

Designing an Autoresponder for Phishing Email Reports

Zeyu Zhang

Master of Science
School of Informatics
University of Edinburgh
2021

Abstract

Deceptions that intend to trick Internet users into providing private or sensitive data are called 'phishing', which is one of the most dangerous cyber-attack for causing financial losses and privacy leak on organizations and individuals. Email is a common vector of phishing, where attackers usually disguise as reliable sources and induce recipients to click on fraudulent links(also called URLs) and input their private information such as bank accounts or socialmedia passwords. Perfect automatic detection is the most effective and efficient method in preventing phishing. However, the common flaws of automatic filters have led to the necessity of supporting human identification of phishing. Email recipients themselves are indeed more suitable for making decisions than machines or other people, because they are clear about their own context.

Nowadays, it is common for organizations to support their members recognizing phishing by encouraging them to report any suspicious emails. However, as a kind of human assistance, not every inquirer can be responded in minutes, while speed can matter a lot because phishers often press their targets to make immediate decisions by injecting a sense of urgency. In addition, as the help desk workers might not be phishing experts either, their feedback is not always helpful that it might be very general but not customized to a certain reported email. Therefore, I designed an automatic responder for phishing email reports, which is technical feasible to implement to return immediate feedback to the inquirer. According to the design, the responder can automatically parse the reported emails and notify the findings through specific colors and texts, ranging from clean(safe) to dangerous. This solution combines the human autonomous decision and more reliable automated detection methods, which cover a wider range of phishing features compared to most user-support tools, including email headers, internal URLs, and email body. Also it focuses on explaining those features to ordinary users. Through user evaluation with 6 non-experts participants, the feedback is demonstrated to be generally comprehensible, helpful and friendly. Finally, the design was improved in terms of function, layout and user understanding based on participants' experience.

Acknowledgements

I would like to thank Kami Vanica firstly for her suggestions on my whole project as a supervisor in the previous several months. It was also kind of Nikolas Pilavakis who provided many useful comments on my progress report. In addition, the TULiPS group helped me a lot in writing the Introduction chapter. Finally, thanks to my six friends in the University of Edinburgh who supported me in the user study and the evaluation.

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.

(Zeyu Zhang)

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Research Goal and Objectives	3
1.3	Novelty	3
2	How Does a Phishing Email Look Like?	5
2.1	Phishing Behaviors and Features	5
2.2	Headers	6
2.3	URLs	8
2.4	Email Contents	10
3	Related Work	12
3.1	Human Education	12
3.2	User Support Tools	14
4	Design Process	18
4.1	User Study & Demand Analysis	18
4.2	Feature Selection	21
4.2.1	Header Feature	22
4.2.2	URL Feature	25
4.2.3	Body Feature	28
4.3	Initial Design of the Autoresponder	29
4.4	Evaluating the Initial Design	31
4.4.1	Feedback Content	31
4.4.2	UI Design	32
4.5	Final Design of the Autoresponder	33

5 Discussion	37
5.1 Contribution	37
5.2 Limitation	38
5.3 Conclusion	39
5.4 Feature Work	39
Bibliography	41

Chapter 1

Introduction

1.1 Motivation

Deceptions that intend to trick Internet users into providing private or sensitive data are called 'phishing'. Instead of a technical threat to cybersecurity, phishing has been considered as more of a 'semantic attack' with the purpose to use people's trust against them [18]. Phishing is harmful for causing serious financial losses on organizations and individuals [16], and it is worth noting that the danger of phishing is even increasing by year due to the development of scamming techniques. Proofpoint reported that 57% of their global respondents have suffered at least one successful phishing attack in 2020, while victims of data loss and credential compromise have a 10% increase compared to the previous year [50]. Email is a common vector of phishing, where attackers usually disguise as reliable sources and induce recipients to click on fraudulent links (also called URLs) and input their private information such as bank accounts or social media passwords[13]. Counted by Talos [57], in the first quarter of 2021, 80% of the emails worldwide are phishing emails, and this figure is essentially the same as in previous quarters, which demonstrates the prevalence of phishing emails.

