential attack vectors, such as Social Media [254].

## Discovery and reporting

Security incident management begins with the initial discovery of ongoing security issues, which are identified through internal, (e.g. security monitoring mechanisms and employee reports) [21, 255, 256] and external sources (e.g. bug bounties or from another organisation) [21, 256, 257]. However, it can be challenging for these external groups to identify who to report to [21]. Internal security monitoring tools, although helpful at detecting security incidents [256], can overload IT staff with a large influx of reports arriving simultaneously for a single incident [20] or from false alarms [258]. These numerous generated reports must be sorted through to determine their legitimacy before action can be taken, taking up valuable staff resources [106].

Phishing is designed to trick both email servers and end-users into thinking that they are looking at a legitimate email, making discovery and accurate validation essential for success. Therefore, security managers recommend concentrating efforts on the identification of phishing through employee training and preventive technical measures [108]. While valuable, these solutions are not perfectly accurate, nor do they protect against more sophisticated phishing attacks, such as *lateral phishing* where the attack originates from a compromised account that is already verified and trusted [52]. Human awareness is therefore seen as an integral component within organisations' security management strategies [252], as it adds a layer of security while complementing the technological controls (e.g. filters). Despite the benefits of training it has been found to not be fully effective with between 15.57% [84] and 69% [17] of people still falling for simulated phishing tests [16, 17, 84], and employees and users are still expected to report phishing [249, 259].

Timely responses greatly aid organisations to react to and mitigate attacks, reducing the number of potential victims who would engage with the phishing communication and therefore reducing or avoiding potential organisational damage [20, 107, 108]. However, the number of phishing reports is still considered too low [107] as people usually only report phishing email when they doubt its safety and need an expert's opinion [18, 19], know the spoofed sender, have a desire to protect other potential victims, or perceive the email to be particularly convincing and therefore dangerous [19]. Counter to this, research has found that users may not report due to a lack of awareness of legitimate reporting channels, concerns of mishandling, and perceived self-efficacy [260]. To promote phishing reporting, prior work has looked at using staff feedback after training simulation to modify policies to better align with staff needs [17].

Phishing reporting can also be treated as a problem of Communal Knowledge Management, where employees report potential phishing to a website that can be publicly accessed [261]. Praising a legitimate reporter and sharing with other staff has been investigated and shown to have a positive impact on encouraging staff phishing reports for a number of scenarios [253].

### **Incident assessment**

Assessing security incidents is non-trivial as it requires trained staff to review, communicate, and identify the causes in order to determine the relative importance of issues [20]. The general public struggles to accurately identify phishing resulting in numerous false-positive reports [22,261] which must be verified and validated [21] for accuracy as well as determine how critical they are based on potential impacts. For example, the impact can be judged based on the number of affected users, the affected services, or the type of users affected [256,262]. Automatically prioritising sophisticated phishing emails helps the Incident Response Teams act on phishing that is more likely to harvest clicks [106].

### **Responding to incidents**

Responding to security incidents often involves coordinating with staff from many areas of an organisation, and with differing expertise [20,258,263]. Organisations are known to use established policies and procedures to help staff follow best practice when responding, which involves investigating the cause, escalating to the required teams while simultaneously documenting all actions taken [256,262]. While these protocols are indeed impactful, they do not necessarily match the actions staff take when handling an incident. Staff responses can be influenced by their attitudes towards the applicable security policy and their interpretation of the policy within their working context [264]. Additionally, the usability of prescribed forensic tools can impact their abilities to follow the best practices [256,265].

While there is minimal research detailing actions around handling phishing reports by experts, automated responses to phishing incidents have proved challenging. It directly depends on the accuracy of the initial report and; therefore, requires expert human validation [22]. While complex, looking at how to backtrack to the origin of a phishing attack and analyse it can help investigating social engineering crime [266].

## Learning from incidents

Learning from past security incidents is also challenging for organisations. Some organisations do not have a formal approach to gather lessons or redistribute those lessons to staff [267, 268], which itself may take considerable time [269]. Organisation risk focusing solely on solutions to incidents without reviewing larger policies or organisational structures [262]. Additionally, reviews may be biased when the focus is placed on rarer large-scale or severe incidents which obscures potential day-to-day lessons [262, 264] and results in overcompensation with security taking an overbearing role in incident management [264].

Best practices for learning primarily focus on the technical aspects and direct cause of incidents [270]; however, security policies may also be causes for learning. For example, Sasse and Brostoff [271] investigated the large number of incidents raised to an industrial organisation's help desk regarding password resets. The research was motivated by the high costs the organisation was incurring in help desk staff time. It concluded that modifying the current unusable password policy had the potential to reduce help desk staff time by 40%.

For phishing incidents, little is known about how to learn from them; however, qualitative and quantitative metrics are used to evaluate security incidents including phishing related incidents handling performance, with the latter being more dominant, such as response time, the number of tickets, and the number of incidents [108].

### 5.2.2 IT incident management

Incident management is a well-established space in Information Technology where service quality, users satisfaction, and system stability are examples of essential organisation requirements.

Many of the processes and policies implemented by organisations will be influenced by their chosen IT governance standard or frameworks such as Control of Business Objectives and Technology (COBIT) [272] and Information Technology Infrastructure Library (ITIL) [273]. Frameworks similar to ITIL dictate the structure of their IT department, guiding how to organise teams, as well as how to handle communication and coordination between teams [274]. ITIL has been found to improve overall service quality [275], customer satisfaction [275], speed of incidents' responses [276], and the number of necessary escalations [276].

However, implementing such frameworks can be challenging [277] because of the

natural tension between theory and practice. For example, frameworks rightly advocate solving the root problems of identified technical issues, but workarounds are much faster and easier for teams to implement [274]. Frameworks also only offer high-level guidance, resulting in a variety of smaller implementation decisions between organisations [276, 278, 279]. With differences in implementation, IT infrastructure's performance is evaluated based on numerous indicators such as customer feedback, internal business processes, and the learning achieved [280].

The choice of tools used to manage IT incidents can have a direct impact on service quality and efficiency [281], which has resulted in research on the development of tailor-made software [281, 282] or the customisation of Off-the-shelf or outdated tools [283–285]. Still, tools alone cannot fix a broken process or workflow [274, 283]. Understanding the issues and finding an optimal workflow process before selecting the software can help organisations decide on the tools that best fit their needs.

The majority of IT incident handling begin with calls to the Help Desk and often take the shape of routine questions and issues that staff can answer confidently. However, around 10% of all calls require further research and escalation to the relevant teams [286, 287]. Research has found that end-user satisfaction is influenced by both the perceived quality of the solution [288] and their beliefs regarding the trustworthiness and level of expertise of staff resolving issues [289]. IT departments rely on several knowledge sources to alleviate pressure on staff and provide information regarding commonly reoccurring incidents. These knowledge sources can take numerous forms, including Internet repositories, cross-organisation shared knowledge base [290], Frequently Asked Questions [291], and peer advice [292].

With peer advice, for example, staff can seek their peers' help when they lack the needed expertise and the situational awareness due to the distribution of information based on the roles [293, 294], both justifying the necessity for hands-off between staff [286]. When issues regarding systems' stability occur, hands-off can involve numerous teams and staff [295] who can provide incident resolution at the right time and in the right context [286].

### **5.2.3 Distributed cognition**

Distributed Cognition theory (DCog) [296] essentially describes the collaboration between multiple agents as a single cognitive systems [294, 297]. This collaboration includes human-computer interactions; thus, DCog is well matched to understand the



relationship between humans, tools and artefacts [298]. In this work, we use DCog as a lens for understanding coordination between embodied agents by analysing the interactions between the people, the problem, and the tools used both in planned and emergent cases [297, 299]. *Cues*, and *Norms* are two key features necessary for supporting work in a distributed context. A cue is defined as a signal that indicates to individuals the required actions and how to enact them. In contrast, norms are the procedures that ensure consistency between individuals' tasks [263, 297].

Similar to other ITSM findings [263, 294, 300, 301], our initial findings from observing the Help Desk showed that their work is highly distributed in nature. We therefore chose to use elements of DCog in our analysis and as a lens in the discussion to understand our results in a wider context.

We are not the first to observe that DCog is part of IT management. Individuals from various units of an organisation collaborate formally and informally to address IT issues, which are characterised by pattern recognition, hypothesis-generating, and testing for uncertain success [300]. For example, Maglio et al. combined distributed cognition framework with joint activity theory to understand a specific problem-solving instance in web-based administration [294]. Botta et al. further expand on this by applying distributed cognition within the context of security management and identifying its influence over organisational processes [263]. However, little work has been done on understanding distributed cognition when resolving phishing incidents by various IS teams.

### 5.3 Participating organisation

The University studied is an internationally recognised UK institution, which supports around 40,000 students and 15,000 academic and administrative staff members. It is distributed over multiple campuses inhabited by several academic schools, each with their own respective local IS management teams that manage their own resources. In total, there are around 1000 IT support staff.

The University's IT service management is guided by ITIL (Information Technology Infrastructure Library), a framework of best practices which is used by organisations worldwide and in diverse sectors and industries [302]. The University has adopted and adapted the ITIL framework for use by all IT services, creating a dedicated Quality Enhancement team to ensure compliance. Since the adoption of ITIL, the University has reported improvements such as the increased clarity of teams' roles

and responsibilities, reductions in services outage, consistent logging of incidents, reduced running costs, and improved customer satisfaction. Improvements such as these are considered significant indicators of a successful implementation of ITIL [274]. Additionally, the University's IS achieved a Service Desk Certification <sup>2</sup>, which is an industry accreditation program specifically designed to certify service desk quality, indicating that the University's ITIL implementation and the workflow detailed within this case is comparative to other organisations using ITIL to inform their phishing practices.

## 5.4 Methodology

The study data was collected from two sources: 1) ethnographic-style observations of the daily work of the Help Desk, and 2) interviews with other University teams. We followed our own University's ethics procedures and at all times ensured that participants were aware that participating in our research study was voluntary.

### 5.4.1 Introductions and setup

Before starting our research project we met with some of the stakeholders, namely the Chief Information Security Officer (CISO) for the University and the Help Desk manager. We explained our project goals and discussed possible project structures that would allow us to conduct the research in a minimally disruptive manner as well as provide insight that might be useful to the University. We also discussed the phishing-related issues they thought were most problematic for their teams. We used these insights as initial scoping for the semi-structured interviews discussed below. Both the CISO and Help Desk manager were very positive about the project and were interested to see what we would find. They were also interested to know what impact would this project have on future research. Their support was vital as they were able to provide all the access needed.

### 5.4.2 Observing the Help Desk

The CISO and Help Desk manager confirmed that the Help Desk was the intended first point of contact for anyone reporting a phishing message. They are responsible for

---

<sup>2</sup><https://www.servicedeskintstitute.com/service-desk-benchmarking/service-desk-certification/>

initial assessment of the report, escalating it if needed, and responding to the reporting user.

Given the Help Desk's central role, we started by observing their workflow. The Help Desk manager allocated a desk for the lead researcher so she could spend time with the team in order to build a good relation with the Help Desk staff and familiarise herself with their work practices. She started by shadowing Help Desk staff while they were doing their daily work. Initially, she only observed and asked about the full range of their normal work practices, then in the second week she started focusing more on phishing-related work practices. The normal mode of observation was to quietly observe the staff doing their work and then ask follow-on or clarifying questions when staff were free. The observations were contextual, bordering on ethnography.

The observations were done over two months, with the researcher taking notes and spending between one and two days a week observing. They conducted focused observation of six Help Desk staff with each observation lasting 4-5 hours. To avoid disturbing the staff, the shadowing time and day was arranged by a senior Help Desk staff based on the staff availability. The researcher also spent time at the provided desk observing the flow of the space and briefly observing different staff as interesting incidents arose. Observed staff included experienced staff, new staff, and an undergraduate computer science student doing work experience.

The researcher was also given limited access to the ticketing system used by the University to help teams track all types of issues within the University. She also attended the training sessions for using the ticketing system. The Help Desk uses the ticketing system to manage communications with users and other teams. They refer to all interactions with other groups as 'calls' and track them through the ticketing system. Calls can be digital, but they can also be a phone call or someone walking into one of the physical Help Desks and asking a question, all of which are logged in the system. The lead researcher was able to use the ticketing system to better understand how phishing calls were handled and passed between teams as well as understanding communications between the Help Desk and end-users. Additionally, the lead researcher was given limited access to the Help Desk's private knowledge base to better understand the observed practices. Throughout the research we took care to use these resources respectfully and quotes used from them in this work have been carefully redacted to protect staff and end-users.

As expected, phishing-related tasks are infrequent and tend to occur in clusters, such as when a single phishing campaign generates many calls in a short time period.

Consequently, the researcher was only able to observe one live reaction to a phishing campaign. Instead, during breaks in work the researcher asked staff about their prior experiences with phishing calls. Because the ticketing system is normally open during work, it was easy for them to pull up prior phishing calls they had handled and discuss them.

### 5.4.3 Interviewing university teams

Observing the Help Desk also gave insight into the work flows of the teams they work with, most of whom use the same ticket tracking system. To better understand the work practices of these other teams, we conducted interviews. Most of the interviews were contextual interviews [303] where during or after the interview the participant showed the lead researcher real phishing handling examples of how they do their work, the systems they use, and metrics from previous attacks. Interviews were mostly conducted in nearby meeting spaces to avoid disturbing other staff.

The Help Desk manager provided introductions to other University teams that deal with various aspects of phishing, even if their involvement was minimal. We were therefore able to interview one or two members from six teams, each of whom work on a range of phishing-related issues including: dealing with users, account resetting, desktop computer rebuilding, best practice management, email relay management, interface with Office365 email, and security. Unfortunately, we were not able to interview the team that manages the network and the virtual team focused on security. In total, we conducted about 25 hours of interview. All the interviewees were experienced staff who worked for the University for more than 4 years and most of them were their team's manager or leader. The lead researcher also constructed a diagram of inter-team workflows from early observation and interview content, she then iteratively improved it by asking staff in following interviews about the accuracy of the identified inter-team interactions, procedures, and their phishing-related roles.

Team interviews started by explaining the project and the general goal of understanding how the University handles phishing reports. We then asked them to explain their team's mission in their own words and their general work practices. We then narrowed in on their phishing-related activities, including their interactions with other teams. The bulk of the interviews involved follow-up unstructured questions on issues or topics were brought up from other team interviews or Help Desk observations.

#### 5.4.4 Periodic review of findings

The shadowing was for understanding the workflow rather than for improving performance; however, the staff were unsure about the data collected. Therefore, the lead researcher setup some feedback sessions with staff where she summarised their findings to demonstrate the type of data she collects and ask for feedback or corrections on the observations. To help guide the research further, approximately twice a month the lead researcher would give a slide presentation detailing their latest interesting observations from the Help Desk shadowing as well as the interviews to our research lab who were encouraged to ask questions and comment. The presentations included the developing diagram of inter-team workflows as well as information flows within teams and between the Help Desk and end-users. The presentations were used to help the lead researcher process observations as well as identifying areas that needed follow-up to understand. As these presentations happened regularly, the lab group was also able to provide needed external clarity.

#### 5.4.5 Interview data analysis

Interviews were audio recorded if the interviewees allowed it, if not, the researcher took detailed notes and completed a write-up immediately after finishing the interview. All audio recordings were transcribed by a researcher, with participant's personal information and team names being substituted for IDs.

Two researchers reviewed all the transcripts and notes. Using open coding as they went, both researchers constructed their own independent codebooks focusing on the process of handling phishing. The researchers then met to discuss the process that had been observed. Through iterative coding and discussion sessions the researchers reached agreement on the workflow for handling phishing as well as the problems and friction points. Following each round of discussion, the two researchers provided feedback to the third researcher so as to guide reporting of results.

During the open coding, it became evident that phishing management at the University was an example of Distributed Cognition as the process of managing phishing clearly involved more than standard escalation of issues, and instead required multiple groups to communicate about their own unique perspectives of the incident and work together to manage it properly. DCog was therefore used to guide the analysis by putting more emphasis on the communications between teams, particularly, points where one team had access to data or resources not available to the other teams and

how that information was being conveyed.

To ensure accuracy of the presented results, an early draft of this paper was shared with stakeholders and their comments were discussed and addressed.

## 5.5 Results

Phishing is handled by multiple teams within the University, including: Security, Quality Enhancement, Help Desk, Mail Relay, Mail Exchange, Network, Incident Response (IRT), and Service Delivery teams. The level of involvement of each team is different; some teams routinely handle phishing-related issues while others are only involved in emergencies or other specific circumstances. End-users are also an important part of this distributed process as they identify, report, and ask advice about phishing by contacting the Help Desk.

### 5.5.1 Phishing campaigns— managing the load

In this section we focus on how the Help Desk manages phishing calls. We observe that their largest problems involve: managing large numbers of reports coming in, deciding what reports need to be escalated to other teams, closing out the report calls efficiently, and using their own judgement on non-standard phishing reports.

The most common phishing attack handled by the Help Desk is *phishing campaigns* where the phisher sends emails to many recipients to increase the odds that one or more will interact with it. The emails are often visually similar but contain variations, such as putting the recipient's name in the email body (e.g. 'Hello Alice,'), using slightly different body text or creating custom URL links for each recipient.

In most cases experienced by the Help Desk, a phishing campaign will use similar subject lines for all the emails. When reporting phishing, users often forward the email, causing the ticketing system to automatically adopt the subject of the phish as the subject of the ticket. If a Help Desk staff member finds a phishing email in their own inbox, they can confirm a campaign by comparing this email to those already reported in the ticketing system. It is one of the cues they use to verify phishing campaigns. Help Desk staff typically use the number of reports with similar-sounding subject lines reported in a short time frame as a signal of a campaign. Often, such sets of reports will happen in the morning due to users checking their email then.

[Help Desk] Morning usually is the peak time for us. We receive calls

between 8 and 9 am because usually, staff will come in the morning check their emails and report phishing.

After determining that they are looking at a phishing campaign, staff then select one or more of the calls to escalate to the appropriate teams. The remaining similar-subject calls are either temporarily ignored by staff or grouped together into a single open call, to reduce clutter in the ticketing queue. Later in the day when new phishing reports have stopped coming in, one member of staff will voluntarily go through and close all phishing calls at once.

The above workflow has naturally evolved as a way for Help Desk staff to manage phishing reports alongside their other service delivery tasks.

### **The number of phishing reports can overwhelm**

The Help Desk's primary goal is to optimise the number of calls processed, either by closing or escalating them, ideally taking less than 15 minutes for most calls. A phishing campaign is problematic for the Help Desk because it generates a large number of phishing calls, each of which must theoretically be handled individually, taking time away from other calls. For 2019, they received phishing-related calls on at least 20 days out of every month, with call counts ranging between 2 and 170 per day. While that number may seem large, it only represents a small percentage of the University reporting phishing. If a theoretical phishing campaign were to target all University staff and students (about 58,000 people) and even 1% were to report it, that would be 580 calls, well above the normally observed number.

Most of the teams, including the Help Desk, agreed that having people report phishing was needed as it is the cue for identifying campaigns. However, they also recognised that the University did not have the resources to look through all the phishing reports. An Quality Enhancement staff member quote explained:

[Quality Enhancement] We want people to report [phishing emails] but we want to be able to manage the load of calls. The problem is we cannot manage them.

The Security team similarly recognised the problem of a lack of resources impacting the Help Desk's capacity for managing reports:

[Security] At the moment we don't have the resource to have someone look at them and triage them which is our main problem.

This overloading problem resulted in the Help Desk adapting their practices to fit the ticketing system's functionality, as we discuss later. Another tactic the Help Desk uses is to get help from other teams, such as the Mail Relay, to block still incoming campaign emails and remove existing phishing from peoples' inboxes. Doing so has positive security impacts, but more practically, it stops the flood of reports making it a strong immediate motivator to react fast.

Help Desk overload also impacts the University's ability to send simulated phishing messages to end-users as part of security measurement and training. Sending such fake phishing emails is currently an industry best practice to understand how well-trained staff are and if the procedures put in place are effective. However, when the University attempted such an exercise, they unintentionally overwhelmed the Help Desk with calls as the Mail Relay's normal work-management strategy of quickly blocking the incoming attack cannot be used on simulated phishing. The overload damaged the Help Desk's ability to do their daily tasks and their ability to provide customised feedback to reporting users, resulting in a loss of a potential user-training opportunity.

[Security] The problem is when people then report [the fake phishing email]; there is no inbuilt system on our email that says 'Wait, this is a fake one, calm down'. So with the issue in the Help Desk, Help Desk was completely swamped and that what happened after we ran the simulation the first time.

As a result, the Security team temporarily stopped sending fake phishing emails and are working on finding ways to better conduct training and testing. Including agreeing to be part of the research project described in this paper as an effort to better understand how phishing reports flow through the organisation and where provisioning is needed.

### **Deciding what to escalate so as to not waste other teams' time**

The main point of the Help Desk is to decide what does and does not need escalation, so other teams only spend time on problems that require their expertise. Hence, one of the Help Desk key tasks is triage, where they sort through reports and identify which calls require escalation to the appropriate team.

In addition to wasting time, escalating unnecessary calls can also result in a polite rebuke from the other team. We observed several situations where a call was escalated, and the other team responded that they had already handled this one in the morning, or that necessary information was not present. Consequently, the Help Desk staff try only to escalate calls when necessary.



[Mail Relay] If they are asking us to block a specific email and they told us 10 times already, then we do not need to see it again ... but what we do not want is the Help Desk passing all the hundred calls to us just saying the same thing 'This person received a phishing email'. We only need to be told once.

Given the need for human validation of potential phish, there exists no formal approach for identifying the 'correct' cue and the applicable actions. Therefore, we detail the range of identified cues and the associated norms made available to Help Desk staff.

*Is the reported phishing email from a University email address?* Compromised University accounts are a serious problem, so if a reported phishing email is from a University email address, the Help Desk will handle it differently. We further detail how compromised accounts are handled in Section 5.5.3.

*Is this a campaign or a one-off attack?* A key criterion for the Help Desk is the number of reports. If there is only a small number of phishing reports, then requesting action to block incoming emails is likely a waste of other teams' time. The attack may be a one-off, or the user could just be confused. Neither case justifies escalation.

Looking at the phishing emails reported during May, we found 9 phishing campaigns with more than 4 phishing reports, 12 campaigns with 2-4 phishing reports, and 40 phishing reports of single phishing emails. Using the helpline strategy, only 21 calls would have been escalated in May.

*Do other teams already know about this phishing campaign?* If other teams are already aware of the problem, then escalating again is unnecessary. So before escalating a call, Help Desk staff check that it is not already being handled. The Help Desk maintains a separate list of phishing emails that are logged with Mail Relay. The list includes the subject lines of phishing calls that have already been escalated.

[Help Desk Knowledge base] Due to the current number of phishing attempts being made on our email service, we really need to monitor what we have sent to Mail Relay so we don't end up inundating them with calls.

Sometimes other teams become aware of a phishing campaign through other sources, such as getting the phishing email themselves or being notified by another University, detailed in Section 5.5.4. When that happens, they notify the Help Desk so that they do not escalate the calls again. However, the various sources of notifications resulted in missing them or delay the response process. In the below example, a call was

escalated to the Mail Relay team, who then deescalated it because the IRT team had already created an earlier call for the same issue which the Help Desk would have been copied in on.

[Mail Relay deescalating a call] The Help Desk were passed a call about this from IRT earlier today.

*Is all the necessary data present?* When reporting, users often forward the problematic email or provide snippets of it in their communication. Nevertheless, the specialised teams such as Mail Relay need the header data from the original email, which is not included when the email is forwarded.

Fig. 5.1 shows two examples of why the original emails (*.eml* files) are needed. Both pictured emails are from the same campaign sent within minutes of each other to two Lecturers in the same University department. To a human, they are obviously part of the same scam, but they are quite different to a computer. The sender emails are different, the use of link text differs, the email text body itself is different, they have different HTML formatting, even the subject has variations. These example variations are consistent to those seen in prior work [18, 106]. Automatically identifying this scam from the email of a whole University is not possible using only the pictured information. However, the *.eml* files contain additional information such as a list of all the email servers the email passed through, their IP addresses, encryption signatures, and any checks performed by the servers.

The easiest way to get the header information is to have the user save and send the original email as an *.eml* file. Reporting tools, such as ones the University is currently pilot testing, will attach this file to the phishing report automatically, but at the time of observation, the only way to get it was for the user to provide it explicitly. To save time, the Help Desk has a pre-written response text, called a *Standard Solution (SS)*—detailed in Section 5.5.5, for asking the user to send the email as an *.eml* file, including instructions on how to do so. So if the *.eml* is missing, the staff managing the call will ask them for it using the SS. The user may then take some time to respond or not respond at all, delaying the report escalation.

*Escalating the call.* Once the Help Desk has all the needed data, they again check that the call has not yet been escalated, and if not, they escalate it. First, the Help Desk staff update the aforementioned list of escalated calls with their phishing incident's subject and then escalate the call to the Mail Relay, Mail Exchange, and IRT teams.

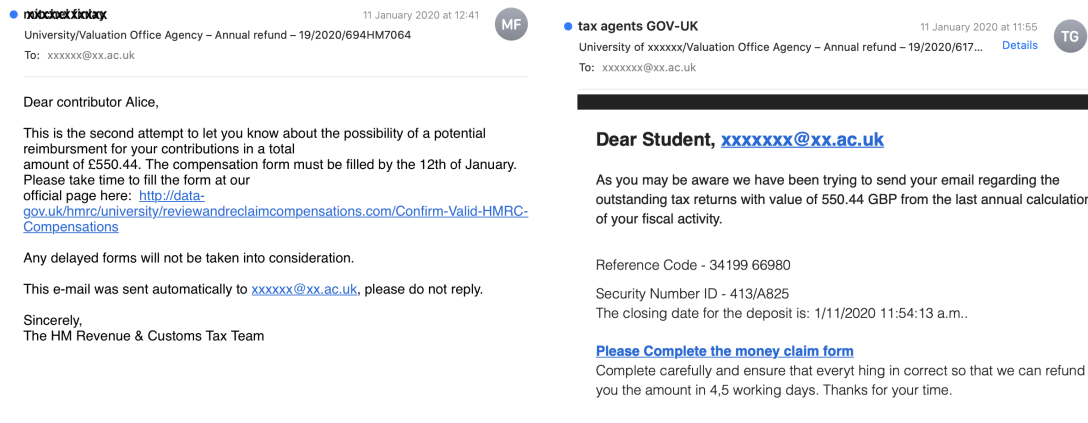


Figure 5.1: Phishing emails sent to two University Lecturers in the same department on the same day. Mildly edited to make them anonymous.

Typically the escalation will be done on a single call so that all three teams can see comments made by the others. The purpose of the escalation to the three teams is to ensure situational awareness across the ITSM department; however, only one team may work on the call depending on the situation.

### Closing calls is also time consuming

Closing a phishing call involves providing feedback to the end-user and marking the call as resolved in the ticket tracking software.

[Help Desk] We would escalate one of the calls to the appropriate team to action in terms of blocking at the relays or removing the offending email from mailboxes. The rest of them we would just contact the user to say delete the email, change your passwords if you've clicked on a link etc.

To do so, a staff member has to look up the phishing SS in a long list, pick the appropriate one based on the user query in Table 5.1, select it, send it to the user, copy it to the call history, and then mark the call resolved. The actions are fairly simple, but when 100+ calls all require the exact same set of actions, it can get repetitive.

To combat this, the Help Desk staff will tag all incoming phishing calls with unclear subjects by adding '[Phishing]' to the subject and leave them open on the system. They will then wait about 30 minutes for any further calls or till the rate at which reports come in slows and then bulk close the calls via an internal communication system. They then select everything either tagged '[Phishing]', or that is clearly part of the day's phishing campaign, and close them out all at once, sending all the reporting

users the same SS and marking all the calls as resolved as one bulk action; thereby, saving everyone quite a bit of call-closing time.

[Help Desk] Sometimes the volume gets beyond us, we are not going through this one by one and contact each user individually with SS. We are gonna just bulk and close these calls with the generic bulk closure message, it goes out with basically goes out and say if you want more information, look at the call in the service portal because sometimes we don't have the capacity to chip through these.

### **Help Desk staff are relied on to use their own judgement to identify non-standard problems**

While the Help Desk heavily relies on standard protocols, procedures, and processes they also encourage their staff to use their own judgement around how to handle each call and provide them with the tools to do so. While seemingly unremarkable, it is important to understand that staff are encouraged to handle calls within minutes and therefore tend to apply the heuristics described above to handle phishing calls quickly. Here we detail a couple of situations where staff noticed something odd and followed up in a non-standard way.

When shadowing a Help Desk staff member, we observed them take a call from a single phishing report which would typically be closed without escalation. The report complained that the forwarded phishing was impersonating their Department Head and asked the Help Desk to block the emails. The Help Desk staff member thought that the phishing email was particularly believable and therefore might constitute a high risk, but they were still uncertain if it was worth escalating. So they used a University provided message trace service on the email's from address and found that the phisher had sent the email to around a hundred users. So they immediately escalated it to the appropriate teams.

In a separate instance, the Help Desk received a large number of calls where users reported a legitimate email from within the University as phishing. The Help Desk team investigated the reasons for that misjudgement and found that an email sending service had been used that made the sent emails look like they might not actually be from the University.

[Help Desk] We have false positive occasionally. It has been assisted now. I am trying to think what is the bulk mail. XX mail is a commercial book mail relay that you can buy into that allows you to send lots and lots of emails to target audiences. Certain colleges, were using [it] for email campaigns and things like that. The problem with that is not coming from the

University email address. It has The University name in some part of it but it is not xx.ac.uk and people were reporting that it is phishing but it is not. We are not having it anymore because changes were made to make the email that is sent to people more obvious that it is part of the University. It should have the University logo, University address, and looks professional.

To address the problem, the University banned the use of email services that do not comply with best practice, such as providing URLs hosted under the University's domain.

### **5.5.2 Converting escalated calls into protections**

In this section we discuss how the Mail Relay and Mail Exchange teams handle escalated calls from the Help Desk along with other teams. Handling an escalated call most often involves finding a reliable way to identify the phish and then applying that method across several systems. Both of which may require hand-offs between several teams to find solutions and implement them. Also, some phish simply cannot be blocked or require several attempts to do so.

A large component of handling an escalated phishing call is ensuring that University users are protected from the reported phishing attack. People have a wide range of skills at identifying phishing [304] with some quickly reporting the phish and other more vulnerable users clicking on malicious links. To protect these vulnerable users, teams can use reports to remove phish from inboxes as well as prevent new phishing from entering the University email system. Doing so primarily falls to the Mail Relay team who handle incoming and outgoing emails and the Mail Exchange team who handle stored emails as well as emails managed by Microsoft's Office365. However, other teams may be required if the situation dictates so. Phishing campaigns can vary dramatically in sophistication with the process being straight forward if all the phishing emails share a unique feature, such as all coming from the same email address. But other phishing campaigns can be more complex (e.g. Fig. 5.1).

#### **Finding a common reliable factor to protect users**

Many features are used to block phishing, but some of the most common are the from address, subject line, and any mail relays listed in the header. The teams use a combination of experience and educated guessing to select potential features and then run practice searches to see how effective it is.

While the main process is fairly basic, doing it reliably is not easily learned as it requires knowledge of email header's structure as well as an understanding of IP addresses and IP address ranges. Feature selection also has to be done carefully so blocking and deleting will not interrupt users' work. A simple example might be a phishing attack that closely mirrors a genuine Dropbox email. If the Mail Relay team chooses to block on features that are also shared by real Dropbox emails, they could inadvertently damage all users' ability to interact with Dropbox. Universities are also complex, with staff working on a wide variety of problems, using a wide variety of tools, and collaborating with a wide variety of people. So blocking whole IP ranges, domains, or tools cannot be done lightly.

The effect of blocking or deleting emails can also range from time consuming to impossible to reverse. The Mail Relay team can trap emails in quarantine and then release them for delivery if they were incorrectly identified as spam. However, when the Mail Exchange team deletes an email, it is permanently gone and cannot be recovered. As a result, the Mail Exchange team has internal procedures around email deletion, such as requiring the manager to approve all deletion commands. They also require specifying a date range, from address and subject line in the deletion commands to limit potential damage.

[Mail Exchange] Within Office 365, you can do what we call it content search. There is one various procedures to ensure we don't do it unless we have the approval to do it but yes we do have the ability to do that and any thing we do is fully auditable.

### **Fixing requires knowledge, resource, and responsibility**

University teams are generally either centred around resources that need to be managed (e.g. Mail Relay) or around specialised knowledge that takes time to accumulate (e.g. Security). Knowledge regarding how to address a phishing incident is often located in many parts of an organisation. This situation is fairly common in IS support as different teams and members often gain in-depth understanding of the systems they work in and then collaborate with others as needed [295]. In terms of phishing, teams need knowledge, both in terms of making non-disruptive changes to running systems and making changes that are likely to have the desired security outcomes. Large computing systems are made of many interconnected components where a single change can have unexpected side effects. To balance this distribution of knowledge and ensure their changes meet the best practices, the University teams actively coordinate with

the resource-based teams to make the actual changes to the system and also to ask for advice and sign-off from other teams, such as Security.

Access to resources, such as systems, services, and channels, is also often restricted to ensure that people with insufficient knowledge do not make well meaning but potentially very problematic changes. For example, it would be much faster and more efficient if the Help Desk could directly go from identifying an accurate phishing report to applying an appropriate rule to the mail relay to stop new phish (and reports) from coming in. However, as the Mail Relay team notes below, knowledge of how to modify the relay safely is not quickly learned.

[Mail Relay] It is so complicated only experts can understand [filter rule construction], I don't think the Help Desk have access to the black list.

Consequently, the Help Desk cannot edit themselves and instead raise a call and wait for the Mail Relay team to take action.

University systems also frequently impact each other, requiring teams to coordinate or ask each other for help addressing an issue. For example, a Mail Relay team member recalled a prior phishing campaign when so many emails were coming in the same time, causing the relay server to experience overload. So despite applying mail filtering rules, the phishing emails were still negatively impacting users' ability to get emails. To solve the problem, they reached out to the Network team to temporarily block all incoming emails from the phisher's IP address range. Doing so reduced the load on the mail relays and resulted in a successful blocking of the malicious emails with minimal impact on University services.

[Mail Relay] We do communicate with Network team in active attacks from a specific IP. Network team will block the traffic to save resources.

In situations requiring specialised knowledge, teams may require consultation, guidance, or sign-off from other teams before enacting a change. For example, phishers will sometimes send emails from an IP address range to make blocking the email harder. IP address allocation is a constantly shifting feature on the Internet as addresses are assigned and reassigned to different Internet Address Registrars. Some ranges are unallocated and can be safely blocked, while other ranges might be used by valid users. The Mail Relay team's knowledge does not necessarily include an in-depth understanding of the current IP Address allocation issues, so when choosing to block IP ranges they may consult with teams like Security or Network who have this knowledge.

[Mail Relay] We consult Security with policy and security. For example, if we have an attack from a country, we will escalate it to them for decision ...

We can't block all the emails from Nigeria. It will damage our reputation when someone from Nigeria tries to contact us. We should just balance the cons and pros of each rule. Many emails will be trapped by blacklists.

While some teams play a direct role in the day-to-day responsibilities of handling phishing, others monitor the ongoing incidents and processes used. For example, the Security team observes the internal processes and advises on change to better promote security practices. They do this by tracking ongoing security incidents handled by the IRT through monitoring of the IRT's resources:

[Security] We have access to the IRT's inbox. So we can see what is coming in and out. So, we know what is happening and we can actually see the [ticket system] calls that are assigned to IRT's inbox. So that how we know what is going on.

### **Some messages cannot be blocked or removed from inboxes**

Sometimes, even with the help of multiple teams, phishing emails cannot be reliably blocked or deleted. During the data collection, a large phishing campaign hit the University. It was sent from different email addresses and had unique subject lines, often involving recipients' names. The Mail Relay applied several blocks on the relays, but the number of from addresses involved made it challenging to fully block or remove them from users' inboxes. Instead, the Quality Enhancement manager relied on the users as the last point of defence and sent an email on to all University members explaining the email characteristics and asking them to delete any email that matched.

[Quality Enhancement email] ... the University is receiving a very significant number of phishing attempts which have the recipient's name in the subject field and some text in the body of the message ... If you have received a message like this can you please delete it without clicking on any links within. If you have received a message like this and clicked on the link, please contact [Help Desk] for advice.

### **Email blocking does not always work on the first try**

Normally, the Mail Relay team will use an escalated call to construct a blocking rule that effectively blocks all emails of that type. But sometimes the blocking rules miss a portion of the campaign.

In the following case, a campaign had happened, the Mail Relay team had been escalated to, and end-users had been sent the SS. The user questioned the block's effectiveness and sent a follow-up call in response to the first phishing report to ensure



that a block was correctly applied on the mail relays. However, they were not satisfied with the response and sent in yet another call.

[End-user] Got another three e-mails through from the xxxx@gmail.com address, so that the block you have put on does not seem to be effective. Could you check what has gone wrong?

In response to the initial call, the Mail Relay team had applied a block based on the phisher's email address, but the phisher was sneaky and bypassed the block. A Help Desk staff correctly identified that the above email was an indication that the Mail Relay team's solution was not working and escalated the call despite it already being on the list of previously escalated calls. Using the new information, the Mail Relay member consulted the staff within the for a solution; thus, Mail Relay team expanded their rules to look for the problematic email address in multiple header locations. Also, given the sneaky attacker behaviour they also chose to block all email traffic originating from the IP range the attacker was using.

[Help Desk escalation] Hi Mail Relay, The block on the below email put in place in call xxxx2241 appear not have worked. Can you advise on what else can be done to try and stop these emails getting through?

[Mail Relay member 1] Any idea why my entry in the access file didn't work?

[Mail Relay member 2] The simple way of blocking an address relies on the return path - which is what the mail system sees - being the same as what is in the from header. That is usually the case, but in this case it was not. They are being rather more sneaky than usual. I have put in a different sort of block which should block anything with 'xxxxxxx@gmail.com' in the From: header. I also identified four network [IP] blocks from which these spams have recently originated (though there may be others I have not found) and blocked all traffic from them.

### **5.5.3 Compromised accounts – a serious reputation and workload problem**

Accounts can be compromised as a result of a successful phishing attack. Once the victim provides their credentials on a fake web page, the attacker can use them to gain access to the victim's accounts. In this section, we discuss University practices of reported compromised accounts. More specifically, we detail the cases in which they hand-off the calls to IRT team or rely on users to revoke the attacker's access to their account. IS teams have strong security awareness with regards to compromised

accounts; thus, reacting on them is independent from reacting to other phishing incidents.

Compromised accounts were found to be mutually understood by all teams, and were unanimously identified as definite security concerns, given that multiple services can be accessed using the University's single authentication system. Among other actions, they can use the accounts to gain access to send out more phishing emails. Similar to findings from previous research [52], staff from Mail Relay, Help Desk, and Mail Exchange believed that this is the most common scenario for compromised accounts in University.

[Mail Relay] Every time an account is compromised, effectively theft of personal information.

Blocking phishing emails is done internally by the spam filter on the mail relays managed by the University's Mail Relay team and also by the default spam filter on Office365 Exchange Server managed by Microsoft. Phishing emails from compromised accounts are treated more seriously than other phishing emails since they do not go through the relay's spam filter and they get a lower spam score from the Exchange filter because they are from the same domain (internal users), which is considered trustworthy.

[Mail Relay] Messages sent internally from one student to another student do not go through any of that relay's scoring. It is all set within Microsoft Office365.

[Security] Add scores to the emails is the things that Office 365 and Microsoft do. They have what is called spam ratings and they basically just go: 'this looks like spam and this does not', based on the huge number of criteria. We use the scores to some extent to block things that have basically the highest score because always entirely they will be.

Compromised accounts are sometimes used to send out spear and loosely targeted phishing emails aimed at specific groups of individuals. Both of these attacks are hard to detect by end users [304]. We observed several cases of these attacks while shadowing the Help Desk staff. In one case, the attacker compromised the victims' account and used it to send several emails to other University members hoping that one of the recipients knows the victim and replies to the attacker.

[Help Desk] The user here reports [Spear Phishing]. So the user was aware of the phishing type. The email said only 'Are you available?'. If the user knows the person who sent the email, they will communicate with the attacker and maybe transfer money to them. Mail Relay told us that the sender accounts are compromised.

Compromised accounts also pose a reputation problem. They can be used to send a malicious email to other organisations with the University's name associated with it. Considering that University is moving toward applying the cryptographic signature of the University's mail relay, other organisations use these signatures in their spam filters, so email coming from a University compromised account is more likely to make it through other organisations' spam filters and harm reputation when their teams have to handle the messages.

[University public Knowledge base] Phishing campaigns lead to compromised accounts and as it stands currently, risks the integrity of the University's reputation, data and ultimately, University business.

Compromised accounts are one of the most critical impacts of phishing campaigns as they can quickly snowball in scale as internal emails can reach more users who can potentially become compromised and phish other users.

Despite their importance, fixing compromised accounts is time-consuming and hard to do at scale as each account must have its password reset individually.

[Mail Relay on behalf of IRT] The problem we have is that often in the time between the initial compromise and the sending of spam and then us trying to delete that phishing from other users mailboxes, even if that is 10 to 15 minute, we get dozens of people who then get compromised. We then have an ongoing problem of phishing going around the University, moving from one person to another person to another person, and we can't keep deleting them all. So ideally, what we want to do is stop getting them in the first place, which is a really difficult thing to do.

### **Reported phishing from new compromised accounts should be escalated**

For every phishing incident, the Help Desk staff look at the sender of the phishing email; if it comes from an internal user, such as a student, it is considered a compromised account and requires more immediate actions beyond those discussed in Section 5.5.1.

The escalation of compromised accounts is similar to the escalation of phishing campaigns in that only new cases should be escalated; however, the Help Desk should escalate one call for every phishing campaign whereas all distinct compromised account reports should be escalated. To ensure the account has not already worked on, the Help Desk staff normally look at the list of blocked compromised accounts provided by the IRT team and only escalate the call to the IRT if the account is not on that list. Unlike the list of phishing emails logged with the Mail Relay, the known

compromised account list is maintained directly by the IRT team. Unlike phishing escalation, Help Desk staff are not expected to verify if the account is compromised before escalating it.

[End-user] Just to make you aware, please see above most likely a phishing email that has been sent from a student account.

[Help Desk] Hi IRT, I can't see this account on our list of compromised accounts.

[IRT] I see no evidence that the account is compromised.

A Mail Relay staff on behalf of the IRT team was positive on hand-off calls that ask if an account is compromised or not.

[Mail Relay] If the Help Desk are asking us to look at a specific email from different account saying is this account compromised, we will look at each individual one of those.

Reporting compromised accounts is necessary since IRT staff can use it to trace other compromised accounts. The example below illustrates how the IRT staff used the report to reset the compromised account, escalate it to Mail Exchange to remove the emails, and also lock other user accounts proactively because they were observed sending the phishing emails from the same location and at the exact time.

[End-user] I received the attached email over the weekend, I think it might have gone to other staff. It seems to be a phishing attempt and comes from what purports to be a student's email address. I did not click the link.

[Help Desk] Hi IRT, see attached email and compromised account.

[IRT] Passwords have been reset. Call also sent to Mail Exchange to delete the emails. User1, user2, user3 are also sending phish around the same time from the same place. Passwords have been reset.

### **Helping users who self-identify potential compromise**

Some users contact the Help Desk looking for advice because they clicked on a phishing email, resulting in a compromised account.

[End-user] I opened a dodgy email and now my email has been hacked and is sending spam all over, what should I do?

In this case, the Help Desk asks the user to change their password, and no escalation is required because in most cases changing the password will lock the attacker out. Time-wise, this is the best-case scenario since the user can solve the problem and do not lose any access to their account. Supposing a compromised account is identified

by someone other than the owner, then any calls have to be escalated to the IRT team who resets the user's accounts with a temporary password and asks the user to change later, which takes more time for everyone.

This scenario aligns with Help Desk strategy of shifting the responsibility to users to reduce future calls when the user re-encounters the same issue.

[Help Desk] We also rely on "shift left". Moving everything to the users. e.g. knowledge on the University website so when users contact the Help Desk, they will receive a link if users can help themselves.

#### **5.5.4 Information flow patterns are not in fixed orders or directions**

While our focus started with the Help Desk, awareness about phishing campaigns and compromised accounts may actually originate from various sources. In this section we detail how teams become aware of an ongoing campaign by discussing the notification procedures within teams or from outside the ticketing system. Teams may become aware of incidents by receiving notification from a number of internal and external channels, including the team's own monitoring systems, a UK-wide educational network called Janet, and the contracted organisation that handles out-of-hours Help Desk calls. Regardless of the source used to identify an ongoing phishing incident, each team works to ensure that all relevant teams are aware of new attacks so as to maintain organisational situational awareness. These cues result in responsive action which may not be reported to helpline staff, but will result in communication between service teams as they coordinate operations for protection.

#### **Phishing awareness can originate from within teams**

Standard monitoring practices by teams can identify suspicious activity within their systems or through their own personal email inboxes. Once aware, teams proactively adjust their systems to protect users from phishing attacks. For example, Mail Exchange can notice phishing incidents if a compromised user exceeded the limit on sending emails.

[Help Desk] Generally, Network look to see when there is strange activity on the network because they have the network logs. So they have some massive traffic ... So what you got here about quota limit, they see some of the other behaviour around that kind of concept that things like massive downloading from users who wouldn't normally do that and then they will feed it in there and say OK here is the thread and start investigating.

Similar to the Help Desk, the team creates a call based on the observed problem and escalate it to other relevant teams to mitigate their own systems. Maintaining vigilance for anomalies allows teams to react immediately.

### **External partners can inform teams**

Apart of the University IS teams, awareness of phishing incidents might originate from external partners. For example, many Universities in the UK gain access to the Internet through the Janet Network, which is a dedicated network infrastructure for the ‘UK research and education community’<sup>3</sup>. If a Janet admin identifies a serious attack, they may choose to block URLs associated with it UK-wide and notify Security or Network teams directly.

[Quality Enhancement] We have for example a huge issue last year where a lot of different Janet organisations were being attacked and Janet is the education network. What happened is gradually different universities started to get attacked but some universities put proactive steps because they were able to share that information to reach a solution.

Janet also informs the Network team about outgoing suspicious communication, so the network team become aware of compromised accounts or machines.

[Help Desk] They are all interlinked because you could end up getting referral from Janet on the basis that someone clicked on a phishing email and gave out some details which is compromised their accounts and it has downloaded some sort of bot that then start trying to connect to all over the world throw your traffic out and trying to contact other hosts which are known to be malicious.

Another external source of information is NorMAN. NorMAN is an Out-Of-Hours service to support University members 24/7. After the work hours, the Help Desk directs all the calls to the NorMAN support desk. NorMAN cannot solve all the calls because they do not have access to the ticket system, so at the end of their work, they provide a report about resolved and not resolved calls. Then, the Help Desk integrate all the calls to the call system and resolves all the calls which could not be handled by NorMAN.

[Help Desk] NorMAN will try to resolve some of the issues but not all of them because they don’t have access to [ticket tracking] services.

---

<sup>3</sup><https://www.jisc.ac.uk/janet>

In the case of phishing attacks, NorMAN staff reply to users without escalating the calls to other teams. Information about any phishing campaigns is included in the NorMAN support report for the Help Desk so they can prepare for the potentially large number of early morning calls and escalate them to the other teams. During shadowing, a first-line staff retrospectively told us how they learned about an ongoing phishing campaign from the manager relaying a NorMAN report.

### **5.5.5 Providing feedback and guidance to end users**

The above sections highlight the importance of user generated phishing reports in detecting and managing phishing attacks. In this section we look at interactions with users, particularly the use of pre-written standard solutions and impacts of bulk closures on users. Standard solutions may save staff time, but potentially at a risk to user satisfaction which is seen as a necessary component of service quality as well as likely influencing users willingness to report future phishing the encounter.

Users contact the Help Desk to ask about guidance around phishing. For example, a user may be uncertain if an email is real or not and is asking for guidance about if they should respond. They may have already taken some action, such as reporting the phishing to their bank, and are asking if that was the correct course of action. Or the phishing email may be threatening them in some way, such as threatening to shut down their account, so they want reassurances that their account will not be deactivated. Some users even engaged with the phishing before realising that it was fraudulent and are seeking guidance about how to best manage the situation.