The advancement of computer science has given birth to many automatic phishing email filters, such as some approaches based on machine learning algorithms [5, 23, 53]. However, Internet users sometimes still see phishing emails in their inboxes rather than the spam boxes because even best detectors are not yet perfect [20]. In this case users themselves need and are suitable to make final judgement because they have the exact contextual knowledge that is not available to the automatic filters or anyone else. For example, only they know their own bank details or maybe account expired date. To support users make safe decisions upon suspicious emails, traditionally they

would be trained by organizations to improve their ability to discriminate phishing, and there are also some computer based anti-phishing tools to help them recognize phishing websites. However, human training can be costly [3], and it is still difficult for ordinary users to recognize some phishing features even with training, such as highly disguised URL domains or redirects [14, 4]. Anti-phishing tools are more reliable in detecting phishing, but warnings from those tools could be too professional to non-experts, therefore frequently ignored [21].

Nowadays, it is common for organizations to encourage their members to report any suspicious emails [37]. This approach could not only enrich the organizations' phishing database, but also allow users to get intuitive advice from the IT help desk. However, as a kind of human assistance, not every inquirer can be responded in minutes, while speed can matter a lot because phishers often press their targets to make immediate decisions by creating a sense of urgency, such as warning serious consequences or informing time limitations [8, 63]. In addition, as the help desk workers might not be phishing experts either, their feedback is not always helpful that it might be very general but not customized to a certain reported email. Therefore, getting timely and customized advice could be very important for users to make security judgments upon suspicious emails. Another issue in user support field is that the existing anti-phishing advice are mostly rigid, which are likely to give a perfunctory impression to users, thus having low effectiveness. This is also considered in the presenting project.

In this study, a prototype of an automatic response is design for end users who report a suspicious phishing email, to give them suggestions while ensuring speed. As speed would not be a big challenge in an auto-responder, the primary problem that is solved is how to design a customized feedback which could effectively help people judge phishing. Such a responder should be based on machine's analysis of the reported email as some typical phishing behaviors can be efficiently identified by automated programs, such as the real sender address and some lexical tricks in links [65, 4]. So I have investigated which machine-oriented phishing features could be understood by ordinary users, as well as how to present them in reasonable layouts. According to my design, every feedback generated by the responder is specifically for the reported email (customized), including the analysis of email headers, links in the email and the email content. To evaluate the work, 6 semi-structured interviews with non-expert participants were conducted to test their attitude on the design and the comprehension on the security information. The results demonstrates the basic success of this study. In

the initial version of the prototype, users can generally understand the terms I provide without prior knowledge, and rather than being perfuncted, they could also feel cared and calm in reading the feedback. However, some design problems such as excessive colors, busy layout, and ineffective summary information were discovered. Finally, an improved prototype was made based on participant feedback.

1.2 Research Goal and Objectives

The primary goal of this project is to design an informative prototype of an auto-responder that will take in potential phishing from an email inbox and automatically generate a useful feedback for the user immediately, which should help people understand phishing features as well as make safe decisions upon suspicious emails in a timely manner. The whole research is based on a university context and aims to give users specific support according to the emails they report. In terms of the practice, I focus on the design process of the user interface rather than the implementation of the working system. My work includes the study of user groups, investigation of their demands, selection of suitable phishing features and make the feedback comprehensible and user-friendly, with more details such as wording choices and specific layouts. In this case, some functions are assumed to work, such as automatically parsing URL domains or detecting redirects, but the feasibility of the autoresponder can be validated by theories or existing works.

To realize the goal, the project will involve the following objectives:

- Select phishing features that are effective for human to make safe decisions.
- Give customized feedback based on the forwarded email.
- Present all the information so that even ordinary users could understand and feel happy to read.
- Support users to make decisions rather than make decisions for them.

1.3 Novelty

As it introduced in the previous sessions, user training, anti phishing tools and reporting to IT help desk are technically valid in supporting people to judge phishing, but

these approaches are not sufficiently comprehensive to cover all the issues of effectiveness, comprehensibility and timeliness. There is a recent study about designing an auto-checker for suspicious URLs, which combine these three issues by interpreting machine-facing features to human beings [3]. However, what they designed only allows URL analysis, but could not include other features that would also frequently appear in phishing emails, such as email headers and contents [65].