Help Desk staff endeavour to respond to every call in a professional way that promotes user satisfaction and resolves any service problems. However, the number of calls and time limitations can impact communicating with users.

#### **Standard solution (SS) design**

While the range of queries can be broad, most users are looking for only a small set of guidance, such as reassurance that the email is indeed phishing, that they have done the right thing by contacting the Help Desk, that they should delete the email, how to ensure that their device is malware free, and how to protect their account if they did interact. The high overlap in needed guidance is a perfect match for a Standard Solution (SS) where several sets of guidance can be written by a qualified person and then re-used. Having an available SS is also helpful for Help Desk staff who may not

themselves be experts in phishing and may feel uncomfortable providing self-written guidance to others. SS also allow them to quickly provide consistent professional guidance with detailed steps.

The benefits of the SSs were acknowledged by Help Desk members we spoke to, with all reporting that they reply to all phishing calls using them. SS were also valued for the time they saved first-line staff. Instead of working 5 minutes on every phishing call and writing a reply to every phishing email they receive, first-line staff can quickly use an SS for common phishing calls.

*SS content design.* Phishing-related SSs were developed by the Help Desk based on the most common reasons for contacting the Help Desk and focus on: 1) thanking users who reported phishing emails, 2) confirming an email is phishing, 3) confirming an email is not phishing, 4) helping users who clicked on links, and 5) informing users about phishing simulations. Contents of phishing-related SS are detailed in Table 5.1.

The wording used in the SS were carefully chosen through a collaboration between the Help Desk and Security teams to ensure that it both matched what users were asking and that the responses were technically accurate.

[Security] And we have occasionally gone to [the Help Desk], the wording you are means the people around being inclined to not report any more. Why you are doing it this way. So, there is some back and forth.

Table 5.1: Phishing Standard Solutions used by the Help Desk when responding to phishing-related calls.

User's query	Summary of SS content
Ask if an email is phishing	Yes it is, staff report informing email team, and provide mitigation steps in case the user clicked on links or opened attachments.
	No it is not, staff manually write the reason(s).
Report phishing	Thank them, staff report informing email team, and provide steps if clicked on links or opened attachments.
Report phishing (simulated attack)	Thank them, inform that it is simulated, educate user, no action required.
Clicked on a link	Ask them to delete the email immediately and follow provided mitigation steps.



*Consistent messaging.* SSs are also helpful because they provide consistent messaging to users which is generally considered an effective approach in public safety communication [305]. The Help Desk manager was a proponent of consistent and accurate communications because they felt it would enable them to develop a trusting relationship with users. Managing relationships with users also partially motivated the creation of multiple SSs addressing different common requests, because they felt that the template variations would signal that the Help Desk is listening to each user and providing custom feedback.

[Help Desk] We focus on the consistency in answering users. It is important to build up the relationship over the years.

[Help Desk] Standard solution is used for specific processes such as phishing. When a user forget their username, we give them technical process.

### **Bulk closing calls has costs**

As discussed earlier in Section 5.5.1, we observed that Help Desk staff typically read the first few phishing calls in a campaign and then assume that the remaining reports with similar subject lines have similar content and should therefore all receive the same SS. They do so to efficiently use valuable first-line staff time. However, bulk closures can result in sounding tone-deaf as well as lead to missing important information.

Some users may be giving or looking for information beyond reporting or inquiring if the email is phishing. The bulk closures can cause these calls to be missed. For example, one of the campaigns was a cyberbully type of phishing attack, where the attacker threatened to share victims' unpleasant secrets with everyone in their address book. After the bulk closure of the emails, a user replied to the SS trying to get an answer to their specific question:

[End-user reply to SS] I asked whether I should be informing police - given that this was an attempt to extort money from me by threats? It is not simply a phishing email.

Bulk closures can also result in missing valuable information from reporting users. For example, the user below noticed that there were different senders in a two week-long campaign.

[End-user] Back in February fake emails of the 'are you on campus?' type were being sent in the name of Miguel, with address miguelxxxx@gmail.com.

I reported this. I just got another one, this time from a different address, miguelyyyy@gmail.com.

The user makes an important point about how the phisher used multiple addresses, which may be helpful to the Mail Relay and Mail Exchange teams. Unfortunately, such emails are at risk of being missed during a campaign because of bulk closure.

## 5.6 Discussion

### 5.6.1 Phishing management is a distributed cognition process

Prior research has found that information service management is, by its very nature, a distributed cognitive (DCog) process [263, 300, 301] that requires hand-offs between distinct teams [294]. In a DCog system the cognitive elements of a problem are solved not by a single person but by offloading them onto an environment made up of a combination of people and technology. Because the technical systems of a large organisation, like a university, are managed by multiple teams who must pass problems between them, they often evidence DCog in their problem solving approaches.

However, that does not necessarily mean that all processes in service management involve DCog as many tasks may only require a single person or small team. In this work we have shown that phishing management, at least in this case, evidences many of the elements of DCog. Phishing touches on many distributed elements of the University's IS operations, people from different teams collaborate to pool their knowledge, unique system perspectives, and technical ability to solve phishing problems which occur on a nearly daily basis. To do so effectively, they also make heavily use of tools like ticketing systems, knowledge bases, jointly maintained lists, and standard solutions, all of which allow them to communicate and share knowledge between members to jointly manage phishing. Ticketing systems allow for quick communication that is archived and visible to all involved teams, enabling joint problem solving. Tools like knowledge bases and jointly maintained lists allow for storing of knowledge for use by other teams or future team members. For example, the maintenance of lists of known compromised accounts by various teams so the Help Desk can accurately respond to users. Standard solutions are also an interesting inter-team example where the Security team worked with the Help Desk to construct user-facing feedback about phishing that is accurate which the Help Desk regularly uses in their communications.

Our work suggests that phishing management might be a DCog system for organisations that divide their IS operations up into different teams and have those teams coordinate through ticketing system like tools. The observation has wider implications

because it allows for a deeper and more nuanced understanding of the workflows than is necessarily presented in this case. Break downs in DCog systems are evidenced partially through challenges that the people involved experience. Such as having a fragmented common ground where each individual has their own set of knowledge that does not necessarily overlap well with the knowledge of others, requiring not only information sharing but also facilitating collaborative operations. In DCog, information also tends to pass through networks of people either directly, or more as we observed, through the passing of a ticket that progressively builds up contextual cues. This dialogue between teams is needed because no one individual has full understanding of the system or the problem, necessitating the information sharing.

By thinking of the problem as a DCog issue, several possible avenues of future work and interpretation open up, some of which we expand on in the following sections. The most obvious space for future work is to look at the role of ticketing systems in supporting the DCog activities around phishing management. These tools were heavily used by staff, but didn't provide an easy way to do things like give everyone access to the full set of phishing reports, or allow for easy highlighting of evidence. Other tools, like the lists of known compromised accounts, were kept separate from the ticketing system making it easier for some teams to miss information, leading inter-team friction.

Phishing is also an interesting DCog issue because of its frequency. Other work on DCog with system administrators [295] talks about how they regularly have to solve a wide variety of issues. Phishing is different in that it is a nearly daily occurrence and while the details of each attack vary, the processes needed to address it are far more stable. Which may mean that the types of support needed here can be more easily built into tools than the standard IT management DCog interactions.

### **5.6.2 The University wants reporting, but can it handle all the reports?**

Everyone across the security community agrees that more phishing reporting is better [1, 246, 247] which is a view that was also shared by University staff as well as the University CISO. It is easy to see where this view comes from since staff reporting is the leading way to learn about security breaches [246]. Humans also have the nice feature that they are not deterministic, which makes their actions harder for attackers to predict and it also means that any sufficiently large campaign is likely to include

several staff who are skilled at phishing detection, or just overly observant that day. Reporting phishing remains one of the best ways IS staff have of finding the attacks that are getting through the automated protections, so of course getting more such reports is important.

The problem though is scale. In order to be effective and identify all phishing that is getting through, all staff need to report any phishing they see. In theory this is a good practice, but it has the side effect of producing a very large number of reports which must be processed by someone. The need to manually process the reports is expensive in staff time. Currently, each report must be processed individually, even if they are bundled together, that still requires a human to individually select each report and then group them. This problem is one that is likely shared by all organisations that manage their phishing reporting through ticket tracking systems. With such a manual process, an obvious future work area is to look at potential automation or human support systems.

Unlike other service disruption type issues that help desks get reports about, phishing reports have a relatively high level of internal similarity. While phishing emails are designed to hide well in inboxes, they stand out as a group in the ticketing system due to the similar subject lines, report times, senders, and URLs. The University Help Desk already makes use of many of these features to manually group such emails. However, important information can also be overlooked due to the number of incoming reports. Phishing emails often have some amount of variance to avoid detection. A ticketing system doesn't have the tools necessary to identify which reports are the most important to look at because they exhibit different features. Reporting users may also be pointing out important information that staff do not have time to look at resulting in a lost opportunity to learn and improve technical defences [267]. For example, pointing out that the emails are still coming in through the filter.

We believe that it is impossible to fully automate the phishing report processing and incident management. Instead we recommend looking into ways to use automation to better support users in managing increasingly complex situations [306]. The use of automation can make the Help Desk's work more efficient while following the best practices [307, 308]. Help Desk staff could use support to better leverage their limited time. Their judgement is needed to both decide if the reported phishing is real [22], and to handle more specific questions users might have. One solution for the large number of reports is to automatically cluster a reports of similar phishing emails as a campaign (e.g. emails in Figure 5.1) and then based on those reports, the system

would auto generate a detailed report about that campaign including the number of reports received and a list of reported compromised accounts. This potential solution can assist the Help Desk so they only need to label the first batch of reported phish and also decide if the escalation is necessary.

There is also some room to look at the effectiveness of AI conversational chatbots to look at the incoming user questions and automatically assign SS based on the question text. Chatbot use has been explored in SOC's to analyse and convey system alerts to security analysts [309] but have not yet been used to generate auto response to security incidents reported by users.

### **5.6.3 Ticketing tools support distributed cognition but have room for improvement**

We observed the critical role played by the ticketing system in facilitating much of the inter-team collaboration. In this situation, the ticketing system itself can be considered an embodied agent with direct impact on the success of the distributed process [297] as it stores the phishing evidence, and it is how the different teams record what they are doing and communicate. It is also how the IRT team monitors what is going on so that they can maintain situational awareness and interject information when needed.

While inter-team collaboration seems reasonably well supported by the ticket tracking tool they use, there are still some potential areas for improvement and future research. In the case of phishing, teams need to see all the incoming reports associated with a campaign to accomplish two tasks: 1) tailor their mitigation to cover all the reported phishing variations, and 2) check if the mitigation has indeed stopped the attack. The second point also includes a need to see future incoming reports.

Ticketing systems are usually built around the standard information service model where a problem might happen, such as a power failure. While a service disruption results in many reports, they are basically identical in terms of information so only one report ever needs to be escalated. Phishing though has some differences because often users each receive a slightly different phishing email. So there is a need to be able to group reports together in a way that is visible to other teams.

Currently the Help Desk spends a great deal of time collecting together similar tickets manually, they also spend time curating tickets to make sure that the ticket they escalate has all the information later teams may need while the suggested solution supports the Help Desk goal to assess the impact of the campaign. An obvious

improvement for the ticketing system would be to leverage the clustering idea to flag a representative number of reports with the features necessary for capturing variations within distinct campaigns. Such a system would save Help Desk time to find the variations, but it would also allow them to escalate in a way where the most useful reports are highlighted to other teams while also giving them visibility of the other reports, indicating the scale of problem. Flagging such information would provide teams with the contextual cues necessary for facilitating dialogues with other teams when further knowledge and resources are needed, aiding collaboration across the distinct systems. As a future research direction, it would also be interesting to investigate the contextual cues necessary for a range of incidents reported to help desk staff, and attempting to address limitations in current systems where break-downs in DCog prevent suitable incident response, not just regarding security incidents.

#### **5.6.4 Best practice, is it helping?**

The University made great efforts to align their structure and practices with industry best practice. They leaned heavily on a framework called ITIL, which provides best practices for IS service management designed to help organisations align practices with business needs. ITIL defines IS services from the customers' point of view to satisfy their needs and to bring value to them without ownership of risks [310]. It provides guidelines rather than rules as it determines 'what should be done' as opposed to 'how it should be done'. Therefore, the implementation of ITIL is different between organisations.

ITIL identifies sets of activities, called processes, that respond to a specific trigger to accomplish specific objectives. The workflow discussed in this paper imply two main processes for handling phishing attacks, namely 'Incident Management' and 'Problem Management'. Incidents are defined as an unplanned event to the service, such as the daily calls to the Help Desk (i.e. queue for printer help, password resets etc.) [310]. In our case, each phishing report, or request for guidance is an example of an incident. The same call can potentially evolve into a 'problem' once the Help Desk receives several different calls regarding that specific attack, which in our case represents the individual phishing campaign. The calls are now considered more critical than the incidents as the impact to the system is greater than one off attacks; and thus, as discussed in Section 5.5.1, should be treated differently. However, given the difficulty in identifying repeating incidents [277] and the need for a quick reaction,

current practices may allow for damage to occur when waiting for more phishing calls before escalation. Another method for triggering a problem should be considered when dealing with phishing to minimise the impact of incidents that cannot be prevented.

The studied University made an effort to develop an environment that encourages phishing reporting in line with guidance [4]. Much of this effort was dictated by ITIL, such as providing details for how to contact the Help Desk on their website, allowing users to contact them via several potential communication channels, providing customised feedback to users, and aiming to quickly resolve users' queries. The Security team also ran campaigns to educate staff and students about phishing and encourage them to report it. However, the number of phishing reports the Help Desk receives is relatively low considering the size of the University. Similar to previous research [19], we noticed some of the reports come to the Help Desk because reporters wanted help in determining the safety of an email, they already suspected that the sender was spoofed or that the email looked too sophisticated.

While prior research attempted to explore the effectiveness of ITIL in general IS operations [275,276], our observations imply that the ITIL framework might not fully fit the workflow of phishing handling. Further research is needed to understand how organisations handle phishing while adhering to the ITIL framework and what barriers might arise from using such a framework.

## 5.7 Limitation

The case study may be suggestive of the situation of organisations but generalising the results requires further research. The case study looked at an academic institution that likely differs from other sectors. Universities also have a wide range of internal structures, so while this case is interesting and instructive, other Universities likely have different structures and may be impacted by things like their size and how centralised IT services are. However, we argue that many aspects of this case have similarities with other organisations; for example, using ticketing systems is quite common across sectors. We therefore believe that many of our high-level findings may be useful in future work around how to better support how IT handles phishing reports.

Both interviews and observations were used to collect data. While observations allow the researcher to observe work practices directly, interviews with participants are complimentary, gaining retrospective accounts of events that have happened across a wide time frame and validating observations made. That said, retrospective interviews

are known to be somewhat biased towards memorable events such as particularly impactful phishing campaigns, which may have caused us to over-sample these events. Interviews also suffer from social desirability bias where participants may provide a version that does not fully reflect reality. To partially counter this issue, we asked every team about what they thought the other teams do and detail communication between them. We also attempted to provide validation of interviews through analysis of calls in the ticketing system. Due to limited access to the system and the use of other communication channels, we were not able to see all interactions between teams.

## 5.8 Conclusions

We explored the process of handling phishing incidents in a large University using a combination of interviews and observations. The University uses industry best practices aligned with ITIL to efficiently react to and prioritise incidents based on their potential impact. One observation is that large phishing campaigns can result in many reports which overwhelm Help Desk staff, making it challenging for them to respond individually to each report. We also find that the Help Desk operates as a kind of report triage, shielding third line staff, such as those that manage the email relays, from being inundated by reports that may not have the data they need to take action. Similar to earlier works [295], we also find that communication among staff in different teams is a key aspect of coordinating phishing attack mitigation. We also believe that managing phishing reports is an example of distributed cognition, where the different teams work together through the ticketing system to coordinate solving a multi-system problem. We believe that although it is impossible to fully automate phishing response, there is potential to better support IT staff through improved tools that allow them to handle the scale and complexity of phishing attacks.





# Chapter 6

## Discussion and Conclusion

In this thesis, two aspects of phishing defences were explored. The first was the design of a URL feature report. To better understand the various features involved in accurate identification of URLs as phishing, I conducted a literature review examining the features used in automation and the features used in human-based detection. I found that only a small number of automated detection features were used in the human-support literature, these notably being features that reveal information beyond the URL simple structure. The review also produced a comprehensive list of features, along with information concerning how human friendly each feature might be. I used this list as a starting point for my next project to support users in reading URLs. To this end, a URL report was created to include a wide variety of necessary features in an understandable format, the initial design of which was inspired by earlier privacy policy work [209]. This report was then iteratively refined with experts and average users. The visual appearance of the report was tested on a wide range of safe and phishing URLs. Next, its effectiveness in assisting users in accurately judging URLs was tested with a group of online participants. The main finding of these studies was the importance of supplying users with simple and inclusive information to help them accurately judge URLs.

The second study took a broad view of combating phishing in an organisational context using an ethnographic method. It revealed that teams communicate in a distributed cognition fashion, with communication occurring amongst teams, end users and artefacts. In distributed cognition, artefacts that enable the storing and sharing of knowledge are a key element of how groups communicate and support the distributed nature of their work. In this research, ticketing systems and procedures were observed to be artefacts that support the communication between the IS teams, making them key to successful phishing mitigation. In this work, I found many challenges related to

successfully applying procedures as well as many places where the volume of phishing reports negatively impacted other workflows, suggesting the potential usefulness of automation in this area.

## 6.1 Discussion

Educating users about phishing features is not sufficient in cases where the techniques used are more complex [26, 61]. For example, if users are not taught about the difference between domains and subdomains, they will not be able to recognise phishing URLs with a brand name in the subdomain [8]. My review in Chapter 3 highlighted that URLs are complicated, meaning that finding a small set of common and reliable features with which to judge URL safety is difficult (bordering on impossible). Moreover, it is not realistic to teach users about features like short links without providing tools to expand such URLs: realistically, users will need to click on short links in some cases, so advising them to avoid all such links is not helpful.

One lesson learned over the course of this thesis involves the need for real-time support that leverages a wide set of available features without requiring users to memorise them all, such as that designed in Chapter 4. Helping average users access relevant information can enable them to make more accurate decisions by teaching them how to leverage their contextual knowledge to identify discrepancies between the facts that the report presents and facts that only they know. Such tools also give users more options and flexibility. Wash et al. [61] observed that experts could identify phishing partially as a result of spotting ‘odd’ details in communications but also because they possessed the ability to compare their expectations with reality – in part through skill at, for example, accurately reading URLs but also through their knowledge of support tools that allow them to look up information like an IP address, which enables them to follow through with their concerns and self-identify whether the concern is valid. Indeed, simplifying technical information can reduce barriers to learning technical details and give users control over the decisions that are in conjunction with their own risk model [226].

One observation supported throughout this thesis is the potential benefit of the URL report for users who are faithfully following flawed heuristics, such as considering simple URLs to be safe [75]. An educational report would help them accurately understand URL destinations and offer a source of fast feedback that may enable them to create more accurate and useful heuristics. Such reports also have the potential to

assist users who need fast feedback and do not want to contact the Help Desk due to the expectation that doing so will result in a delayed and generic response [24]. That said, one RQ that should also be asked involves the impact that such a report could have on the Help Desk's call load: phishing reports are necessary to IS teams so they can understand the impact of such as attack, but reporting a phishing attempt is not the same as asking whether an email is safe. The latter requires far more manual work from the help desk, whilst the former is more of a drop box than a back-and-forth conversation. Indeed, in light of the above, further work is required to establish the effectiveness of using report-like interventions in organisations to help overloaded staff whilst also collecting considerable numbers of phishing reports.

Similar to prior IT research [294], there is an effort to take the complex set of tasks associated with IT management and break them down into different teams to allow those teams to build a depth of knowledge in their given areas. This separation, however, makes phishing management challenging: information must be passed between teams to enable a coordinated response. I observed that this coordination was an example of distributed cognition. However, the case of phishing differs somewhat from other IT issues, since end users play a much larger role in the distributed cognition here than they might in other cases. Phishing is mostly reported by end users, and so identifying the range of phishing requires end users to report phishing attacks and to provide useful information in those reports – for example, the *.eml* files in which they noticed the deception taking place. In many ways, this makes end users an important part of the distributed cognition process. They are not simply reporters, as they provide vital data and commentary on whether the implemented solutions are working properly. Their actions also impact the training given by the university by allowing the university to understand whether users have successfully learned how to spot phishing or if they followed the proper steps when they did fall for phishing attacks. In sum, since users are an important part of this distributed cognition system, it may be interesting for future work to consider them in that capacity when building tools to improve phishing management and when thinking about the type of feedback to give to users.

The observation that handling phishing reports is a distributed cognition system also allows for potentially broader impacts and considerations. Whilst users are clearly a component of the phishing process, this may also turn out to be true for other security and privacy issues related to IT management. Password resets [311], for example, also involve users and IT managers working together, often through shared artefacts like policies, tools and guidance. Whilst a less compelling example than the back-and-

forth required to address phishing issues, it may be interesting to begin thinking about other places where users play an active role in how IT management is performed and how to best support the entire system, rather than thinking of users and administrators as separate groups.

One lesson emphasised throughout this thesis is the need for more phishing-related support, both for individuals (as discussed above) and organisations. The tools used in IS systems are generalised to handle the majority of IT-related incidents. Whilst I observed how the organisation at hand adapted its practices to fit its resources, the benefits of reporting phishing emails were limited due to a lack of tools (Section 5.6). On a similar note, improving reactive measures has the potential to protect everyone within the organisation, as some people are highly susceptible to phishing attacks and continue to click on phishing links even after training [84]. A first step toward supporting IS teams in making use of every phishing report is to improve ticketing tools. Whilst fixing this problem requires many components, as many teams are involved in handling phishing incidents, it would positively impact staff time. As reported by Avanan, staff in the Security Operation Centres spend approximately 22% of their time dealing with phishing emails, so supporting them would potentially reduce the time they spend on this task [312]. The case study in Chapter 5 helps to understand this problem and set the requirements to solve it, which will be discussed in Section 6.2.

The incident response workflow needs to better contextualise problems so that research groups (e.g. machine learning, software engineering) can help to solve them. However, researching IS processes and procedures is challenging due to the busyness of the IS staff and the sensitivity of the information they might share [313–315], which may block the initiation of such research. Over the course of the case study, I found that, although IS teams were welcoming and friendly, they were different to individually recruited participants. I also observed that some staff were hesitant to share information, so the periodic review of findings positively impacted participants' willingness to share more information, especially during the observation of the help desk staff. I found that showing them how the data collected was interpreted has inspired them to show more cases, which enabled me to better understand their procedures. Similar to previous research [315], this observation highlights the importance of involving participants in reading, discussing and approving the final result.

## 6.2 Future Work

As outlined in Chapter 2, existing research has largely focused on what causes users to fall for phishing, as well as different ideas regarding how to better support them (e.g. through training). However, there is a large research gap around the phishing ecosystem: we know very little about many ecosystem elements, such as what the case study (Chapter 5) highlights. Suggested directions for future research include understanding how people interpret feedback from the help desk and investigating its impact on users' future reporting behaviour and exploring the best interventions for encouraging and facilitating phishing reporting, such as using machine learning techniques to customise automatic responses to reporters. Surprisingly, during the case study, the majority of the reports handled were part of campaigns that had already been dealt with; thus, we know little about how individual support staff judge those reports and what resources they need. Retrospective interviews following the critical decision method was effective in understanding how experts detect phishing spam [61]; therefore, such a method can help understand how an individual staff member makes a decision regarding the safety of reported phishing emails, as well as how staff work together as a group to identify a phishing campaign. An additional interesting point suggested by our focus group participants would be to establish whether the URL report can be useful to the help desk if integrated into its reporting process. Specifically, IT experts use non-conclusive features to make decisions about email safety [61], making the report potentially suitable for this user group. Focus group sessions with the help desk should explore the benefits of using such a report and the best methods to adjust it to fit their needs.

Prior work has found that users require professional advice regarding communications about which they are suspicious but unsure [244]. However, to be highly effective, advice requires a fast response from experts [24]. The prototype of the URL report in Chapter 4 is an example of a type of support that can be automatically generated and therefore provided to users quickly. However, I was not able to test the report longitudinally in this thesis. An obvious potential direction for future work would be to do so and explore the long-term effects of using such a report. Such work could explore several issues, such as how to better integrate the report into users' workflow, possibly by implementing it as a plugin or a mobile app, which were suggested by our focus group participants. These approaches would work well in the context of a long-term study to understand the impact of the long-term usage of the report on users'

safety assessments. Moreover, whilst I focused on URLs, this kind of support can be extended in future work to simplify other types of technical information, such as email headers, attachments and other hard-to-fake indicators that are used by experts [61].

Whilst the security community encourages reporting of phishing emails, we noticed that it was challenging for IS staff to make use of every phishing email report due to issues like volume of reports. Reporting is important, but it is not effective if staff are only able to use a fraction of the campaign to design their remediations. Reviewing every phishing report requires extra human effort, and organisations are aware that their time and resources are limited; hence, they weigh benefits against risks when dealing with security incidents [39]. Although some phishing campaigns are simple and easy to block, others are sophisticated enough that they should be investigated thoroughly in order to identify their important characteristics [108]. My results suggest that there is plenty of room for support through automation. IS staff could use assistance in grouping together similar reports so that they can respond to all users at once rather than grouping these reports manually, thereby saving time. They also need to determine whether new, similar attacks are continuing to come in so that they can identify whether remediation is working. Many analysts find automation to be an effective solution for the majority of incident response challenges [108], and I agree that it has potential to help here. Specifically, machine learning approaches could be used to cluster emails based on their characteristics, thus resulting in groups of campaigns, as observed. By automatically clustering phishing reports, the system can provide the various involved teams with customised support that is tailored to their roles whilst still linking back to the same set of evidence (i.e. reports). For example, the help desk needs to respond to everyone who reports a phishing attempt, which takes time, and they tend to wait until phishing reports stop coming in before they write the response. This delay makes the process easier for the help desk but means that their advice comes late for end users. Automatically clustering phishing reports can enable the help desk to write a response for a given cluster and then have the system automatically send that response to all reports matching that cluster, allowing for faster response time. The system could also assist help desk staff in constructing more customised responses based on a combination of the content of the reported phishing and prewritten standard solutions – for example, not including the advice ‘don’t click links’ if no links exist. Similarly, clustering could help third-line teams, which need to take reactive measures in response to phishing. Currently, the help desk only escalates one representative report, and combing through that report to find useful common features takes time. With

automatically created clusters, the help desk could instead escalate a cluster, which would automatically include new reports as they come in. The system could also look for commonalities and differences across the reports and highlight them to the team – for example, if the emails all came from the same ‘from’ address or all passed through the same relay. In short, there are many ways that automatic clustering of reports could be used to support staff and improve end users’ experiences.

If such a system could be built, it would be interesting to test it with IS teams and explore its impact on the challenges we observed in Chapter 5. Another interesting study would be to investigate the role of such a system in the distributed cognition observed in the phishing-handling workflow.

### **6.3 Limitations**

In this thesis, I focused on examining how to provide support to individuals and understand the workflow of phishing reports. However, the phishing ecosystem literature described earlier in chapter 2 has many gaps that are important to explore, some of which are recommended for exploration in future work, as discussed above.

The thesis’s findings indicate the level of support that would be helpful for both users and IS staff, but there are also limitations to consider. In this work, I conducted focus groups, online studies and an ethnographic study. However, a longitudinal study is needed to test the results at scale. The URL feature report, when tested, significantly improved user ability to judge URL safety, but I did not have the chance to implement it and see how it impacted people over a longer time frame. Additionally, in this thesis, I showed that there is a need for support, but I could not build such support, as doing so would require many components and extensive work. That said, the findings I present here suggest that such an amount of work will be worthwhile, since it has the potential to help all IS teams involved in the process of phishing handling.

### **6.4 Conclusion**

In this thesis, I presented the results of three studies, two of which aimed to explore URLs’ phishing features to support humans. I found that reading URLs is not an easy task and that the current features used in user trainings are not sufficient. For example, whilst hosting websites is efficient for small businesses and personal websites, phishers use them to bypass filters, as phishing websites inherit the host features. Thus, I



proposed a URL report that provides real-time information about a given URL. The design aims to empower users to compare the information to the context in which they receive the URL (e.g. the web-hosting possibility for reputable companies against personal websites). By testing the report's visual appearance and participants' abilities to comprehend it and use it accurately, I showed that automated assistance can help users when they are in doubt and do not have access to experts. Although the report is similar to the other user support tools described in Section 2.2, there are several areas in which they differ: namely, the report's information is understandable to average users and can efficiently be used with their current background knowledge [316].

The third study investigated the phishing response workflow in an academic organisation and sought to understand the challenges that organisation faced when dealing with reported phishing emails. The findings suggest that distributed cognition is an essential element of the successful processing of phishing reports. The significant perceived difficulty in our case was the unworkable number of phishing reports: whilst the organisation adapted its procedures to fit its resources, the study's findings suggest the potential for automation.

This research can benefit researchers focused on phishing, since it centres on several points of the phishing life cycle and suggests future directions. The phishing ecosystem needs to be viewed as a whole when considering a viable solution.

# **Appendix A**

## **Survey on the URL Report**

## Consent

### Study Information

#### What will I be asked to do:

You will be asked to fill out a survey from a laptop, tablet or desktop device. The survey is approximately 30 minutes in length. The survey is primarily composed of a sequence of links (URLs) and an interface designed to help you understand where the links will lead. You will be asked to look at the interface and use it to answer questions about the links. You do not need to visit any of the links, only look at them. You will also be asked some basic demographic questions. Questions marked with a red star are mandatory - you will need to answer them in order to complete the survey. You can stop taking part in this survey at any time.

#### How will my data be used:

The purpose of the research is to provide more support for people who are uncertain about the safety of various communications. The data you provide will be used to improve anti-phishing tools and generally help people better understand potentially malicious communications sent to them. Provided data will be stored on the Qualtrics platform and secure computers at the xxxxxxxxxxxxxxxx. Anonymised data with personally identifying features, such as IDs and identifying comments, removed may be shared with other researchers or stored on other platforms.

#### Compensation:

You will receive £ 3.50 as compensation for completing this survey along with the knowledge that you have helped science.

#### Data protection rights:

Your data will be processed in accordance with Data Protection Law. The xxxxxx xxxxx is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit [www.ico.org.uk](http://www.ico.org.uk). Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at xxxxxxxxxxxxxxxx.

**Ethics:**

This project complies with the ethics procedure at the xxxxxxxxxxxxxxxxxxxxxxxx  
xxxxxxxxx (tracking number xxxxxxx).

If you have any questions or concerns, please contact either xxxxxxxi  
(xxxxxxxxxxxxxxxxxxxxx) or xxxxxxxxxxxxxxxxxxxx (xxxxxxxx). If you have  
any complaints you should contact the ethics panel (xxxxx).

I attest that I am over 18 years old, have read the above description, and wish to take  
part in the survey.

- Yes, take me to the survey.
- No, do not continue with the survey.

**Familiarity**

1. How familiar are you with the following terms about websites?

	Extremely familiar	Very familiar	Moderately familiar	Slightly familiar	Not familiar at all
Page Rank	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Web hosting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Toadfish	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Domain	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
URL redirection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Domain age	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
IP address	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Blacklist	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Search Engine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Domain popularity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Domain location	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
URL manipulation tricks	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Website category	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2. How familiar are you with each of the following organisations?

Qualtrics Survey Software

	I have heard of it and I have an account with them.	I have heard of it and I previously had an account with them.	I have heard of it but I don't have an account with them.	I have not heard of it.
BestChange	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Facebook	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tripod	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Microsoft	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bittrex	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
eBay	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### URL Warning

For the following questions you will be given information on an organization along with a link (URL).

You will be asked to use the provided information to determine if the URL really does go to that organization or not. All needed information to decide is provided.

Some of the links (URLs) you will see were once real attack URLs. We have made efforts to only select those that are now blocked by most browsers. But please, do not type them in to your browser for your own safety.

3. What should you do in the following questions?

- Read provided information and answer questions.
- Type in URLs and answer questions.
- Click on URLs and answer questions.

**4. If not “read provided information and answer questions” chosen in Q3, show this and return to Q3.**

**You have chosen incorrectly, please re-read the text provided and choose the correct option.**

### URL Prior Knowledge

4. What website does this URL lead to?

<https://paypal.com/google.com>

- Paypal's websites
- Amazon's website
- Google's website
- Other

5. What website does this URL lead to?

<https://facebook.profile.com>

- Profile's website
- PayPal's website
- Facebook's websites
- Other

6. What website does this URL lead to?

<https://www.nytimes.com:443/>

- 443 website
- The New York Times' website
- www's website
- Other

## Beginning of Branching

The following part of the survey includes questions based on the condition. Each condition is represented by two blocks as different sets of URLs are presented. Therefore, we have the following six blocks.

### Block 1: Summary Groups 1 and 2

1. Summary Judgment Group
2. Summary Understanding
3. Report Usability

### Block 2: Full Report Group 1 and 2

1. Priming
2. Full Report Judgment Group
2. Full Report Understanding
3. Report Usability

### Block 3: Control Groups 1 and 2

1. Control Judgment Group
2. Control Understanding

The different sets of questions which are included in each block, e.g., Priming or

## Auto filled information in all the conditions


The questions in this sub-block are presented in a loop. The participant is shown six different images and scenarios, but has to answer the same set of questions for each. The scenario text is adapted grammatically by replacing placeholders with content from the following table.

Group2	Group 1	Company Description	Company name
<a href="https://resolutioncenter.ebay.com/policies/?id=123">https://resolutioncenter.ebay.com/policies/?id=123</a>	<a href="https://itmurl.com/www.ebay.co.uk/item=30327559652">https://itmurl.com/www.ebay.co.uk/item=30327559652</a>	An auction and consumer to consumer sales website	eBay
<a href="https://www.bestcnange.ru/exchangers/mkt=en&amp;id=234">https://www.bestcnange.ru/exchangers/mkt=en&amp;id=234</a>	<a href="https://www.bestchange.ru/exchangers/mkt=en&amp;id=234">https://www.bestchange.ru/exchangers/mkt=en&amp;id=234</a>	An electric currency trading (e.g. Bitcoin)	BestChange
<a href="https://www.365onmicrosoft.com/login/?lang=en-GB">https://www.365onmicrosoft.com/login/?lang=en-GB</a>	<a href="https://email.microsoftonline.com/login/?mkt=en-GB">https://email.microsoftonline.com/login/?mkt=en-GB</a>	A computer software and hardware builder	Microsoft
<a href="https://international.bittrex.com/account/?id=2423">https://international.bittrex.com/account/?id=2423</a>	<a href="https://international.bitrèx.com/account/?id=2423">https://international.bitrèx.com/account/?id=2423</a>	An electric currency exchange (e.g. Bitcoin)	Bittrex
<a href="https://fb.me/messages/t/788720331154519">https://fb.me/messages/t/788720331154519</a>	<a href="https://l.facebook.com/l.php?u=http%3A%2F%2F67.23.238.165">https://l.facebook.com/l.php?u=http%3A%2F%2F67.23.238.165</a>	A website to connect and share with other people.	Facebook
<a href="https://webmasterq.tripod.com/pricing?plan=free-ad">https://webmasterq.tripod.com/pricing?plan=free-ad</a>	<a href="https://www.tripod.lycos.com/pricing/?plan=free-ad">https://www.tripod.lycos.com/pricing/?plan=free-ad</a>	A free website builder.	Tripod

### Summary Judgment Group

7. Imagine that you want to visit \${Company name}, which is \${Company description}. Using the report provided below, do you think the URL: "[\\${Company URL}](#)" leads to a page owned by the above company or is it a malicious URL?

[See the report in a new tab](#)

 **The image name corresponds to the image shown to the participant, as displayed in the following list of images.**

- Malicious
- Trustworthy

8. How confident are you in your decision above?

- Not confident at all
- Less confident
- Somewhat confident
- Confident
- Very confident

9. Indicate the **three** features that most **influenced** your decision about if the URL is malicious or trustworthy Note that some features may not appear on some interfaces.

- Manipulation tricks
- Domain
- Domain popularity
- Domain age
- My own prior experience reading URLs
- Colours (Red, Green, Yellow)
- Search result
- Host multiple sites
- Goes to

10. Any comments (optional):



## Report Summary

<https://www.bestchange.ru/exchangers/mkt=en&id=234>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>1</b>	<b>No Match</b>	<b>1 month</b>	<b>Low</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

Best Change - Phishing

## Report Summary

<https://www.bestchange.ru/exchangers/mkt=en&id=234>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>0</b>	<b>Partial match</b>	<b>12 years</b>	<b>High</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

Best Change - Safe

## Report Summary

<https://international.bittrex.com/account/?id=2423>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>1</b>	<b>No Match</b>	<b>5 months</b>	<b>Low</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

Bittrex - Phishing

## Report Summary

<https://international.bittrex.com/account/?id=2423>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>0</b>	<b>Partial match</b>	<b>5 years</b>	<b>High</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

Bittrex - Safe

## Report Summary

<https://resolutioncenter.ebay.com/policies/?id=123>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>0</b>	<b>Match</b>	<b>24 years</b>	<b>High</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

Ebay - Safe

## Report Summary

<https://itmurl.com/www.ebay.co.uk/item=30327559652>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>0</b>	<b>No Match</b>	<b>1 month</b>	<b>Low</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

Ebay - Phishing

## Report Summary

<https://l.facebook.com/l.php?u=http%3A%2F%2F67.23.238.165>

Goes to:

[https://67.23.238.165/&h=AT3y8f6\\_tofRPamyd-PMIIZT9D64tAzKlzhBZ0zDRZVTibJU...](https://67.23.238.165/&h=AT3y8f6_tofRPamyd-PMIIZT9D64tAzKlzhBZ0zDRZVTibJU...)

⚠ We cannot guarantee the safety of danger of this link.

**Used Manipulation tricks**

**1**

**Search Result**

**No Match**

**Domain Age**

**No data**

**Domain Popularity**

**Low**

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

Facebook - Phishing

## Report Summary

<https://fb.me/messages/t/788720331154519>

Goes to:

<https://www.facebook.com/>

⚠ We cannot guarantee the safety of danger of this link.

**Used Manipulation tricks**

**0**

**Search Result**

**Match**

**Domain Age**

**22 years**

**Domain Popularity**

**High**

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

Facebook - Safe

## Report Summary

<https://email.microsoftonline.com/login/?mkt=en-GB>

⚠ We cannot guarantee the safety of danger of this link.

Used Manipulation  
tricks

1

Search Result

No Match

Domain Age

17 years

Domain  
Popularity

High

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

Microsoft - Safe

## Report Summary

<https://www.365onmicrosoft.com/login/?langua=en-GB>

⚠ We cannot guarantee the safety of danger of this link.

Used Manipulation  
tricks

1

Search Result

No Match

Domain Age

1 month

Domain  
Popularity

Low

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

Microsoft - Phishing

## Report Summary

[https://www.tripod.lycos.com/pricing/?plan=free\\_ad](https://www.tripod.lycos.com/pricing/?plan=free_ad)

**⚠ We cannot guarantee the safety of danger of this link.**

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>0</b>	<b>Partial match</b>	<b>24 years</b>	<b>High</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

Tripod - safe

## Report Summary

[https://webmasterq.tripod.com/pricing?plan=free\\_ad](https://webmasterq.tripod.com/pricing?plan=free_ad)

**⚠ We cannot guarantee the safety of danger of this link.**

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>0</b>	<b>No Match</b>	<b>Multiple sites hosted here</b>	<b>Multiple sites hosted here</b>

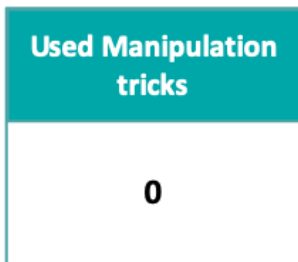
Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

Tripod - Phishing

## Summary Understanding

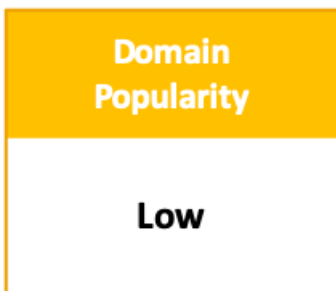
In this section, we will show you parts of a report and we will ask you to chose the interpretation that makes most sense for you.

11. What does "Manipulation tricks" mean?



- The Internet community uses these tricks to confuse attackers.
- An attacker has crafted the URL using these tricks to confuse people.
- The organisation uses these manipulations to highlight how good they are at security.
- The description is confusing.
- Other

12. What does "Domain popularity" mean?



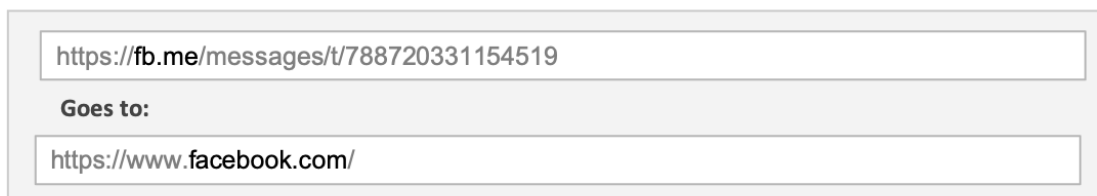
- It is based on the number of other popular websites that are linked to this website.
- It is based on the number of people who visit this website compared to other websites.
- It is based on the quality of the site compared to similar sites.
- The description is confusing.
- Other

13. What does "Multiple sites hosted here" mean?



- This domain belongs to an organisation that allows other people to create their own websites.
- This domain is known to host malicious websites.
- The domain is a discussion forum where people can post links to other websites.
- The description is confusing.
- Other

14. What does the following mean?



- The report is recommending that I visit the second URL instead because it is safer.
- When I click on the first URL (fb.me), the second URL (facebook) will open.
- The first URL sends personal data to the second URL.
- The description is confusing.
- Other

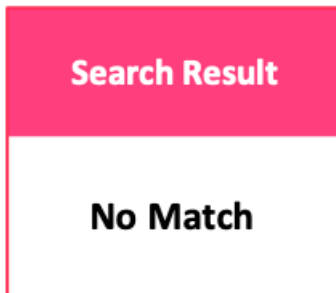


15. What does "Domain age" mean?



- It indicates how long ago this URL was reported as phishing.
- It indicates how long the organisation has had an online presence at this domain.
- It indicates how long ago the physical organisation was established.
- The description is confusing.
- Other

16. What does "Search result" mean?



- This exact URL does not appear in any list of malicious URLs.
- This exact URL does not appear in any list of safe URLs.
- This exact URL does not appear in Google's top search results.
- The description is confusing.
- Other

## Report Usability

17. Please tell us how strongly you agree or disagree with the following statements.

Qualtrics Survey Software


	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
It is hard to find information in the report.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can learn a lot about phishing using this report.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
If this report were easily accessible via a website, I would use it when I see a suspicious communication.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Completing this survey taught me something useful about URLs.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am confident of my understanding of the report.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would imagine that most people would learn to use this report very quickly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found the report very cumbersome to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I needed to learn a lot of things before I could get going with this report.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

18. Any comments (optional):

### Full Report Judgement Group

19. Imagine that you want to visit "\${Im://Field/2}", which is "\${Im://Field/5}". Using the report provided below, do you think the URL: "[\\${Im://Field/3}](#)" leads to a page owned by the above company or is it a malicious URL?

[See the report in a new tab](#)

 **The image name corresponds to the image shown to the participant, as displayed in the following list of images.**

- Malicious
- Trustworthy

20. How confident are you in your decision above?

- Not confident at all
- Less confident
- Somewhat confident
- Confident
- Very confident

21. Indicate the **three** features that most **influenced** your decision about if the URL is malicious or trustworthy. Note that some features may not appear on some interfaces.

- Manipulation tricks
- Domain
- Domain popularity
- Domain age
- My own prior experience reading URLs
- Colours
- Search result
- Host multiple sites
- Goes to

22. Any comments (optional):

## Priming

Please read the following report carefully and answer the questions below.

NOTE: The information provided in this report is fictional. We use the example to see how readable the report design is.

### Report Summary

https://readingability.com/1vcs37

**Goes to:**

https://www.letmebeyoureyes.com/users/userID-12975/en/

⚠ We cannot guarantee the safety of danger of this link.

Used Manipulation tricks

0

Search Result

No Match

Domain Age

28 years

Domain Popularity

Moderate

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

### Manipulation Tricks



Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

No Tricks Used

Of the known tricks attackers use, none appear in this URL.

### URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	letmebeyoureyes.com
<b>Location</b> The physical address where the domain owner claims they live or conduct business.	Wonderland
<b>Categorization</b> Indicates the type of this website, such as shopping or travel. New or phishing domains may not be categorized.	Fiction
<b>Domain Age</b> The date when the domain was first registered.	22-08-1990 <b>28 years</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>does not match</b> your URL: https://www.youtube.com/account/....