This solution draws on and combine the advantages of user support methods, such as the high comprehension of user education, the professionalism of anti-phishing tools, and reasonable usability and customization potential of email reporting. In addition, this research continues the existing comprehension work which aims to let non-experts understand professional phishing features, but this solution can be more comprehensive that, in addition to URLs, the analysis of email headers and contents are also included. Finally, this study is also committed to improving users' experience when reading safety recommendations, which reduce their stereotypes of such services through friendly language and information layout, so that further improve the effectiveness of this method in helping user judge phishing.

Chapter 2

How Does a Phishing Email Look Like?

2.1 Phishing Behaviors and Features

The common goal of most phishing attacks today is to let victims click on a URL in the email, to direct them into a website owned by the phisher and induce them to enter private information through it. Phishers use various types of disguise to make victims mistakenly believe that they are receiving a message from legitimate and related organizations, such as their banks, social media platforms, energy supporters or governments. Basically, an email can be divided into a header and a body. The header stores the tracking information of the conversation, such as the sender and recipient address, and the subject if there is one. The body shows the specific contents of the email. Nowadays, whether it is for humans or automatic filters, judging phishing emails is based on specific sets of phishing features [40]. A phishing feature can be understood as a variable in an email that phishing would show some common patterns in, such as the number of internal links or the domain age of the URL [35]. In other words, a phishing feature is a small element of an email that could indicate it is phishing or not. Phishing features can be established because phishing emails have some typical behaviors, which are not the same as most legitimate emails. Existing study shows that both email header and body contain valid phishing features, which means that malicious emails have special behaviors in both header and body [41]. While the URL in the email, as a part of the email body, has been found particularly numerous and complex behaviors because it is the bridge between the victim and the malicious website [58]. Therefore, the following chapters will introduce how phishing emails

would behave in terms of email headers, URLs, and contents. Note that the contents here are the text in the email body except the URLs.

It is worth knowing that there are indeed various phishing features found so far, but many of them are actually not common and are mentioned in very few works [4]. In addition, there are also a big number of features which are limited in human-centered phishing prevention because they are difficult to be understood by non-experts, or are confusing to explain, such as some DNS features or the page rank [4, 3]. Moreover, to identify phishing emails, a small set of common features could be sufficient to achieve reasonable outcome. Many automatic filtering algorithms can achieve an accuracy of over 90% with less than 10 widely-used features[22, 19, 40, 68], and some even reach more than 95% [40, 68]. If similar features can be understood by users, the effect may be even better with their own contexts. Therefore, this essay will only cover some typical phishing features that are considered to be potentially comprehensible to ordinary users. This chapter will focus on explaining how would attackers disguise their emails, and the specific feature selection for the security feedback will be introduced in subsequent chapters.

2.2 Headers

Nowadays, the world email communication is based on the Simple Mail Transfer Protocol (SMTP) [24]. When a user receives a new email, the information displayed at the top is usually the email subject, sender, recipient, and CC, which are read from the email header. In fact, in addition to these information, a complete email header also records many other contents, such as a unique Message-ID that acts as an identifier, and the security authentication result upon the email [53]. Checking the complete header could help to detect phishing emails. However, ordinary users may not do so usually, because the complete header is in the source file of the email which requires additional operations to open, and the information is too professional to understand without prior knowledge. Such situation leads to a vulnerability to phishing attacks.

no-reply@accounts.google.com

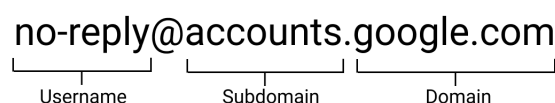
The diagram shows the email address 'no-reply@accounts.google.com' with three brackets underneath. The first bracket is under 'no-reply' and labeled 'Username'. The second bracket is under 'accounts' and labeled 'Subdomain'. The third bracket is under 'google.com' and labeled 'Domain'.