[See the report in a new tab](#)

23. The popularity of the page the URL leads to is:

- Popular
- Somewhat popular
- Not popular
- Not found in the report
- Other

24. According to the report, what is the webpage that will open once you clicked on the link (URL)?

- <https://www.youtube.com/account/activate>
- <http://www.letmebeyeureyes.com/users/userID-12975/en/>
- <http://readingability.com/1vcs37>
- <https://www.amazon.com>
- Not found in the report
- Other

25. Where is the owner of this domain located?

- Australia
- United States
- Wonderland
- United Kingdom
- Not found in the report
- Other

26. Which of the following is the domain?

- yahoo.com
- letmebeyeureyes.com
- youtube.com
- Not found in the report
- Other

27. The popularity of the domain the URL leads to is:

- Popular
- Somewhat popular
- Not popular
- Not found in the report
- Other

28. Does this domain host other websites?

- Yes
- No
- Not found in the report
- Other

29. According to the report, if you were to search Google for the URL, what would be the top result?

- <https://www.youtube.com/account/activate>
- <https://www.amazon.com>
- <http://www.letmebeyoureyes.com/users/userID-12975/en/>
- <http://readingability.com/1vcs37>
- Not found in the report
- Other

30. How old is the website domain?

- Less than month
- 1-6 months
- 6-12 months
- More than a year
- Not found in the report
- Other

31. How many manipulation tricks does this URL have?

- 0
- 5
- 8
- 11
- Not found in the report
- Other

32. Which category does this URL belong to?

- Information Technology
- Fiction
- Shopping
- Business
- Not found in the report
- Other

## Report Summary

<https://www.bestchange.ru/exchangers/mkt=en&id=234>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>1</b>	<b>No Match</b>	<b>1 month</b>	<b>Low</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue



## Manipulation Tricks

Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

<b>Similar to a popular domain:</b> 'bestchange.ru' is similar to popular domain "bestchange.ru".	<a href="https://www.bestchange.ru/">https://www.bestchange.ru/...</a>
--	--

## URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	bestchange.ru
<b>Location</b> The physical address where the domain owner claims they live or conduct business.	GDPR masked
<b>Domain Age</b> The date when the domain was first registered.	16-08-2019 <b>1 month</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>does not match</b> your URL: <a href="http://uniespro.blogspot.com/">http://uniespro.blogspot.com/</a>

Best Change - Phishing



## Report Summary

<https://international.bittrex.com/account/?id=2423>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
0	Partial match	5 years	High

Color Code: ✖ Known Issue ! Possible Issue ✔ No Issue

## Manipulation Tricks


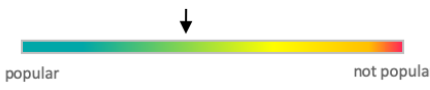
Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

### No Tricks Used

Of the known tricks attackers use, none appear in this URL

## URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	bittrex.com
<b>Location</b> The physical address where the domain owner claims they live or conduct business.	United States
<b>Categorization</b> Indicates the type of this website, such as shopping or travel. New or phishing domains may not be categorized.	Finance and Banking
<b>Domain Age</b> The date when the domain was first registered.	2014-01-25 <b>5 years</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	 popular not popular
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	 popular not popular
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>partially match</b> your URL: <a href="https://international.bittrex.com/">https://international.bittrex.com/..</a>

Bittrex - Safe

## Report Summary

<https://resolutioncenter.ebay.com/policies/?id=123>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>0</b>	<b>Match</b>	<b>24 years</b>	<b>High</b>

Color Code: ✖ Known Issue ! Possible Issue ✔ No Issue

## Manipulation Tricks



Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

### No Tricks Used

Of the known tricks attackers use, none appear in this URL

## URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	ebay.com
<b>Location</b> The physical address where the domain owner claims they live or conduct business.	United States
<b>Categorization</b> Indicates the type of this website, such as shopping or travel. New or phishing domains may not be categorized.	Shopping
<b>Domain Age</b> The date when the domain was first registered.	1995-08-04 <b>24 years</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	 popular not popular
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	 popular not popular
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>fully match</b> your URL: <a href="https://resolutioncenter.ebay.com/">https://resolutioncenter.ebay.com/</a>

Ebay - Phishing

## Report Summary

<https://fb.me/messages/t/788720331154519>

Goes to:

<https://www.facebook.com/>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>0</b>	<b>Match</b>	<b>22 years</b>	<b>High</b>

Color Code: ✖ Known Issue ⚠ Possible Issue ✔ No Issue

## Manipulation Tricks



Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

### No Tricks Used

Of the known tricks attackers use, none appear in this URL

## URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	facebook.com
<b>Categorization</b> Indicates the type of this website, such as shopping or travel. New or phishing domains may not be categorized.	Social Networking
<b>Domain Age</b> The date when the domain was first registered.	1997-03-29 <b>22 years</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	 popular not popular
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	 popular not popular
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>Fully match</b> your URL: <a href="https://facebook.com/">https://facebook.com/</a>

Facebook - Safe

## Report Summary

<https://www.365onmicrosoft.com/login/?langua=en-GB>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>1</b>	<b>No Match</b>	<b>1 month</b>	<b>Low</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue



## Manipulation Tricks

Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

<b>Similar to a popular domain:</b> '365onmicrosoft.com' is similar to the popular domain "microsoft.com".	<a href="https://www.365onmicrosoft.com/...">https://www.365onmicrosoft.com/...</a>
---	---

## URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	365onmicrosoft.com
<b>Location</b> The physical address where the domain owner claims they live or conduct business.	Canada
<b>Domain Age</b> The date when the domain was first registered.	14-08-2019 <b>1 month</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	 popular not popular
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	 popular not popular
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>does not match</b> your URL: <a href="https://social.microsoft.com/...">https://social.microsoft.com/...</a>

Microsoft - Phishing

## Report Summary

[https://webmasterq.tripod.com/pricing?plan=free\\_ad](https://webmasterq.tripod.com/pricing?plan=free_ad)

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
0	No Match	Multiple sites hosted here	Multiple sites hosted here

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue


## Manipulation Tricks

Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

<b>No Tricks Used</b> Of the known tricks attackers use, none appear in this URL
---

## URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	tripod.com
<b>Location</b> The physical address where the domain owner claims they live or conduct business.	United States
<b>Domain Age</b> The date when the domain was first registered	⚠ <b>Note:</b> tripod.com provides web space for other people to put their webpages online. So the age and popularity of the site cannot be accurately determined.
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	<b>Usually only small companies and personal websites are hosted by other domains.</b>
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	 popular <span style="float: right;">not popular</span>
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>does not match</b> your URL: <a href="https://promaster.com/Category...">https://promaster.com/Category...</a>

Tripod - Phishing

## Report Summary

<https://www.bestchange.ru/exchangers/mkt=en&id=234>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>0</b>	<b>Partial match</b>	<b>12 years</b>	<b>High</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

## Manipulation Tricks

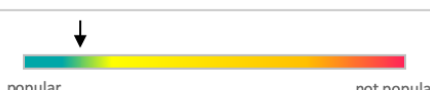
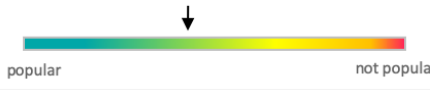
Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

### No Tricks Used

Of the known tricks attackers use, none appear in this URL

## URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	bestchange.ru
<b>Categorization</b> Indicates the type of this website, such as shopping or travel. New or phishing domains may not be categorized.	Finance and Banking
<b>Domain Age</b> The date when the domain was first registered.	2007-06-12 <b>12 years</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	 popular not popular
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	 popular not popular
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>partially match</b> your URL: <a href="https://bestchange.com/">https://bestchange.com/</a>

Best Change - Safe

## Report Summary

<https://l.facebook.com/l.php?u=http%3A%2F%2F67.238.165>

Goes to:

[https://67.238.238.165/&h=AT3y8f6\\_tofRPamyd-PMIIZT9D64tAzKlzhBZ0zDRZVTibJU...](https://67.238.238.165/&h=AT3y8f6_tofRPamyd-PMIIZT9D64tAzKlzhBZ0zDRZVTibJU...)

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b> 1	<b>Search Result</b> No Match	<b>Domain Age</b> No data	<b>Domain Popularity</b> Low
--------------------------------------	----------------------------------	------------------------------	---------------------------------

Color Code: ✖ Known Issue ! Possible Issue ✔ No Issue


## Manipulation Tricks

Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

<b>IP address</b> The organization name is represented by its IP address "67.238.238.165" not the human readable representation.	<b>Equivalent to:</b> <a href="https://negociosparana.com.br">https://negociosparana.com.br</a>
---	--

## URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	negociosparana.com.br
<b>Location</b> The physical address where the domain owner claims they live or conduct business.	Brazil
<b>Categorization</b> Indicates the type of this website, such as shopping or travel. New or phishing domains may not be categorized.	Business
<b>Domain Age</b> The date when the domain was first registered.	No data
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>does not match</b> your URL: <a href="https://www.mathworks.com/">https://www.mathworks.com/...</a>

Facebook - Phishing

## Report Summary

<https://email.microsoftonline.com/login/?mkt=en-GB>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>1</b>	<b>No Match</b>	<b>17 years</b>	<b>High</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue



## Manipulation Tricks

Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

<b>Similar to a popular domain:</b> '365onmicrosoft.com' is similar to the popular domain "microsoft.com".	<a href="https://click.email.microsoftonline.com/...">https://click.email.microsoftonline.com/...</a>
---	---

## URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	microsoftonline.com
<b>Location</b> The physical address where the domain owner claims they live or conduct business..	United States
<b>Categorization</b> Indicates the type of this website, such as shopping or travel. New or phishing domains may not be categorized.	Information Technology
<b>Domain Age</b> The date when the domain was first registered.	09-07-2002 <b>17 years</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	 popular not popular
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	 popular not popular
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>does not match</b> your URL: <a href="https://answers.microsoft.com/...">https://answers.microsoft.com/...</a>

Microsoft - Safe



## Report Summary

[https://www.tripod.lycos.com/pricing/?plan=free\\_ad](https://www.tripod.lycos.com/pricing/?plan=free_ad)

**⚠ We cannot guarantee the safety of danger of this link.**

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>0</b>	<b>Partial match</b>	<b>24 years</b>	<b>High</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

## Manipulation Tricks


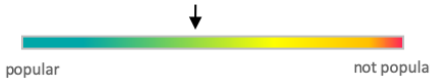
Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

### No Tricks Used

Of the known tricks attackers use, none appear in this URL

## URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	lycos.com
<b>Location</b> Phishing website owners are likely to be registered in countries different from the legitimate website.	GDPR masked
<b>Categorization</b> Indicates the type of this website, such as shopping or travel. New or phishing domains may not be categorized.	Search Engines and Portals
<b>Domain Age</b> When the domain was first registered.	1995-04-13 <b>24 years</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	 popular not popular
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	 popular not popular
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>partially match</b> your URL: <a href="https://tripod.lycos.co.uk/...">https://tripod.lycos.co.uk/...</a>

Tripod - Safe

## Report Summary

<https://international.bittrex.com/account/?id=2423>

**⚠ We cannot guarantee the safety of danger of this link.**

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>1</b>	<b>No Match</b>	<b>5 months</b>	<b>Low</b>

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue



## Manipulation Tricks

Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

<b>Non-English Characters</b> This URL contains <b>Vietnamese</b> characters. If you do not expect a Vietnamese website, do not risk it.	<a href="https://international.bittrex.com/...">https://international.bittrex.com/...</a>
---	---

## URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	bittrex.com
<b>Domain Age</b> The date when the domain was first registered.	2019-03-28 <b>5 months</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	 popular <span style="float: right;">not popular ↓</span>
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	 popular <span style="float: right;">not popular ↓</span>
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>does not match</b> your URL: <a href="https://www.americancentury.com/...">https://www.americancentury.com/...</a>

Bittrex - Phishing

## Report Summary

<https://itmurl.com/www.ebay.co.uk/item=30327559652>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
0	No Match	1 month	Low

Color Code: ✘ Known Issue ! Possible Issue ✔ No Issue

## Manipulation Tricks



Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

### No Tricks Used

Of the known tricks attackers use, none appear in this URL

## URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	itmurl.com
<b>Location</b> The physical address where the domain owner claims they live or conduct business.	United Kingdom
<b>Categorization</b> Indicates the type of this website, such as shopping or travel. New or phishing domains may not be categorized.	Newly Registered Domain
<b>Domain Age</b> The date when the domain was first registered.	19-08-2019 <b>1 month</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	 popular → not popular ↓
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	 popular → not popular ↓
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>does not match</b> your URL: <a href="https://stackoverflow.com/signin/...">https://stackoverflow.com/signin/...</a>

Ebay - Safe

## Full-report Understanding

In this section, we will show you parts of a report and we will ask you to chose the interpretation that makes most sense for you.

37. What does "Manipulation tricks" mean?

### Manipulation Tricks

Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.


<b>Non-English Characters</b> This URL contains <b>Vietnamese</b> characters. If you do not expect a Vietnamese website, do not risk it.	<a href="https://international.bittrex.com/">https://international.bittrex.com/...</a>
---	--

- The organisation uses these manipulations to highlight how good they are at security.
- An attacker has crafted the URL using these tricks to confuse people.
- The Internet community uses these tricks to confuse attackers.
- The description is confusing.
- Other

38. What does "PageRank" mean?

### PageRank

Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.



popular not popular ↓


- The similarity between this page and the official expected website.
- How often this website has been visited by the general population.
- How many popular pages, such as Facebook, are connected to this page.
- The description is confusing.
- Other

39. What does "Categorization" mean?

<b>Categorization</b> Indicates the type of this website, such as shopping or travel. New or phishing domains may not be categorized.	Search Engines and Portals
--	----------------------------

- The type of the top level domain of the URL such as government (gov), education (edu), or commercial (com).
- The type of the organization, like financial or social media.
- If the URL is a special type, like having Chinese letters, is a shortener, or is an auto-redirector.
- The description is confusing.
- Other


40. What does "Domain popularity" mean?

<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	 popular <span style="float: right;">not popular</span>
--	--

- It is based on the number of people who visit this website compared to other websites.
- It is based on the number of other popular websites that are linked to this website.
- It is based on the quality of the site compared to similar sites.
- The description is confusing.
- Other

41. What does the warning below mean?

Qualtrics Survey Software

 **Note:** tripod.com provides web space for other people to put their webpages online. So the age and popularity of the site cannot be accurately determined.

**Usually only small companies and personal websites are hosted by other domains.**

- This domain is known to host malicious websites.
- The domain is a discussion forum where people can post links to other websites.
- This domain belongs to an organisation that allows other people to create their own websites.
- The description is confusing.
- Other

42. What does the following mean?

<https://fb.me/messages/t/788720331154519>

**Goes to:**

<https://www.facebook.com/>

- When I click on the first URL (fb.me), the second URL (facebook) will open.
- The first URL sends personal data to the second URL.
- The report is recommending that I visit the second URL instead because it is safer.
- The description is confusing.
- Other

## 43. What does "Domain age" mean?

<b>Domain Age</b> When the domain was first registered.	1995-04-13 <b>24 years</b>
--	-------------------------------

- It indicates how long ago the physical organisation was established.
- It indicates how long ago this URL was reported as phishing.
- It indicates how long the organisation has had an online presence at this domain.
- The description is confusing.
- Other

## 44. What does "Search result" mean?

<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>does not match</b> your URL: <a href="https://promaster.com/Category...">https://promaster.com/Category...</a>
---	--

- This exact URL does not appear in any list of safe URLs.
- This exact URL does not appear in any list of malicious URLs.
- This exact URL does not appear in Google's top search results.
- The description is confusing.
- Other

## 45. What does "Location" mean?

<b>Location</b> Phishing website owners are likely to be registered in countries different from the legitimate website.	United states
--	---------------

- The country the domain servers (computers) are located in.
- The country I live in.
- The country the domain owners are located in.
- The description is confusing.
- Other

46. What does "Domain" mean?

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	tripod.com
---	------------

- The main part of the website address.
- The category in which this URL belongs.
- The legal domain which this website falls in, such as Japanese law.
- The description is confusing.
- Other



## Control Judgement Group

51. Imagine that you want to visit \${Company name}, which is \${Company description}. Do you think the URL: "\${Company URL with the domain highlighted}" leads to a page owned by the above company or is it a malicious URL?

- Malicious
- Trustworthy

52. How confident are you in your decision above?

- Not confident at all
- Less confident
- Somewhat confident
- Confident
- Very confident

53. Indicate the **three** features that most **influenced** your decision about if the URL is malicious or trustworthy. Note that some features may not appear on some interfaces.

- "\${domain}" part of the URL
- My own prior experience reading URLs
- "https://" part of the URL
- "\${subdomain}" part of the URL
- I know \${company name}'s URL
- Other:

54. Any comments (optional):

## Control Understanding

In this section, we will show you a report and sections from it. Please read the following report carefully and answer the questions below about which features of the report make the most sense for you.

NOTE: The information provided in this report is fictional. We use the example to see how readable the report design is.

### Report Summary

<https://readingability.com/1vcs37>

**Goes to:**

<https://www.letmebeyoureyes.com/users/userID-12975/en/>

⚠ We cannot guarantee the safety of danger of this link.

<b>Used Manipulation tricks</b>	<b>Search Result</b>	<b>Domain Age</b>	<b>Domain Popularity</b>
<b>0</b>	<b>No Match</b>	<b>28 years</b>	<b>Moderate</b>

Color Code: ✗ Known Issue ⚠ Possible Issue ✓ No Issue

### Manipulation Tricks



Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.

**No Tricks Used**

Of the known tricks attackers use, none appear in this URL.

### URL Facts

Facts about the URL to help you compare between what you know with what this URL have.

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	letmebeyoureyes.com
<b>Location</b> The physical address where the domain owner claims they live or conduct business.	Wonderland
<b>Categorization</b> Indicates the type of this website, such as shopping or travel. New or phishing domains may not be categorized.	Fiction
<b>Domain Age</b> The date when the domain was first registered.	22-08-1990 <b>28 years</b>
<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	
<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	
<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>does not match</b> your URL: <a href="https://www.youtube.com/account/...">https://www.youtube.com/account/...</a>


[See the report in a new tab](#)

55. What does "Manipulation tricks" mean?

Manipulation Tricks	
Manipulation tricks are used to hide where a URL really goes. Below are the tricks that appear in this URL.	
<b>Non-English Characters</b> This URL contains <b>Vietnamese</b> characters. If you do not expect a Vietnamese website, do not risk it.	<a href="https://international.bittrex.com/">https://international.bittrex.com/...</a>

- The organisation uses these manipulations to highlight how good they are at security.
- The Internet community uses these tricks to confuse attackers.
- An attacker has crafted the URL using these tricks to confuse people.
- The description is confusing.
- Other

56. What does "PageRank" mean?

<b>PageRank</b> Indicates how often popular web pages link to this page. Different parts of a domain can have different page ranks.	 popular <span style="float: right;">not popular</span>
--	---


- The similarity between this page and the official expected website.
- How often this website has been visited by the general population.
- How many popular pages, such as Facebook, are connected to this page.
- The description is confusing.
- Other

57. What does "Categorization" mean?

<b>Categorization</b> Indicates the type of this website, such as shopping or travel. New or phishing domains may not be categorized.	Search Engines and Portals
--	----------------------------


- The type of the top level domain of the URL such as government (gov), education (edu), or commercial (com).
- The type of the organization, like financial or social media.
- If the URL is a special type, like having Chinese letters, is a shortener, or is an auto-redirector.
- The description is confusing.
- Other

58. What does "Domain popularity" mean?

<b>Domain Popularity</b> Global rank that indicates how often all pages associated with a domain are visited relative to other domains.	
--	--

- It is based on the quality of the site compared to similar sites.
- It is based on the number of other popular websites that are linked to this website.
- It is based on the number of people who visit this website compared to other websites.
- The description is confusing.
- Other

59. What does the warning below mean?

 **Note:** tripod.com provides web space for other people to put their webpages online. So the age and popularity of the site cannot be accurately determined.

**Usually only small companies and personal websites are hosted by other domains.**

- This domain belongs to an organisation that allows other people to create their own websites.
- The domain is a discussion forum where people can post links to other websites.
- This domain is known to host malicious websites.
- The description is confusing.
- Other

60. What does the following mean?

<a href="https://fb.me/messages/t/788720331154519">https://fb.me/messages/t/788720331154519</a>
<b>Goes to:</b>
<a href="https://www.facebook.com/">https://www.facebook.com/</a>

- The report is recommending that I visit the second URL instead because it is safer.
- When I click on the first URL (fb.me), the second URL (facebook) will open.
- The first URL sends personal data to the second URL.
- The description is confusing.
- Other

61. What does "Domain age" mean?

<b>Domain Age</b> When the domain was first registered.	1995-04-13 <b>24 years</b>
--	-------------------------------

- It indicates how long the organisation has had an online presence at this domain.
- It indicates how long ago this URL was reported as phishing.
- It indicates how long ago the physical organisation was established.
- The description is confusing.
- Other

62. What does "Search result" mean?

<b>Top Search Result</b> Top result when we googled the URL you gave us. Legitimate URLs should appear on the top search results.	The search result <b>does not match</b> your URL: <a href="https://promaster.com/Category...">https://promaster.com/Category...</a>
---	--

- This exact URL does not appear in any list of safe URLs.
- This exact URL does not appear in Google's top search results.
- This exact URL does not appear in any list of malicious URLs.
- The description is confusing.
- Other

63. What does "Location" mean?

<b>Location</b> Phishing website owners are likely to be registered in countries different from the legitimate website.	United states
--	---------------

- The country the domain servers (computers) are located in.
- The country the domain owners are located in.
- The country I live in.
- The description is confusing.
- Other

64. What does "Domain" mean?

<b>Domain</b> Primary web address of the group that maintains this website. It should match the organization you expect.	tripod.com
---	------------

- The main part of the website address.
- The category in which this URL belongs.
- The legal domain which this website falls in, such as Japanese law.
- The description is confusing.
- Other









## Demographics

74. Prolific ID:

Qualtrics Survey Software

75. Age (in years):

76. Sex:

- Male
- Female
- Non-binary
- Prefer not to say
- Prefer to self describe

77. What is the highest degree you have obtained?

- High school degree or equivalent
- Some college, no degree
- Bachelor's degree
- Master's degree
- Professional degree (vocational training)
- Doctorate degree
- Other (please specify)



# Bibliography

- [1] Verizon. 2020 data breach investigations report. Technical report, Verizon Trademark Services LLC, 2020. [Online] <https://vz.to/3vKNI1K>. Accessed Jun. 2020.
- [2] FBI's Internet Crime Complaint Center (IC3). 2017 internet crime report. Technical report, The Federal Bureau of Investigation (FBI), Internet Crime Complaint Center, 2017. [Online] <https://pdf.ic3.gov/2017%5FIC3Report.pdf>. Accessed Aug. 2020.
- [3] FBI. 2019 internet crime report, data reflects an evolving threat and the importance of reporting. Technical report, The Federal Bureau of Investigation, Internet Crime Complaint Center, 2020. [Online] <https://www.fbi.gov/news/stories/2019-internet-crime-report-released-021120>. Accessed Aug. 2020.
- [4] Phishing attacks: defending your organisation. <https://www.ncsc.gov.uk/guidance/phishing>, Feb. 2018. Accessed Feb. 2019.
- [5] Constantinos Patsakis and Anargyros Chrysanthou. Analysing the fall 2020 Emotet campaign. *CoRR*, abs/2011.06479, 2020.
- [6] Emotet malware disrupted, Feb 2021.
- [7] Gilchan Park, Lauren M. Stuart, Julia M. Taylor, and Victor Raskin. Comparing machine and human ability to detect phishing emails. In *2014 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2014, October 5-8, 2014*, pages 2322–2327, San Diego, CA, USA, 2014. IEEE.
- [8] Sara Albakry, Kami Vaniea, and Maria K. Wolters. What is this URL's destination? empirical evaluation of users' URL reading. In *CHI '20: CHI Conference*

- on Human Factors in Computing Systems*, pages 1–12, Honolulu, HI, USA, Apr. 2020. ACM.
- [9] Kholoud Althobaiti, Nicole Meng, and Kami Vaniea. I don't need an expert! making URL phishing features human comprehensible. In *CHI '21: CHI Conference on Human Factors in Computing Systems*, pages 1–17, Yokohama, Japan, May. 2021. ACM.
- [10] Hongpeng Zhu. Online meta-learning firewall to prevent phishing attacks. *Neural Computing and Applications*, 32(23):17137–17147, 2020.
- [11] Tiliang Zhang, Hua Zhang, and Fei Gao. A malicious advertising detection scheme based on the depth of URL strategy. In *Proceedings of the 2013 Sixth International Symposium on Computational Intelligence and Design - Volume 02*, ISCID '13, page 57–60, USA, 2013. IEEE Computer Society.
- [12] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason I. Hong, and Elizabeth Nunge. Anti-phishing Phil: the design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 3rd Symposium on Usable Privacy and Security, SOUPS*, volume 229, pages 88–99, Pittsburgh, Pennsylvania, USA, Jul. 2007. ACM.
- [13] Rakesh M. Verma and Ayman El Aassal. Comprehensive method for detecting phishing emails using correlation-based analysis and user participation. In *Proceedings of the Seventh ACM Conference on Data and Application Security and Privacy, CODASPY 2017, Scottsdale, AZ, USA, March 22-24, 2017*, pages 155–157. ACM, 2017.
- [14] Devdatta Akhawe and Adrienne Porter Felt. Alice in Warningland: A large-scale field study of browser security warning effectiveness. In *22nd USENIX Security Symposium (USENIX Security 13)*, pages 257–272, Washington, D.C., Aug. 2013. USENIX Association.
- [15] Anjum N. Shaikh, Antesar M. Shabut, and M. Alamgir Hossain. A literature review on phishing crime, prevention review and investigation of gaps. In *10th International Conference on Software, Knowledge, Information Management & Applications, SKIMA 2016*, pages 9–15, Chengdu, China, 2016. IEEE.

- [16] Katelin A. Moul. Avoid phishing traps. In *ACM SIGUCCS Annual Conference, SIGUCCS*, pages 199–208, New Orleans, LA, USA, 2019. ACM.
- [17] Wayne D. Kearney and Hennie A. Kruger. Phishing and organisational learning. In *Security and Privacy Protection in Information Processing Systems - 28th IFIP*, volume 405, pages 379–390, Auckland, New Zealand, Jul. 2013. Springer.
- [18] E.E.H. Lastdrager, Pieter H. Hartel, and Marianne Junger. Apatate: Anti-phishing analysing and triaging environment (Poster). In *36th IEEE Symposium on Security and Privacy*, page 2, United States, May. 2015. IEEE Computer Society.
- [19] Pavlo Burda, Luca Allodi, and Nicol Zannone. Don't forget the human: a crowdsourced approach to automate response and containment against spear phishing attacks. In *European Symposium on Security and Privacy*, page 6, Virtual Conference, 2020. IEEE.
- [20] Rodrigo Werlinger, Kasia Muldner, Kirstie Hawkey, and Konstantin Beznosov. Preparation, detection, and analysis: the diagnostic work of IT security incident response. *Information Management & Computer Security*, 18(1):26–42, 2010.
- [21] Erka Koivunen. “why wasn't I notified?”: Information security incident reporting demystified. In *Information Security Technology for Applications - 15th Nordic Conference on Secure IT Systems, NordSec, Revised Selected Papers*, volume 7127 of *Lecture Notes in Computer Science*, pages 55–70, Espoo, Finland, Oct. 2010. Springer.
- [22] Martin Husák and Jakub Cegan. Phigaro: Automatic phishing detection and incident response framework. In *Ninth International Conference on Availability, Reliability and Security, ARES*, pages 295–302, Fribourg, Switzerland, Sep. 2014. IEEE Computer Society.
- [23] Heather Richter Lipford and Mary Ellen Zurko. Someone to watch over me. In *Proceedings of the 2012 New Security Paradigms Workshop, NSPW '12*, page 67–76, New York, NY, USA, 2012. Association for Computing Machinery.
- [24] James Nicholson, Lynne M. Coventry, and Pam Briggs. Introducing the cybersurvival task: Assessing and addressing staff beliefs about effective cyber protection. In *Fourteenth Symposium on Usable Privacy and Security, SOUPS*,

- August 12-14, 2018*, pages 443–457, Baltimore, MD, USA, 2018. USENIX Association.
- [25] Cofense PhishMe. Enterprise phishing resiliency and defense report. Technical report, PhishMe, Inc, 2017. [Online] <https://cofense.com/wp-content/uploads/2017/11/Enterprise-Phishing-Resiliency-and-Defense-Report-2017.pdf>. Accessed Aug. 2020.
- [26] Matheesha Fernando and Nalin Asanka Gamagedara Arachchilage. Why Johnny can't rely on anti-phishing educational interventions to protect himself against contemporary phishing attacks? *CoRR*, abs/2004.13262:1–12, 2020.
- [27] Doyen Sahoo, Chenghao Liu, and Steven C. H. Hoi. Malicious URL detection using machine learning: A survey. <http://arxiv.org/abs/1701.07179>, 2019.
- [28] LOGOPHILIA LIMITED. Phishing. <https://wordspy.com/index.php?word=phishing>, Aug. 2003. Accessed Nov. 2020.
- [29] Marc Rader and Shawon Rahman. Exploring historical and emerging phishing techniques and mitigating the associated security risks. *CoRR*, abs/1512.00082, 2015.
- [30] Daniel Jampen, Gürkan Gür, Thomas Sutter, and Bernhard Tellenbach. Don't click: towards an effective anti-phishing training. A comparative literature review. *Human-centric Computing and Information Sciences*, 10:33, 2020.
- [31] WHO. Beware of criminals pretending to be WHO. <https://www.who.int/about/communications/cyber-security>, 2020. Accessed Nov. 2020.
- [32] Danny Palmer. Phishing alert: GDPR-themed scam wants you to hand over passwords, credit card details. <https://zd.net/3u2pFKV>, May. 2018. Accessed Nov. 2020.
- [33] Elmer EH Lastdrager. Achieving a consensual definition of phishing based on a systematic review of the literature. *Crime Science*, 3(1):9, 2014.
- [34] Ammar Almomani, Brij B. Gupta, Samer Atawneh, Andrew Meulenberg, and Eman Almomani. A survey of phishing email filtering techniques. *IEEE Communications Surveys Tutorials*, 15(4):2070–2090, 2013.

- [35] Adam Oest, Yeganeh Safaei, Adam Doupé, Gail-Joon Ahn, Brad Wardman, and Gary Warner. Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis. In *2018 APWG Symposium on Electronic Crime Research, eCrime 2018, May 15-17, 2018*, pages 1–12, San Diego, CA, USA, 2018. IEEE.
- [36] Brij B. Gupta, Aakanksha Tewari, Ankit Kumar Jain, and Dharma P. Agrawal. Fighting against phishing attacks: state of the art and future challenges. *Neural Computing and Applications*, 28(12):3629–3654, 2017.
- [37] Photon Research Team. The ecosystem of phishing: From Minnows to Marlins. <https://www.digitalshadows.com/blog-and-research/the-ecosystem-of-phishing/>, Feb. 2020. Accessed Nov. 2020.
- [38] Ahmed AlEroud and Lina Zhou. Phishing environments, techniques, and countermeasures: A survey. *Computers and Security*, 68:160–196, 2017.
- [39] Rana Alabdan. Phishing attacks survey: Types, vectors, and technical approaches. *Future Internet*, 12(10):168, 2020.
- [40] Swapan Purkait. Phishing counter measures and their effectiveness - literature review. *Information Management & Computer Security*, 20(5):382–420, 2012.
- [41] Sidharth Chhabra, Anupama Aggarwal, Fabrício Benevenuto, and Ponnurangam Kumaraguru. Phi.sh/\$ocial: the phishing landscape through short URLs. In *The 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference, CEAS*, pages 92–101, Perth, Australia, Sep. 2011. ACM.
- [42] Kayode Sakariyah Adewole, Nor Badrul Anuar, Amirrudin Kamsin, Kasturi Dewi Varathan, and Syed Abdul Razak. Malicious accounts: Dark of the social networks. *Journal of Network and Computer Applications*, 79:41–67, 2017.
- [43] Francois Mouton, Mercia M. Malan, Louise Leenen, and H.S. Venter. Social engineering attack framework. In *2014 Information Security for South Africa*, pages 1–9, 2014.
- [44] Daniela A. S. de Oliveira, Harold Rocha, Huizi Yang, Donovan Ellis, Sandeep Dommaraju, Melis Muradoglu, Devon Weir, Adam Soliman, Tian Lin, and Natalie C. Ebner. Dissecting spear phishing emails for older vs young adults: On



- the interplay of weapons of influence and life domains in predicting susceptibility to phishing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, May 06-11, 2017*, pages 6412–6424. ACM, 2017.
- [45] Kathryn Parsons, Agata McCormac, Malcolm Robert Pattinson, Marcus A. Butavicius, and Cate Jerram. The design of phishing studies: Challenges for researchers. *Computers and Security*, 52:194–206, 2015.
- [46] Marcus A. Butavicius, Kathryn Parsons, Malcolm R. Pattinson, and Agata McCormac. Breaching the human firewall: Social engineering in phishing and spear-phishing emails. *CoRR*, abs/1606.00887, 2016.
- [47] Zinaida Benenson, Freya Gassmann, and Robert Landwirth. Unpacking spear phishing susceptibility. In *Financial Cryptography and Data Security - FC International Workshops, Malta Revised Selected Papers*, volume 10323 of *Lecture Notes in Computer Science*, pages 610–627. Springer, Apr. 2017.
- [48] Rami Mustafa A. Mohammad, Fadi A. Thabtah, and Lee McCluskey. Tutorial and critical analysis of phishing websites methods. *Computer Science Review*, 17:1–24, 2015.
- [49] Markus Jakobsson and Sid Stamm. Invasive browser sniffing and countermeasures. In *Proceedings of the 15th International Conference on World Wide Web, WWW '06*, page 523–532, New York, NY, USA, 2006. Association for Computing Machinery.
- [50] Xiao Han, Nizar Kheir, and Davide Balzarotti. Phisheye: Live monitoring of sandboxed phishing kits. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*, pages 1402–1413. ACM, 2016.
- [51] Kjell Hausken and Gregory Levitin. Review of systems defense and attack models. *International Journal of Performability Engineering*, 8(4):355–366, 2012.
- [52] Grant Ho, Asaf Cidon, Lior Gavish, Marco Schweighauser, Vern Paxson, Stefan Savage, Geoffrey M. Voelker, and David A. Wagner. Detecting and characterizing lateral phishing at scale. In *28th USENIX Security Symposium*, pages 1273–1290, Santa Clara, CA, USA, Aug. 2019. USENIX Association.

- [53] Lorenzo Franceschi-Bicchierai. How hackers broke into John Podesta and Colin Powell's Gmail accounts. <https://bit.ly/20773Kw>, 2016. Accessed Aug. 2020.
- [54] Grant Ho, Aashish Sharma, Mobin Javed, Vern Paxson, and David A. Wagner. Detecting credential spearphishing in enterprise settings. In *26th USENIX Security Symposium, USENIX Security*, pages 469–485, Vancouver, BC, Canada, Aug. 2017. USENIX Association.
- [55] Asaf Cidon, Lior Gavish, Itay Bleier, Nadia Korshun, Marco Schweighauser, and Alexey Tsitkin. High precision detection of business email compromise. In *28th USENIX Security Symposium, USENIX Security*, pages 1291–1307, Santa Clara, CA, USA, Aug. 2019. USENIX Association.
- [56] Mahmoud Khonji, Youssef Iraqi, and Andrew Jones. Mitigation of spear phishing attacks: A content-based authorship identification framework. In *6th International Conference for Internet Technology and Secured Transactions, ICITST*, pages 416–421, Abu Dhabi, UAE, Dec. 2011. IEEE.
- [57] Gianluca Stringhini and Olivier Thonnard. That ain't you: Blocking spearphishing through behavioral modelling. In *Detection of Intrusions and Malware, and Vulnerability Assessment - 12th International Conference, DIMVA, Proceedings*, volume 9148 of *Lecture Notes in Computer Science*, pages 78–97, Milan, Italy, Jul. 2015. Springer.
- [58] Sevtap Duman, Kubra Kalkan-Cakmakci, Manuel Egele, William K. Robertson, and Engin Kirda. Emailprofiler: Spearphishing filtering with header and stylometric features of emails. In *40th Annual Computer Software and Applications Conference, COMPSAC*, pages 408–416, Atlanta, GA, USA, Jun. 2016. IEEE Computer Society.
- [59] Mohammed Al-Janabi, Ed de Quincey, and Peter Andras. Using supervised machine learning algorithms to detect suspicious URLs in online social networks. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, July 31 - August 03, 2017*, pages 1104–1111, Sydney, Australia, 2017. ACM.

- [60] Issa Qabajeh, Fadi A. Thabtah, and Francisco Chiclana. A recent review of conventional vs. automated cybersecurity anti-phishing techniques. *Computer Science Review*, 29:44–55, 2018.
- [61] Rick Wash. How experts detect phishing scam emails. *Proceedings of the ACM on Human Computer Interaction*, 4(CSCW2):160:1–160:28, 2020.
- [62] H. Atashzar, A. Torkaman, M. Bahrololum, and M. H. Tadayon. A survey on web application vulnerabilities and countermeasures. In *6th International Conference on Computer Sciences and Convergence Information Technology (IC-CIT)*, pages 647–652, Nov. 2011.
- [63] Kieran Rendall, Antonia Nisioti, and Alexios Mylonas. Towards a multi-layered phishing detection. *Sensors*, 20(16):4540, 2020.
- [64] Jason Hong. The state of phishing attacks. *Communications of the ACM*, 55(1):74–81, 2012.
- [65] Lorrie Faith Cranor. A framework for reasoning about the human in the loop. In *Usability, Psychology, and Security, UPSEC'08*, pages 1–15, San Francisco, CA, sUSA, Apr. 2008. USENIX Association.
- [66] Verizon. 2017 data breach investigations report. Technical report, Verizon Trademark Services LLC, 2017. [Online] <https://vz.to/3h8SYbE>. Accessed Jun. 2018.
- [67] Robert W. Reeder, Adrienne Porter Felt, Sunny Consolvo, Nathan Malkin, Christopher Thompson, and Serge Egelman. An experience sampling study of user reactions to browser warnings in the field. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI*, page 512, Montreal, QC, Canada, Apr. 2018. ACM.
- [68] Taimur Bakhshi, Maria Papadaki, and Steven Furnell. Social engineering: assessing vulnerabilities in practice. *Information Management & Computer Security*, 17(1):53–63, 2009.
- [69] Rachna Dhamija, J. D. Tygar, and Marti A. Hearst. Why phishing works. In *Proceedings of the Conference on Human Factors in Computing Systems, CHI*, pages 581–590, Montréal, Québec, Canada, Apr. 2006. ACM.

- [70] Melanie Volkamer, Karen Renaud, Benjamin Reinheimer, and Alexandra Kunz. User experiences of TORPEDO: tooltip-powered phishing email detection. *Computer Security*, 71:100–113, 2017.
- [71] Jim Blythe, L. Jean Camp, and Vaibhav Garg. Targeted risk communication for computer security. In *Proceedings of the 16th International Conference on Intelligent User Interfaces, IUI*, pages 295–298, Palo Alto, CA, USA, Feb. 2011. ACM.
- [72] Rennie Naidoo. Analysing urgency and trust cues exploited in phishing scam designs. In *10th International Conference on Cyber Warfare and Security, IC-CWS*, pages 216–222, The University of Venda and The Council for Scientific and Industrial Research, South Africa, 2015. Academic Conferences International Limited.
- [73] Ahmed Abbasi, Fatemeh Zahedi, and Yan Chen. Impact of anti-phishing tool performance on attack success rates. In *2012 IEEE International Conference on Intelligence and Security Informatics, ISI 2012, Washington, DC, USA, June 11-14, 2012*, pages 12–17. IEEE, 2012.
- [74] Markus Jakobsson, Alex Tsow, Ankur Shah, Eli Blevis, and Youn-Kyung Lim. What instills trust? A qualitative study of phishing. In *Financial Cryptography and Data Security, 11th International Conference, FC, and 1st International Workshop on Usable Security, USEC, Scarborough, Trinidad and Tobago, February 12-16. Revised Selected Papers*, volume 4886 of *Lecture Notes in Computer Science*, pages 356–361. Springer, 2007.
- [75] Mohamed Alsharnouby, Furkan Alaca, and Sonia Chiasson. Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human-Computer Studies*, 82:69–82, 2015.
- [76] Hazim Almuhiemedi, Adrienne Porter Felt, Robert W. Reeder, and Sunny Consolvo. Your reputation precedes you: History, reputation, and the chrome malware warning. In *Tenth Symposium on Usable Privacy and Security, SOUPS*, pages 113–128, Menlo Park, CA, USA, Jul. 2014. USENIX Association.
- [77] Julie S. Downs, Mandy B. Holbrook, and Lorrie Faith Cranor. Decision strategies and susceptibility to phishing. In *Proceedings of the 2nd Symposium on*

- Usable Privacy and Security, SOUPS*, volume 149, pages 79–90, Pittsburgh, Pennsylvania, USA, July 12-14 2006. ACM.
- [78] Gamze Canova, Melanie Volkamer, Clemens Bergmann, and Benjamin Reinheimer. NoPhish app evaluation: Lab and retention study. In *Internet Society, 8 February 2015*, volume 453 of *USEC '15*, pages 1–10, San Diego, CA, USA, 2015. The Internet Society.
- [79] Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Theodore Pham. School of phish: A real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security, SOUPS '09*, pages 1–12, New York, NY, USA, 2009. ACM.
- [80] Tara Seals. Cost of user security training tops \$290k per year. <https://www.infosecurity-magazine.com/news/cost-of-user-security-training>, 2017. Accessed Nov. 2020.
- [81] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason I. Hong. Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology*, 10(2):7:1–7:31, 2010.
- [82] Hermann Ebbinghaus. Memory: a contribution to experimental psychology. *Annals of neurosciences*, 20(4):155–156, Oct. 2013.
- [83] Kholoud Althobaiti, Ghaidaa Rummani, and Kami Vaniea. A review of human- and computer-facing URL phishing features. In *European Symposium on Security and Privacy Workshops, EuroS&P Workshops*, pages 182–191, Stockholm, Sweden, Jun. 2019. IEEE.
- [84] Hossein Siadati, Sean Palka, Avi Siegel, and Damon McCoy. Measuring the effectiveness of embedded phishing exercises. In *10th USENIX Workshop on Cyber Security Experimentation and Test, CSET 2017, August 14, 2017*, page 8, Vancouver, BC, Canada, 2017. USENIX Association.
- [85] Mahmoud Khonji, Youssef Iraqi, and Andrew Jones. Phishing detection: A literature survey. *IEEE Communications Surveys Tutorials*, 15(4):2091–2121, 2013.

- [86] Masayuki Higashino. A design of an anti-phishing training system collaborated with multiple organizations. In *Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services, iiWAS 2019, December 2-4, 2019*, pages 589–592, Munich, Germany, 2019. ACM.
- [87] Collin Jackson, Daniel R. Simon, Desney S. Tan, and Adam Barth. An evaluation of extended validation and picture-in-picture phishing attacks. In *Financial Cryptography and Data Security, 11th International Conference, FC, and 1st International Workshop on Usable Security, USEC, February 12-16, 2007. Revised Selected Papers*, volume 4886 of *Lecture Notes in Computer Science*, pages 281–293, Scarborough, Trinidad and Tobago, 2007. Springer.
- [88] Joshua Reynolds, Deepak Kumar., Zane Ma, Rohan Subramanian, Meishan Wu, Martin Shelton, Joshua Mason, Emily Stark, and Michael Bailey. Measuring identity confusion with uniform resource locators. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI '20*, pages 1–12, Honolulu, HI, USA, 2020. ACM.
- [89] Netcraft Ltd. Internet security and data mining. <https://www.netcraft.com/>, 2019. Accessed Jun. 2020.
- [90] Min Wu, Robert C. Miller, and Simson L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *Proceedings of the Conference on Human Factors in Computing Systems, CHI*, pages 601–610, Montréal, Québec, Canada, Apr. 2006. ACM.
- [91] Jun Yang, Pengpeng Yang, Xiaohui Jin, and Qian Ma. Multi-classification for malicious URL based on improved semi-supervised algorithm. In *IEEE International Conference on Computational Science and Engineering, CSE 2017, and IEEE International Conference on Embedded and Ubiquitous Computing, EUC, Volume 1*, pages 143–150, Guangzhou, China, Jul. 2017. IEEE Computer Society.
- [92] Kholoud Althobaiti, Kami Vaniea, and Serena Zheng. Faheem: Explaining URLs to people using a Slack bot. In *2018 Symposium on Digital Behaviour Intervention for Cyber Security (AISB)*, pages 1–8, Liverpool, UK, Apr. 2018. University of Liverpool.

- [93] Angela Sasse. Scaring and bullying people into security won't work. *IEEE Security & Privacy*, 13(3):80–83, 2015.
- [94] Jason Steer. Defending against spear-phishing. *Computer Fraud & Security*, 2017(8):18–20, 2017.
- [95] Elie Bursztein, Borbala Benko, Daniel Margolis, Tadek Pietraszek, Andy Archer, Allan Aquino, Andreas Pitsillidis, and Stefan Savage. Handcrafted fraud and extortion: Manual account hijacking in the wild. In *Proceedings of the 2014 Internet Measurement Conference, IMC 2014, Vancouver, BC, Canada, November 5-7, 2014*, pages 347–358. ACM, 2014.
- [96] Jeremiah Onaolapo, Enrico Mariconti, and Gianluca Stringhini. What happens after you are pwned: Understanding the use of leaked webmail credentials in the wild. In *Proceedings of the 2016 ACM on Internet Measurement Conference, IMC 2016, Santa Monica, CA, USA, November 14-16, 2016*, pages 65–79. ACM, 2016.
- [97] Katiana Krawchenko. The phishing email that hacked the account of John Podesta. <https://www.cbsnews.com/news/the-phishing-email-that-hacked-the-account-of-john-podesta/>, Oct. 2016. Accessed Aug. 2020.
- [98] ProofPoint Phishing Report. State of the phish. Technical Report 1, Proofpoint, Inc., 2019. [Online] <https://bit.ly/201n180>. Accessed May. 2019.
- [99] Edwin Donald Frauenstein and Rossouw von Solms. Phishing: How an organization can protect itself. In *Information Security South Africa Conference 2009, School of Tourism & Hospitality, Proceedings ISSA2009*, pages 253–268, University of Johannesburg, Johannesburg, South Africa, Jul. 2009. ISSA, Pretoria, South Africa.
- [100] Nikolaos Tsalis, Nikos Virvilis, Alexios Mylonas, Theodore K. Apostolopoulos, and Dimitris Gritzalis. Browser blacklists: The utopia of phishing protection. In *E-Business and Telecommunications - 11th International Joint Conference, ICETE, Revised Selected Papers*, volume 554 of *Communications in Computer and Information Science*, pages 278–293, Vienna, Austria, Aug. 2014. Springer.

- [101] Ankit Kumar Jain and Brij B. Gupta. A novel approach to protect against phishing attacks at client side using auto-updated white-list. *EURASIP Journal Information Security*, 2016:9, 2016.
- [102] Nalin Asanka Gamagedara Arachchilage and Melissa Cole. Designing a mobile game for home computer users to protect against phishing attacks. *CoRR*, abs/1602.03929, 2016.
- [103] Jemal H. Abawajy. User preference of cyber security awareness delivery methods. *Behaviour and Information Technology*, 33(3):236–247, 2014.
- [104] Nalin Asanka Gamagedara Arachchilage, Steve Love, and Carsten Maple. Can a mobile game teach computer users to thwart phishing attacks? *CoRR*, abs/1511.01622, 2015.
- [105] IsecT Ltd. ISO/IEC 27002:2013 - information technology - security techniques - code of practice for information security controls (second edition). [Online] <https://www.iso27001security.com/html/27002.html>, 2013. Accessed May. 2020.
- [106] Amber van der Heijden and Luca Allodi. Cognitive triaging of phishing attacks. In *28th USENIX Security Symposium, USENIX Security*, pages 1309–1326, Santa Clara, CA, USA, 2019. USENIX Association.
- [107] Verizon. 2019 data enterprise phishing resiliency and defense report breach investigations report. Technical report, Verizon Trademark Services LLC, 2019. [Online] <https://vz.to/2RukvJC>. Accessed Jun. 2020.
- [108] Faris Bugra Kokulu, Ananta Soneji, Tiffany Bao, Yan Shoshitaishvili, Ziming Zhao, Adam Doupé, and Gail-Joon Ahn. Matched and mismatched socs: A qualitative study on security operations center issues. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS*, pages 1955–1970, London, UK, Nov. 2019. ACM.
- [109] Verizon. 2018 data breach investigations report. Technical report, Verizon Trademark Services LLC, 2018. [Online] <https://vz.to/2Rzk8Zw>. Accessed Aug. 2019.
- [110] Krutika Rani Sahu and Jigyasu Dubey. A survey on phishing attacks. *International Journal of Computer Applications*, 88(10):42–45, 2014.



- [111] Pradeepthi K V and Kannan A. Performance study of classification techniques for phishing URL detection. In *2014 Sixth International Conference on Advanced Computing (ICoAC)*, pages 135–139, 2014.
- [112] Aakanksha Tewari, Ankit Kumar Jain, and Brij B. Gupta. Recent survey of various defense mechanisms against phishing attacks. *Journal of Information Privacy and Security*, 12:3–13, 2016.
- [113] Rishi Vaidya. Cyber security breaches survey 2019. Technical report, University of Portsmouth, Social research institute, 2019. [Online] [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/950063/Cyber\\_Security\\_Breaches\\_Survey\\_2019\\_-\\_Main\\_Report\\_-\\_revised\\_V2.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950063/Cyber_Security_Breaches_Survey_2019_-_Main_Report_-_revised_V2.pdf). Accessed Oct. 2021.
- [114] Srishti Gupta and Ponnurangam Kumaraguru. Emerging phishing trends and effectiveness of the anti-phishing landing page. In *2014 APWG Symposium on Electronic Crime Research, eCrime*, pages 36–47, Birmingham, AL, USA, Sep. 2014. IEEE.
- [115] Surbhi Gupta, Abhishek Singhal, and Akanksha Kapoor. A literature survey on social engineering attacks: Phishing attack. In *International Conference on Computing, Communication and Automation (ICCCA)*, pages 537–540. IEEE, 2016.
- [116] V Suganya. A review on phishing attacks and various anti phishing techniques. *International Journal of Computer Applications*, 139, 04 2016.
- [117] Zuochao Dou, Issa Khalil, Abdallah Khreishah, Ala I. Al-Fuqaha, and Mohsen Guizani. Systematization of knowledge (SOK): A systematic review of software-based web phishing detection. *IEEE Communications Surveys Tutorials*, 19(4):2797–2819, 2017.
- [118] Gaurav Varshney, Manoj Misra, and Pradeep K. Atrey. A survey and classification of web phishing detection schemes. *Security and Communication Networks*, 9(18):6266–6284, 2016.
- [119] Dharmaraj Rajaram Patil and B. Patil. Survey on malicious web pages detection techniques. *International Journal of u- and e- Service, Science and Technology*, 8(5):195–206, 2015.

- [120] S. B. Rathod and T. M. Pattewar. A comparative performance evaluation of content based spam and malicious URL detection in e-mail. In *IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS)*, pages 49–54, 2015.
- [121] Yury Zhauniarovich, Issa Khalil, Ting Yu, and Marc Dacier. A survey on malicious domains detection through DNS data analysis. *ACM Computing Surveys*, 51(4):67:1–67:36, 2018.
- [122] A. Haider and R. Singh. Phishing URL detection using neural network optimized by cultural algorithm. *International Journal of Computer Sciences and Engineering*, 6(7):860–863, 2018.
- [123] Guang Xiang, Jason I. Hong, Carolyn Penstein Rosé, and Lorrie Faith Cranor. CANTINA+: A feature-rich machine learning framework for detecting phishing web sites. *ACM Transactions on Information and System Security*, 14(2):21:1–21:28, 2011.
- [124] Mahmoud Khonji, Youssef Iraqi, and Andrew Jones. Lexical URL analysis for discriminating phishing and legitimate e-mail messages. In *6th International Conference for Internet Technology and Secured Transactions, ICITST 2011, Abu Dhabi, UAE, December 11-14, 2011*, pages 422–427. IEEE, 2011.
- [125] Aaron Blum, Brad Wardman, Tamar Solorio, and Gary Warner. Lexical feature based phishing URL detection using online learning. In *Proceedings of the 3rd ACM Workshop on Security and Artificial Intelligence, AISec*, pages 54–60, Chicago, Illinois, USA, 2010. ACM.
- [126] Michael Darling, Greg Heileman, Gilad Gressel, Aravind Ashok, and Prabakaran Poornachandran. A lexical approach for classifying malicious URLs. In *International Conference on High Performance Computing & Simulation, HPCS*, pages 195–202, Amsterdam, Netherlands, Jul. 2015. IEEE.
- [127] Ammar Yahya Daeef, R. Badlishah Ahmad, Yasmin Yacob, and Ng Yen Phing. Wide scope and fast websites phishing detection using URLs lexical features. In *3rd International Conference on Electronic Design (ICED)*, pages 410–415, Phuket, Thailand, 2016.

- [128] Mohammed Nazim Feroz and Susan Mengel. Phishing URL detection using URL ranking. In *2015 IEEE International Congress on Big Data, New York City, NY, USA, June 27 - July 2, 2015*, pages 635–638. IEEE Computer Society, 2015.
- [129] Anh Le, Athina Markopoulou, and Michalis Faloutsos. PhishDef: URL names say it all. In *INFOCOM 2011. 30th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies*, pages 191–195, Shanghai, China, 10-15 April 2011. IEEE.
- [130] V. Santhana Lakshmi and M.S. Vijaya. Efficient prediction of phishing websites using supervised learning algorithms. In *International Conference on Communication Technology and System Design*, volume 30, pages 798–805, 2012.
- [131] Routhu Srinivasa Rao and Alwyn R. Pais. Detecting phishing websites using automation of human behavior. In *Proceedings of the 3rd ACM Workshop on Cyber-Physical System Security, CPSSAsiaCCS 2017, Abu Dhabi, United Arab Emirates, April 2, 2017*, pages 33–42. ACM, 2017.
- [132] J Erkkila. Why we fall for phishing. In *Proceedings of the 2011 CHI Conference on Human Factors in Computing Systems, CHI '11*, pages 1–8, ancouver, BC, Canada, 2011. ACM.
- [133] Aiping Xiong, Robert W. Proctor, Weining Yang, and Ninghui Li. Is domain highlighting actually helpful in identifying phishing web pages? *Human Factors*, 59(4):640–660, 2017.
- [134] RFC1738: Uniform Resource Locators (url).
- [135] Ram B. Basnet and Tenzin Doleck. Towards developing a tool to detect phishing URLs : A machine learning approach. In *IEEE International Conference on Computational Intelligence Communication Technology*, pages 220–223, Ghaziabad, India, 2015.
- [136] Ram B. Basnet, Andrew H. Sung<sup>2</sup>, and Quingzhong Liu. Learning to detect phishing URLs. *International Journal of Research in Engineering and Technology, IJRET*, 3(6):11–24, 2014.
- [137] Youness Mourtaji, Mohammed Bouhorma, and Alghazzawi. Perception of a new framework for detecting phishing web pages. In *Proceedings of the*

*Mediterranean Symposium on Smart City Application, SCAMS '17*, New York, NY, USA, 2017. Association for Computing Machinery.

- [138] L. A. T. Nguyen, B. L. To, H. K. Nguyen, and M. H. Nguyen. A novel approach for phishing detection using URL-based heuristic. In *2014 International Conference on Computing, Management and Telecommunications (ComManTel)*, pages 298–303, 2014.
- [139] Kang-Leng Chiew, Ee Hung Chang, San-Nah Sze, and Wei King Tiong. Utilisation of website logo for phishing detection. *Computers and Security*, 54:16–26, 2015.
- [140] Hassan Y. A. Abutair and Abdelfettah Belghith. Using case-based reasoning for phishing detection. In *The 8th International Conference on Ambient Systems, Networks and Technologies (ANT 2017) / The 7th International Conference on Sustainable Energy Information Technology (SEIT 2017), 16-19 May 2017, Madeira, Portugal*, volume 109 of *Procedia Computer Science*, pages 281–288. Elsevier, 2017.
- [141] Chunlin Liu, Lidong Wang, Bo Lang, and Yuan Zhou. Finding effective classifier for malicious URL detection. In *Proceedings of the 2nd International Conference on Management Engineering, Software Engineering and Service Sciences, ICMSS 2018*, pages 240–244, New York, NY, USA, 2018. Association for Computing Machinery.
- [142] Senhao Wen, Zhiyuan Zhao, and Hanbing Yan. Detecting malicious websites in depth through analyzing topics and web-pages. In *Proceedings of the 2nd International Conference on Cryptography, Security and Privacy, ICCSP*, pages 128–133, Guiyang, China, Mar. 2018. ACM.
- [143] Dongsong Zhang, Zhijun Yan, Hansi Jiang, and Taeha Kim. A domain-feature enhanced classification model for the detection of Chinese phishing e-business websites. *Information and Management*, 51(7):845–853, 2014.
- [144] Giovanni Bottazzi, Emiliano Casalicchio, Davide Cingolani, Fabio Marturana, and Marco Piu. Mp-shield: A framework for phishing detection in mobile devices. In *15th International Conference on Computer and Information Technology, CIT; 14th International Conference on Ubiquitous Computing and Communications, IUCC; 13th International Conference on Dependable, Autonomic and*

- Secure Computing, DASC; 13th International Conference on Pervasive Intelligence and Computing, PICom*, pages 1977–1983, Liverpool, United Kingdom, Oct. 2015. IEEE.
- [145] Rakesh M. Verma and Keith Dyer. On the character of phishing URLs : Accurate and robust statistical learning classifiers. In *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy, CODASPY*, pages 111–122, San Antonio, TX, USA, Mar. 2015. ACM.
- [146] Varsharani Ramdas Hawanna, VY Kulkarni, and RA Rane. A novel algorithm to detect phishing URLs. In *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*, pages 548–552. IEEE, 2016.
- [147] N. A. Azeez and A. Oluwatosin. CyberProtector: Identifying compromised URLs in electronic mails with bayesian classification. In *International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 959–965, Dec. 2016.
- [148] P. Singh, Y. P. S. Maravi, and S. Sharma. Phishing websites detection through supervised learning networks. In *2015 International Conference on Computing and Communications Technologies (ICCCT)*, pages 61–65, Chennai, India, 26-27 Feb. 2015. IEEE.
- [149] Jin-Lee Lee, Dong-Hyun Kim, and Lee Chang-Hoon. Heuristic-based approach for phishing site detection using URL features. In *Third International Conference on Advances in Computing, Electronics and Electrical Technology-CEET*, pages 131–135, 2015.
- [150] Firdous Kausar, Bushra Al-Otaibi, Asma Al-Qadi, and Nwayer Al-Dossari. Hybrid client side phishing websites detection approach. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 5(7):132–140, 2014.
- [151] Wenwu Chen, Wei Zhang, and Yang Su. Phishing detection research based on LSTM recurrent neural network. In *Data Science - 4th International Conference of Pioneering Computer Scientists, Engineers and Educators, ICPCSEE 2018, Zhengzhou, China, September 21-23, 2018, Proceedings, Part I*, volume 901 of *Communications in Computer and Information Science*, pages 638–645. Springer, 2018.

- [152] Ankit Kumar Jain and B. B. Gupta. PHISH-SAFE: URL features-based phishing detection system using machine learning. In *Cyber Security*, pages 467–474, Singapore, 2018. Springer Singapore.
- [153] S. Gupta. Efficient malicious domain detection using word segmentation and BM pattern matching. In *2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, pages 1–6, 2016.
- [154] Guolin Tan, Peng Zhang, Qingyun Liu, Xinran Liu, Chungze Zhu, and Fenghu Dou. Adaptive malicious URL detection: Learning in the presence of concept drifts. In *17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications / 12th IEEE International Conference On Big Data Science And Engineering, TrustCom/BigDataSE*, pages 737–743, New York, NY, USA, Aug. 2018. IEEE.
- [155] Sushma Nagesh Bannur, Lawrence K. Saul, and Stefan Savage. Judging a site by its content: learning the textual, structural, and visual features of malicious web pages. In *Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence, AISec*, pages 1–10, Chicago, IL, USA, Oct. 2011. ACM.
- [156] Davide Canali, Marco Cova, Giovanni Vigna, and Christopher Kruegel. Prophiler: a fast filter for the large-scale detection of malicious web pages. In *Proceedings of the 20th International Conference on World Wide Web, WWW*, pages 197–206, Hyderabad, India, 2011. ACM.
- [157] Bhagyashree E. Sananse and Tanuja K. Sarode. Phishing URL detection: A machine learning and web mining-based approach. *International Journal of Computer Applications*, 123(13):46–50, Aug. 2015.
- [158] Min-Sheng Lin, Chien-Yi Chiu, Yuh-Jye Lee, and Hsing-Kuo Pao. Malicious URL filtering - A big data application. In *Proceedings of the 2013 International Conference on Big Data, 6-9 October 2013*, pages 589–596, Santa Clara, CA, USA, 2013. IEEE.
- [159] Hsing-Kuo Pao, Yan-Lin Chou, and Yuh-Jye Lee. Malicious URL detection based on Kolmogorov Complexity Estimation. In *2012 IEEE/WIC/ACM International Conferences on Web Intelligence, WI*, pages 380–387, Macau, China, Dec. 2012. IEEE Computer Society.

- [160] Goutam Chakraborty and Tsai Tzung Lin. A URL address aware classification of malicious websites for online security during web-surfing. In *2017 International Conference on Advanced Networks and Telecommunications Systems, ANTS*, pages 1–6, Bhubaneswar, India, Dec. 2017. IEEE.
- [161] Haotian Liu, Xiang Pan, and Zhengyang Qu. Learning based malicious web sites detection using suspicious URLs. *Last accessed January, 2016*.
- [162] Justin Ma, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. Beyond blacklists: learning to detect malicious web sites from suspicious URLs. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009*, pages 1245–1254. ACM, 2009.
- [163] R. Kumar, X. Zhang, H. A. Tariq, and R. U. Khan. Malicious URL detection using multi-layer filtering model. In *2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pages 97–100, Chengdu, China, Dec. 2017. IEEE.
- [164] Gamze Canova, Melanie Volkamer, Clemens Bergmann, Roland Borza, Benjamin Reinheimer, Simon Stockhardt, and Ralf Tenberg. Learn to spot phishing URLs with the Android NoPhish app. In *Information Security Education Across the Curriculum - 9th IFIP WG 11.8 World Conference, WISE9, Hamburg, Germany, May 26-28, 2015, Proceedings*, volume 453 of *IFIP Advances in Information and Communication Technology*, pages 87–100. Springer, 2015.
- [165] S. Carolin Jeeva and Elijah Blessing Rajasingh. Intelligent phishing URL detection using association rule mining. *Human-centric Computing and Information Sciences*, 6:10, 2016.
- [166] Ying Xue, Yang Li, Yuangang Yao, Xianghui Zhao, Jianyi Liu, and Ru Zhang. Phishing sites detection based on URL correlation. In *4th International Conference on Cloud Computing and Intelligence Systems, CCIS*, pages 244–248, Beijing, China, Aug. 2016. IEEE.
- [167] A. R. Nagaonkar and U. L. Kulkarni. Finding the malicious URLs using search engines. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 3692–3694, New Delhi, India, 16-18 March 2016. IEEE.

- [168] Nalin Asanka Gamagedara Arachchilage, Steve Love, and Konstantin Beznosov. Phishing threat avoidance behaviour: An empirical investigation. *Computers in Human Behavior*, 60:185–197, 2016.
- [169] Rami Mustafa A. Mohammad, Fadi A. Thabtah, and Lee McCluskey. Intelligent rule-based phishing websites classification. *IET Information Security*, 8(3):153–160, 2014.
- [170] Sonu Gupta and Shelly Sachdeva. Invitation or bait? detecting malicious URLs in facebook events. In *2018 Eleventh International Conference on Contemporary Computing, IC3*, pages 1–6, Noida, India, Aug. 2018. IEEE Computer Society.
- [171] Samuel Marchal, Kalle Saari, Nidhi Singh, and N. Asokan. Know your phish: Novel techniques for detecting phishing sites and their targets. In *36th International Conference on Distributed Computing Systems, ICDCS*, pages 323–333, Nara, Japan, Jun. 2016. IEEE.
- [172] Justin Ma, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. Learning to detect malicious URLs. *ACM Transactions on Intelligent Systems and Technology*, 2(3):30:1–30:24, 2011.
- [173] Samuel Marchal, Giovanni Armano, Tommi Grondahl, Kalle Saari, Nidhi Singh, and N. Asokan. Off-the-Hook: An efficient and usable client-side phishing prevention application. *IEEE Transactions on Computers*, 66(10):1717–1733, 2017.
- [174] Mohammed Nazim Feroz and Susan Mengel. Examination of data, rule generation and detection of phishing URLs using online logistic regression. In *2014 IEEE International Conference on Big Data, Big Data 2014, Washington, DC, USA, October 27-30, 2014*, pages 241–250. IEEE Computer Society, 2014.
- [175] Ke-Wei Su, Kuo-Ping Wu, Hahn-Ming Lee, and Te-En Wei. Suspicious URL filtering based on logistic regression with multi-view analysis. In *Eighth Asia Joint Conference on Information Security, AsiaJCIS*, pages 77–84, Seoul, Korea, Jul. 2013. IEEE Computer Society.
- [176] Mahmood Moghimi and Ali Yazdian Varjani. New rule-based phishing detection method. *Expert Systems with Applications*, 53:231–242, 2016.



- [177] Hossein Shirazi, Bruhadeshwar Bezawada, and Indrakshi Ray. “Kn0w Thy Domain Nam”: Unbiased phishing detection using domain name based features. In *Proceedings of the 23rd ACM on Symposium on Access Control Models and Technologies, SACMAT 2018, Indianapolis, IN, USA, June 13-15, 2018*, pages 69–75. ACM, 2018.
- [178] R Hamsa Veni, A Hariprasad Reddy, and C Kesavulu. Identifying malicious web links and their attack types in social networks. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 3(4):1060–1066, 2018.
- [179] Sujata Garera, Niels Provos, Monica Chew, and Aviel D. Rubin. A framework for detection and measurement of phishing attacks. In *Proceedings of the 2007 ACM Workshop on Recurring Malcode, WORM ’07*, page 1–8, New York, NY, USA, 2007. Association for Computing Machinery.
- [180] Gamze Canova, Melanie Volkamer, Clemens Bergmann, and Roland Borza. NoPhish: An anti-phishing education app. In *Security and Trust Management - 10th International Workshop, STM 2014, Wroclaw, Poland, September 10-11, 2014. Proceedings*, volume 8743 of *Lecture Notes in Computer Science*, pages 188–192. Springer, 2014.
- [181] Binod Gyawali, Tamar Solorio, Manuel Montes-y-Gómez, Brad Wardman, and Gary Warner. Evaluating a semisupervised approach to phishing URL identification in a realistic scenario. In *The 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference, CEAS 2011, Perth, Australia, September 1-2, 2011, Proceedings*, pages 176–183. ACM, 2011.
- [182] Justin Ma, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. Identifying suspicious URLs : an application of large-scale online learning. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML 2009, June 14-18, 2009*, volume 382, pages 681–688, Montreal, Quebec, Canada, 2009. ACM.
- [183] Jianyi Zhang, Yang Pan, Zhiqiang Wang, and Biao Liu. URL based gateway side phishing detection method. In *Trustcom/BigDataSE/ISPA, August 23-26, 2016*, pages 268–275, Tianjin, China, 2016. IEEE.

- [184] Huiping Yao and Dongwan Shin. Towards preventing QR code based attacks on Android phone using security warnings. In *8th ACM Symposium on Information, Computer and Communications Security, ASIA CCS '13*, pages 341–346, Hangzhou, China, May. 2013. ACM.
- [185] Pawan Prakash, Manish Kumar, Ramana Rao Kompella, and Minaxi Gupta. PhishNet: Predictive blacklisting to detect phishing attacks. In *INFOCOM. 29th International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies*, pages 346–350, San Diego, CA, USA, 2010. IEEE.
- [186] Bo Hong, Wei Wang, Liming Wang, Guanggang Geng, Yali Xiao, Xiaodong Li, and Wei Mao. A hybrid system to find & fight phishing attacks actively. In *Proceedings of the 2011 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2011, Campus Scientifique de la Doua, Lyon, France, August 22-27, 2011*, pages 506–509. IEEE Computer Society, 2011.
- [187] Bandar Alghamdi, Jason Watson, and Yue Xu. Toward detecting malicious links in online social networks through user behavior. In *IEEE/WIC/ACM International Conference on Web Intelligence - Workshops, WI Workshops*, pages 5–8, Omaha, NE, USA, Oct. 2016. IEEE Computer Society.
- [188] A. K. Jain and B. B. Gupta. Comparative analysis of features based machine learning approaches for phishing detection. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 2125–2130, New Delhi, India, 16-18 March 2016. IEEE.
- [189] Fang Lv, Bailing Wang, Junheng Huang, Yushan Sun, and Yuliang Wei. A proactive discovery and filtering solution on phishing websites. In *International Conference on Big Data, Big Data*, pages 2348–2355, Santa Clara, CA, USA, 2015. IEEE Computer Society.
- [190] Shuang Hao, Nick Feamster, and Ramakant Pandrangi. Monitoring the initial DNS behavior of malicious domains. In *Proceedings of the 11th ACM SIGCOMM Internet Measurement Conference, IMC '11*, pages 269–278, Berlin, Germany, 2011. ACM.
- [191] Gaurav Varshney, Manoj Misra, and Pradeep K. Atrey. A phish detector using lightweight search features. *Computer Security*, 62:213–228, 2016.

- [192] Yue Zhang, Serge Egelman, Lorrie Cranor, and Jason Hong. Phinding phish: Evaluating anti-phishing tools. In *the 14th Annual Network & Distributed System Security Symposium NDSS*, San Diego, CA, 2007. Carnegie Mellon University.
- [193] Lung-Hao Lee, Kuei-Ching Lee, Hsin-Hsi Chen, and Yuen-Hsien Tseng. POSTER: proactive blacklist update for anti-phishing. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*, pages 1448–1450. ACM, 2014.
- [194] Weining Yang, Aiping Xiong, Jing Chen, Robert W. Proctor, and Ninghui Li. Use of phishing training to improve security warning compliance: Evidence from a field experiment. In *Proceedings of the Hot Topics in Science of Security: Symposium and Bootcamp, HoTSoS*, pages 52–61, Hanover, MD, USA, Apr. 2017. ACM.
- [195] Samuel Marchal, Jérôme François, Radu State, and Thomas Engel. PhishStorm: Detecting phishing with streaming analytics. *IEEE Transactions on Network and Service Management*, 11(4):458–471, 2014.
- [196] Chris Grier, Kurt Thomas, Vern Paxson, and Chao Michael Zhang. @Spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM Conference on Computer and Communications Security, CCS 2010, October 4-8, 2010*, pages 27–37, Chicago, Illinois, USA, 2010. ACM.
- [197] Neha Gupta, Anupama Aggarwal, and Ponnurangam Kumaraguru. bit.ly/malicious: Deep dive into short URL based e-crime detection. In *APWG Symposium on Electronic Crime Research, eCrime*, pages 14–24, Birmingham, AL, USA, Sep. 2014. IEEE.
- [198] N. S. Gawale and N. N. Patil. Implementation of a system to detect malicious URLs for Twitter users. In *2015 International Conference on Pervasive Computing (ICPC)*, pages 1–5, Pune, India, 8-10 Jan. 2015. IEEE.
- [199] Sangho Lee and Jong Kim. WarningBird: A near real-time detection system for suspicious URLs in Twitter stream. *IEEE Transactions on Dependable and Secure Computing*, 10(3):183–195, 2013.

- [200] Ivan Torroledo, Luis David Camacho, and Alejandro Correa Bahnsen. Hunting malicious TLS certificates with deep neural networks. In *Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security, CCS 2018, Toronto, ON, Canada, October 19, 2018*, pages 64–73. ACM, 2018.
- [201] Ksenia Tsyganok, Evgeny Tumoyan, and Lyudmila K. Babenko. Development the method of detection the malicious pages interconnection in the internet. In *The 6th International Conference on Security of Information and Networks, SIN '13, Aksaray, Turkey, November 26-28, 2013*, pages 433–435. ACM, 2013.
- [202] Onur Catakoglu, Marco Balduzzi, and Davide Balzarotti. Automatic extraction of indicators of compromise for web applications. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*, pages 333–343. ACM, 2016.
- [203] Mitsuaki Akiyama, Takeshi Yagi, and Mitsutaka Itoh. Searching structural neighborhood of malicious URLs to improve blacklisting. In *IEEE/IPSJ International Symposium on Applications and the Internet Searching*, 2011.
- [204] Cheng Hsin Hsu, Polo Wang, and Samuel Pu. Identify fixed-path phishing attack by STC. In *The 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference, CEAS 2011, Perth, Australia, September 1-2, 2011, Proceedings*, pages 172–175. ACM, 2011.
- [205] Ye Cao, Weili Han, and Yueran Le. Anti-phishing based on automated individual white-list. In *Proceedings of the 4th Workshop on Digital Identity Management*, pages 51–60, Alexandria, VA, USA, Oct. 2008. ACM.
- [206] Hiba Zuhair Zeydan, Ali Selamat, and Mazleena Salleh. Feature selection for phishing detection: a review of research. *Journal of Intelligent Systems Technologies and Applications IJISTA*, 15(2):147–162, 2016.
- [207] Microsoft. Microsoft security intelligence report. Technical report, Microsoft, 2018. [Online] <https://www.microsoft.com/en-us/security/intelligence-report>. Accessed Aug. 2018.
- [208] Emma J. Williams, Joanne Hinds, and Adam N. Joinson. Exploring susceptibility to phishing in the workplace. *International Journal of Human Computer Studies*, 120:1–13, 2018.

- [209] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A “Nutrition Label” for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security, SOUPS*, pages 1–a12, Mountain View, California, USA, Jul. 2009. ACM.
- [210] Mattia Mossano, Kami Vaniea, Lukas Aldag, Reyhan Düzgün, Peter Mayer, and Melanie Volkamer. Analysis of publicly available anti-phishing webpages: contradicting information, lack of concrete advice and very narrow attack vector. In *European Symposium on Security and Privacy Workshops, EuroS&P Workshops*, pages 130–139, Genoa, Italy, Sep. 2020. IEEE.
- [211] Iulia Ion, Rob Reeder, and Sunny Consolvo. “...No one Can Hack My Mind”: Comparing expert and non-expert security practices. In *Eleventh Symposium On Usable Privacy and Security, SOUPS*, pages 327–346, Ottawa, Canada, Jul. 2015. USENIX Association.
- [212] Robert W. Reeder, Iulia Ion, and Sunny Consolvo. 152 simple steps to stay safe online: Security advice for non-tech-savvy users. *IEEE Security & Privacy*, 15(5):55–64, 2017.
- [213] Timothy Kelley and Bennett I. Bertenthal. Attention and past behavior, not security knowledge, modulate users’ decisions to login to insecure websites. *Information and Computer Security*, 24(2):164–176, 2016.
- [214] Rashid Tahir, Ali Raza, Faizan Ahmad, Jehangir Kazi, Fareed Zaffar, Chris Kanich, and Matthew Caesar. It’s all in the name: Why some URLs are more vulnerable to typosquatting. In *Conference on Computer Communications, INFOCOM 2018, April 16-19, 2018*, pages 2618–2626, Honolulu, HI, USA, 2018. IEEE.
- [215] Evgeniy Gabrilovich and Alex Gontmakher. The homograph attack. *Communications of the ACM*, 45(2):128, 2002.
- [216] Gabor Szathmari. Why outdated anti-phishing advice leaves you exposed (part 2). <https://blog.ironbastion.com.au/why-outdated-anti-phishing-advice-leaves-you-exposed-part-2/>, Jul. 2020.

- [217] Janos Szurdi, Balazs Kocso, Gabor Cseh, Jonathan Spring, Márk Félegyházi, and Chris Kanich. The long “Taile” of typosquatting domain names. In *Proceedings of the 23rd USENIX Security Symposium*, pages 191–206, San Diego, CA, USA, Aug. 2014. USENIX Association.
- [218] Florian Quinkert, Tobias Lauinger, William K. Robertson, Engin Kirda, and Thorsten Holz. It’s not what it looks like: Measuring attacks and defensive registrations of homograph domains. In *7th Conference on Communications and Network Security, CNS 2019, June 10-12, 2019*, pages 259–267, Washington, DC, USA, 2019. IEEE.
- [219] Maria Sameen, Kyunghyun Han, and Seong Oun Hwang. Phishhaven- an efficient real-time AI phishing URLs detection system. *IEEE Access*, 8:83425–83443, 2020.
- [220] Lucian Constantin. Attackers host phishing pages on Azure. [Online] <https://securityboulevard.com/2019/03/attackers-host-phishing-pages-on-azure/>, Mar. 2019. Accessed Jun. 2019.
- [221] Krishna Bhargava, Douglas Brewer, and Kang Li. A study of URL redirection indicating spam. In *Sixth conference on e-mail and anti-spam CEAS*, pages 1–4, California, USA, 2009. Steve Sheng’s Publications.
- [222] Charles A. O’Reilly. Individuals and information overload in organizations: Is more necessarily better? *The Academy of Management Journal*, 23(4):684–696, 1980.
- [223] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. Standardizing privacy notices: An online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’10*, page 1573–1582, New York, NY, USA, 2010. Association for Computing Machinery.
- [224] Gaurav Misra, Nalin Asanka Gamagedara Arachchilage, and Shlomo Berkovsky. Phish phinder: A game design approach to enhance user confidence in mitigating phishing attacks. In *Eleventh International Symposium on Human Aspects of Information Security & Assurance, HAISA, Proceedings*, pages 41–51, Adelaide, Australia, Nov. 2017. University of Plymouth.

- [225] Nalin Asanka Gamagedara Arachchilage and Steve Love. Security awareness of computer users: A phishing threat avoidance perspective. *Computers in Human Behavior*, 38:304–312, 2014.
- [226] Iacovos Kirlappos and Martina Angela Sasse. Security education against phishing: A modest proposal for a major rethink. *IEEE Security and Privacy*, 10(2):24–32, 2012.
- [227] Ulrike Meyer and Vincent Drury. Certified phishing: Taking a look at public key certificates of phishing websites. In *Fifteenth Symposium on Usable Privacy and Security, SOUPS*, pages 210–223, Santa Clara, CA, USA, Aug. 2019. USENIX Association.
- [228] Let’s Encrypt. Free SSL/TLS certificates. <https://letsencrypt.org/>, 2019. Accessed Dec. 2020.
- [229] Joseph Johnson. UK: number of internet users who are students 2011-2019. <https://www.statista.com/statistics/940040/number-of-student-internet-users-in-the-uk/>, May. 2019.
- [230] Bernhard Jenny and Nathaniel Vaughn Kelso. Color design for the color vision impaired. *Cartographic Perspectives*, 58:61–67, 2007.
- [231] Dan J. Graham, Jacob L. Orquin, and Vivianne H.M. Visschers. Eye tracking and nutrition label use: A review of the literature and recommendations for label enhancement. *Food Policy*, 37(4):378–382, 2012.
- [232] Nuttapon Sanglerdsinlapachai and Arnon Rungsawang. Using domain top-page similarity feature in machine learning-based web phishing detection. In *Third International Conference on Knowledge Discovery and Data Mining, WKDD*, pages 187–190, Phuket, Thailand, 9-10 January 2010. IEEE.
- [233] Fortinet. Web filter categories. <https://www.fortiguard.com/webfilter/categories>, Jan. 2021. Accessed Aug. 2020.
- [234] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 30(1-7):107–117, 1998.
- [235] LLC OpenDNS. PhishTank: Join the fight against phishing. <https://www.phishtank.com/>, 2019. Accessed Dec. 2020.

- [236] OpenPhish. OpenPhish: Phishing intelligence. <https://openphish.com>, 2019. Accessed Dec. 2020.
- [237] CMBuild. Archive of dmoz.org. <https://dmoz-odp.org/Reference/>, 2013. Accessed Dec. 2020.
- [238] Philipp Koehn, Huda Khayrallah, Kenneth Heafield, and Mikel L. Forcada. Findings of the WMT 2018 shared task on parallel corpus filtering. In *Proceedings of the Third Conference on Machine Translation: Shared Task Papers, WMT 2018, October 31 - November 1, 2018*, pages 726–739, Belgium, Brussels, 2018. Association for Computational Linguistics.
- [239] Ryan T. Wright and Kent Marett. The influence of experiential and dispositional factors in phishing: An empirical investigation of the deceived. *Journal of Management Information Systems*, 27(1):273–303, 2010.
- [240] Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. Privacy attitudes of Mechanical Turk workers and the U.S. public. In *10th Symposium on Usable Privacy and Security, SOUPS*, pages 37–49, Menlo Park, CA, USA, Jul. 2014. USENIX Association.
- [241] Sara Albakry and Kami Vaniea. Automatic phishing detection versus user training, is there a middle ground using XAI? In *Proceedings of the SICSA Workshop on Reasoning, Learning and Explainability*, volume 2151 of *CEUR Workshop Proceedings*, pages 1–2, Aberdeen, Scotland, UK, Jun. 2018. CEUR-WS.org.
- [242] Patrickson Weanquoi, Jaris Johnson, and Jinghua Zhang. Using a game to teach about phishing. In *Proceedings of the 18th Annual Conference on Information Technology Education and the 6th Annual Conference on Research in Information Technology*, page 75, Rochester, New York, USA, Oct. 2017. ACM.
- [243] Ponnurangam Kumaraguru, Yong Rhee, Alessandro Acquisti, Lorrie Faith Cranor, Jason I. Hong, and Elizabeth Nunge. Protecting people from phishing: the design and evaluation of an embedded training email system. In *Proceedings of the Conference on Human Factors in Computing Systems, CHI*, pages 905–914, San Jose, California, USA, Apr. 2007. ACM.
- [244] Elissa M. Redmiles, Amelia R. Malone, and Michelle L. Mazurek. I think they’re trying to tell me something: Advice sources and selection for digital



- security. In *Symposium on Security and Privacy, SP, May 22-26*, pages 272–288, San Jose, CA, USA, 2016. IEEE.
- [245] Stephen Waddell. Catchphish: A URL and anti-phishing research platform. Master's thesis, University of Edinburgh, Mar. 2020.
- [246] Media & Sport Department for Digital, Culture. Official statistics cyber security breaches survey 2020– chapter 5: Incidence and impact of breaches or attacks. Technical report, National Cyber Security Centre, May. 2020. [Online] <https://www.gov.uk/government/publications/cyber-security-breaches-survey-2020/cyber-security-breaches-survey-2020>. Accessed Jan. 2021.
- [247] ProofPoint Phishing Report. State of the phish– an in-depth look at user awareness, vulnerability and resilience. Technical Report 1, Proofpoint, Inc., 2020. [Online] <https://www.proofpoint.com/sites/default/files/gtd-pfpt-us-tr-state-of-the-phish-2020.pdf>. Accessed Jan. 2021.
- [248] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. How I learned to be secure: a census-representative survey of security advice sources and behavior. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, October 24-28*, pages 666–677, Vienna, Austria, 2016. ACM.
- [249] Report phishing sites. <https://www.us-cert.gov/report-phishing>, 2020. Accessed May. 2020.
- [250] SparkCMS by Baunfire.com. Report phishing. <https://education.apwg.org/report-cybercrime/>, 2020. Accessed 2019.
- [251] Eben M. Haber and Eser Kandogan. Security administrators: A breed apart. In *SOUPS Workshop on Usable IT Security Management (USM)*, pages 3–6, Pittsburgh, PA, USA, 2007. CiteSeerX.
- [252] Edwin Donald Frauenstein and Rossouw von Solms. An enterprise anti-phishing framework. In *Information Assurance and Security Education and Training - 8th IFIP WG 11.8 World Conference on Information Security Education, WISE 8, Proceedings*, volume 406 of *IFIP Advances in Information*

- and Communication Technology*, pages 196–203, Auckland, New Zealand, Jul. 2013. Springer.
- [253] Rand Abu Hammour, Yousef Al Gharaibeh, Malik Qasaimeh, and Raad S. Al-Qassas. The status of information security systems in banking sector from social engineering perspective. In *Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems, DATA*, pages 14:1–14:7, Dubai, UAE, Dec. 2018. ACM.
- [254] Heidi Wilcox and Maumita Bhattacharya. A framework to mitigate social engineering through social media within the enterprise. In *2016 11th Conference on Industrial Electronics and Applications (ICIEA)*, pages 1039–1044, Hefei, China, 2016. IEEE.
- [255] George Grispos, William Bradley Glisson, David Bourrie, Tim Storer, and Stacy Miller. Security incident recognition and reporting (SIRR): an industrial perspective. In *23rd Americas Conference on Information Systems, AMCIS*, pages 1–10, Boston, MA, USA, Aug. 2017. Association for Information Systems.
- [256] Stefan Metzger, Wolfgang Hommel, and Helmut Reiser. Integrated security incident management - concepts and real-world experiences. In *Sixth International Conference on IT Security Incident Management and IT Forensics, IMF*, pages 107–121, Stuttgart, Germany, May. 2011. IEEE Computer Society.
- [257] Thomas Maillart, Mingyi Zhao, Jens Grossklags, and John Chuang. Given enough eyeballs, all bugs are shallow? revisiting Eric Raymond with bug bounty programs. *Journal of Cybersecurity*, 3(2):81–90, 2017.
- [258] Ankit Shah, Rajesh Ganesan, Sushil Jajodia, and Hasan Cam. Understanding tradeoffs between throughput, quality, and cost of alert analysis in a CSOC. *IEEE Transactions on Information Forensics and Security*, 14(5):1155–1170, 2019.
- [259] Phishing attacks: dealing with suspicious emails and messages. <http://bit.ly/3tTwQpC>, Dec. 2018. Accessed Aug. 2019.
- [260] Youngsun Kwak, Seyoung Lee, Amanda Damiano, and Arun Vishwanath. Why do users not report spear phishing emails? *Telematics Informatics*, 48:101343, 2020.

- [261] Matthew L. Jensen, Alexandra Durcikova, and Ryan T. Wright. Combating phishing attacks: A knowledge management approach. In *50th Hawaii International Conference on System Sciences, HICSS*, pages 1–10, Hilton Waikoloa Village, Hawaii, USA, January 4-7, 2017. ScholarSpace / AIS Electronic Library (AISeL).
- [262] Atif Ahmad, Justin Hadgkiss, and Anthonie B. Ruighaver. Incident response teams - challenges in supporting the organisational security function. *Computer Security*, 31(5):643–652, 2012.
- [263] David Botta, Kasia Muldner, Kirstie Hawkey, and Konstantin Beznosov. Toward understanding distributed cognition in IT security management: the role of cues and norms. *Cognition, Technology & Work*, 13(2):121–134, 2011.
- [264] Albese Demjaha, Tristan Caulfield, M. Angela Sasse, and David J. Pym. 2 fast 2 secure: A case study of post-breach security changes. In *European Symposium on Security and Privacy Workshops, EuroS&P Workshops*, pages 192–201, Stockholm, Sweden, Jun. 2019. IEEE.
- [265] Suhaila Ismail, Arniyati Ahmad, and Mohd Afizi Mohd Shukran. New method of forensic computing in a small organization. *Australian Journal of Basic and Applied Sciences*, 5(9):2019e25, 2011.
- [266] Christina Lekati. Complexities in investigating cases of social engineering: How reverse engineering and profiling can assist in the collection of evidence. In *11th International Conference on IT Security Incident Management & IT Forensics (IMF)*, pages 107–109, Hamburg, Germany, 2018. IEEE.
- [267] Ying He and Chris Johnson. Challenges of information security incident learning: An industrial case study in a Chinese healthcare organization. *Informatics for Health and Social Care*, 42(4):393–408, 2017. Pmid: 28068150.
- [268] Piya Shedden, Atif Ahmad, and A B. Ruighaver. Organisational learning and incident response: Promoting effective learning through the incident response process. In *Proceedings of the 8th Australian Information Security Management Conference*, pages 131–142, Perth, Australia, 2010. Edith Cowan University.
- [269] Marshall A Kuypers, Thomas Maillart, and Elisabeth Paté-Cornell. An empirical analysis of cyber security incidents at a large organization. *Department of*

*Management Science and Engineering, Stanford University, School of Information, UC Berkeley*, 30:1–22, 2016.

- [270] Inger Anne Tøndel, Maria B. Line, and Martin Gilje Jaatun. Information security incident management: Current practice as reported in the literature. *Computers and Security*, 45:42–57, 2014.
- [271] M. A. Sasse, S. Brostoff, and D. Weirich. Transforming the 'weakest link' – a human/computer interaction approach to usable and effective security. *BT Technology Journal*, 19(3):122–131, Jul. 2001.
- [272] COBIT: Control Objectives for Information Technologies. [Online]<https://www.isaca.org/resources/cobit>, 2020. Accessed 10 Oct. 2020.
- [273] ITIL-IT service management. <https://www.axelos.com/best-practice-solutions/itil>, 2021. Accessed 10 Oct. 2020.
- [274] Carol Pollard and Aileen Cater-Steel. Justifications, strategies, and critical success factors in successful ITIL implementations in U.S. and Australian companies: An exploratory study. *Information Systems Management*, 26(2):164–175, 2009.
- [275] BC Potgieter, JH Botha, and C Lew. Evidence that use of the ITIL framework is effective. In *18th Annual conference of the national advisory committee on computing qualifications*, pages 423–427, Tauranga, NZ, 2005. CiteSeerX, CiteSeerX.
- [276] V R Palilingan and J R Batmetan. Incident management in academic information system using ITIL framework. *IOP Conference Series: Materials Science and Engineering*, 306:012110, Feb. 2018.
- [277] M. Jäntti. Examining challenges in IT service desk system and processes: A case study. In *The Seventh International Conference on Systems(ICONs)*, pages 105–108, 2012.
- [278] Xiaojun Tang and Yuki Todo. A study of service desk setup in implementing IT service management in enterprises. *Technology and Investment*, 4(3):190–196, 2013.

- [279] Abtin Refahi Farjadi Tehrani and Faras Zuheir Mustafa Mohamed. A CBR-based approach to ITIL-based service desk. *Journal of Emerging Trends in Computing and Information Sciences*, 2(10):476—484, 2011.
- [280] Norshidah Mohamed and Jasber Kaur A. P. Gian Singh. A conceptual framework for information technology governance effectiveness in private organizations. *Information Management and Computer Security*, 20(2):88–106, 2012.
- [281] Lynne M. Coventry and T. B. Kane. The automation of helpdesks. In *Proceedings of the 1st International Workshop on Intelligent User Interfaces, IUI*, pages 219–222, Orlando, Florida, USA, Jan. 1993. ACM.
- [282] Robert Prince, Jianwen Su, Hong Tang, and Yonggang Zhao. The design of an interactive online help desk in the alexandria digital library. In *Proceedings of the international joint conference on Work activities coordination and collaboration*, pages 217–226, San Francisco, California, USA, 1999. ACM.
- [283] Matthew J. Conlon. Overhaul your helpdesk ticketing system. In *Proceedings of the 35th Annual ACM SIGUCCS Conference on User Services*, pages 37–40, Orlando, Florida, USA, Oct. 2007. ACM.
- [284] Christian J. Sinnett and Tammy Barr. Building a champagne helpdesk on a beer budget. In *Proceedings of the 32nd Annual ACM SIGUCCS Conference on User Services*, pages 351–353, Baltimore, MD, USA, Oct. 2004. ACM.
- [285] Rachael Cottam, Jeff Goff, and Peter Nguyen. Extending the centralized helpdesk functionality to improve decentralized support. In *ACM SIGUCCS Annual Conference, SIGUCCS '12*, pages 153–156, Memphis, TN, USA, Oct. 2012. ACM.
- [286] Nikhil Sharma. Sensemaking handoff: When and how? In *People Transforming Information - Information Transforming People - Proceedings of the 71st ASIS&T Annual Meeting, ASIST*, volume 45 of *Proceedings of the Association for Information Science and Technology*, pages 1–12, Columbus, OH, USA, 2008. Wiley.
- [287] Lena Mamykina and Catherine G. Wolf. Evolution of contact point: a case study of a help desk and its users. In *CSCW Proceeding on the ACM Conference on*

*Computer Supported Cooperative Work*, pages 41–48, Philadelphia, PA, USA, Dec. 2000. ACM.

- [288] Lex S. Van Velsen, Michaël F. Steehouder, and Menno D. T. De Jong. Evaluation of user support: Factors that affect user satisfaction with helpdesks and helplines. *IEEE Transactions on Professional Communication*, 50(3):219–231, 2007.
- [289] Christopher L. Carr, Patrick J. Bateman, and Saral J. Navlakha. They call for help, but don't always listen: The development of the user-help desk knowledge application model. In *Learning from the past & charting the future of the discipline. 14th Americas Conference on Information Systems, AMCIS*, page 387, Toronto, Ontario, Canada, Aug. 2008. Association for Information Systems.
- [290] Kevin F. White, Wayne G. Lutters, and Anita Komlodi. Towards virtualizing the helpdesk: assessing the relevance of knowledge across distance. In *Proceedings of the 2nd ACM Symposium on Computer Human Interaction for Management of Information Technology, CHIMIT*, page 3, San Diego, California, USA, Nov. 2008. ACM.
- [291] Christine Halverson, Thomas Erickson, and Mark S. Ackerman. Behind the help desk: evolution of a knowledge management system in a large organization. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, pages 304–313, Chicago, Illinois, USA, Nov. 2004. ACM.
- [292] Tracy Ann Sykes. Support structures and their impacts on employee outcomes: A longitudinal field study of an enterprise system implementation. *MIS Quarterly*, 39(2):437–495, 2015.
- [293] Sharoda A. Paul and Madhu C. Reddy. Understanding together: sensemaking in collaborative information seeking. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, pages 321–330, Savannah, Georgia, USA, Feb. 2010. ACM.
- [294] Paul P. Maglio, Eser Kandogan, and Eben Haber. Distributed cognition and joint activity in collaborative problem solving. In *Annual Meeting of the Cognitive Science Society*, volume 25, pages 758–763. ACM, 2003.

- [295] Eser Kandogan, Paul P. Maglio, Eben M. Haber, and John Bailey. *Taming information technology: Lessons from studies of system administrators*. Human Technology Interaction Series. Oxford University Press, 2012.
- [296] Edwin Hutchins. The social organization of distributed cognition. In *Perspectives on socially shared cognition*, pages 283–307. American Psychological Association, 1991.
- [297] J.S Busby. Error and distributed cognition in design. *Design Studies*, 22(3):233–254, 2001.
- [298] James D. Hollan, Edwin Hutchins, and David Kirsh. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, 7(2):174–196, 2000.
- [299] Mark Perry. Chapter 8 - distributed cognition. In *HCI Models, Theories, and Frameworks*, Interactive Technologies, pages 193–223. Morgan Kaufmann, San Francisco, 2003.
- [300] David Botta, Rodrigo Werlinger, André Gagné, Konstantin Beznosov, Lee Iversen, Sidney S. Fels, and Brian D. Fisher. Towards understanding IT security professionals and their tools. In *Proceedings of the 3rd Symposium on Usable Privacy and Security, SOUPS*, volume 229, pages 100–111, Pittsburgh, Pennsylvania, USA, 2007. ACM.
- [301] Rodrigo Werlinger, Kirstie Hawkey, and Konstantin Beznosov. An integrated view of human, organizational, and technological challenges of IT security management. *Information Management & Computer Security*, 17(1):4–19, 2009.
- [302] Joe Hertvik. Who uses ITIL in 2020? <https://www.bmc.com/blogs/who-uses-itil/>, Jun. 2020. Accessed June. 2019.
- [303] Karen Holtzblatt and Sandra Jones. Conducting and analyzing a contextual interview (excerpt). In Ronald M. Baecker, Jonathan Grudin, William A.S. Buxton, and Saul Greenberg, editors, *Readings in Human-Computer Interaction*, Interactive Technologies, pages 241–253. Morgan Kaufmann, San Francisco, CA, USA, 1995.

- [304] Sabina Kleitman, Marvin K. H. Law, and Judy Kay. It's the deceiver and the receiver: Individual differences in phishing susceptibility and false positives with item profiling. *Plos One*, 13(10):1–29, Oct. 2018.
- [305] Annie Saunders. Online solutions: looking to the future of knowledgebase management. In *Proceedings of the 32nd Annual ACM SIGUCCS Conference on User Services*, pages 194–197, Baltimore, MD, USA, 2004. ACM.
- [306] Rob Barrett, Paul P. Maglio, Eser Kandogan, and John H. Bailey. Usable autonomous computing systems: The system administrators' perspective. *Advanced Engineering Informatics*, 19(3):213–221, 2005.
- [307] Sonia Chiasson, PC Van Oorschot, and Robert Biddle. Even experts deserve usable security: Design guidelines for security management systems. In *SOUPS Workshop on Usable IT Security Management (USM)*, pages 1–4, Pittsburgh, PA, USA, 2007. CiteSeerX.
- [308] Rodrigo Werlinger, Kirstie Hawkey, and Konstantin Beznosov. Security practitioners in context: their activities and interactions. In *Extended Abstracts Proceedings of the Conference on Human Factors in Computing Systems, CHI*, pages 3789–3794, Florence, Italy, Apr. 2008. ACM.
- [309] Vihanga Heshan Perera, Amila Nuwan Senarathne, and Lakmal Rupasinghe. Intelligent SOC chatbot for security operation center. In *International Conference on Advancements in Computing (ICAC)*, pages 340–345, 2019.
- [310] Abhinav Krishna Kaiser. *Become ITIL Foundation Certified in 7 Days*, volume 1st edition. Apress, 2017.
- [311] Anne Adams and Martina Angela Sasse. Users are not the enemy. *Communications of the ACM*, 42(12):40–46, Dec. 1999.
- [312] Don Byrne. Six key mistakes leading to SOC burnout. <https://www.avanan.com/blog/six-key-mistakes-leading-soc-burnout>, Mar. 2021. Accessed April. 2021.
- [313] Mohammad Tahaei, Alisa Frik, and Kami Vaniea. Privacy champions in software teams: Understanding their motivations, strategies, and challenges. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*,



- CHI '21, pages 1—15, New York, NY, USA, 2021. Association for Computing Machinery.
- [314] Daniel Votipka, Desiree Abrokwa, and Michelle L. Mazurek. Building and validating a scale for secure software development self-efficacy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–20, New York, NY, USA, 2020. Association for Computing Machinery.
- [315] Andrew G Kotulic and Jan Guynes Clark. Why there aren't more information security research studies. *Information & Management*, 41(5):597–607, 2004.
- [316] Linfeng Li and Marko Helenius. Usability evaluation of anti-phishing toolbars. *Journal in Computer Virology*, 3(2):163–184, 2007.