Figure 2.1: Example email address

Ordinary users usually judge the sender only by the address in the "From" tag of

the email, and this is indeed the most intuitive sender information. However, email addresses could easily be used to confuse users. As figure 2.1 shows, an email address contains a username, a domain name and optionally some subdomains. It should be clear that only the domain indicates the source of the email, and the domain name should match the sender declared in the email, while username and subdomain names can be anything set by the domain owner [12]. For example, the address of the customer service emails sent by Paypal is 'service@paypal.com' and the domain name is 'paypal.com'. However, users who are not familiar with domains may consider an address such as 'paypal@notice-access-273.com' as Paypal as well, while in fact it is from 'notice-access-273.com'. There may also be a confusing sender with the address 'service@paypal.team-123.com', but it is still not Paypal because 'paypal' here is a subdomain of 'team-123.com'. Placing the organization name in a location other than the domain is a common feature of phishing emails [68], but this may not be an illegible trick for users who have some knowledge about domains.

Some more advanced phishers may tamper with the domain in the sender address, which is also known as 'Email Spoofing' [49]. In this kind of deceptions phishers often set the sender to the address with a exact domain of a formal organization, such as 'Bank of Scotland ;info@emails.bankofscotland.co.uk;', to make users believe that they have actually received an email from that party. Email Spoofing is difficult to prevent with early versions of SMTP as the initial protocol did not add any address authentication measures, which made it very easy for phishers to steal email domains from legitimate organizations [64]. As a solution of this issue, several authentication-based security extensions have been launched on SMTP [24], such as Sender Policy Framework (SPF), Domain Key Identified Mail (DKIM) and the updated Domain-Based Message Authentication, Reporting, and Conformance (DMARC), which combines The first two methods. Organizations can deploy DMARC to protect their own domain names, so that every email sent by their domains will be verified to block misappropriation. However, although email authentication can effectively prevent email spoofing in theory, the number of protected domains is not yet reassuring. The 2019 Global DMARC Adoption Report shows that almost 80% of companies worldwide do not have any DMARC policies, which means that the email addresses of many organizations are still at risk of being used without permission [46]. In addition, the action to emails that do not pass the DMARC is set by the deployer, such as rejecting, accepting, or marking. However, not all organizations would completely reject such emails, which leads to the possibility that even phishing emails flagged by DMARC

may still appear in users' inboxes [32]. Therefore, it could be important to check the authentication result in the complete email header. It is also worth noting that the Message-ID in the header contains the real sender's domain name [54], which is added by the mail server, so is hard to be altered by attackers. For example, a phishing email could have a sender of 'Bank of Scotland ;info@emails.bankofscotland.co.uk;', which seems to be from the domain 'emails.bankofscotland.co.uk', but the Message-ID might be '2060000.1012684767@spa.spawn.se7en.org', which indicates that it is actually from 'wn.se7en.org'. Therefore, for a domain that is not protected by email authentication, checking whether it is the same as the one displayed in Message-ID could be a useful method to know if it is spoofed.

2.3 URLs

The essence of phishing attacks is to let victims interact with malicious information, and the URLs to websites owned by phishers are the key to achieve such interactions. Therefore, phishers always try their best to make their links look reliable and safe, in order to increase the victims' chance of clicking [11]. According to the figure 2.2, common components of URLs include protocols, domains (hostname), and path name, as well as some others like port numbers that are not usually displayed to users. The domains can be further divided into a top-level domain (TLD), a domain name, and potentially several subdomains. Similar to the email address, the domain name in a URL is also the part that indicates its attribution, and the domain name/TLD (domain.TLD) cannot be completely the same as any of the existing legal ones. However, although this restriction prevents phishers from using a website address that seems exactly like a legitimate organization as email spoofing, they still have various tricks to deceive users.

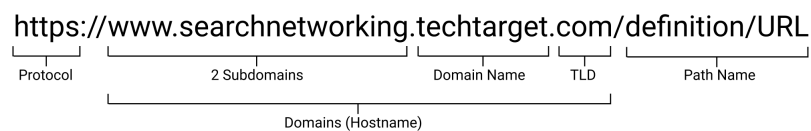


Figure 2.2: Example URL

First, since there are no limitations on the subdomains, path name and other fields in the URL, phishers often put confusing contents in those locations [50, 13]. This kind of tricks is similar to the ones in the email header mentioned in the previous sec-

tion. For example, 'https://paypal.bank-service.com' is actually a link to the 'paypal' subdomain under 'bank-service.com'. Studies have also concluded that some common TLDs like 'com' could be used as a subdomain or path field in phishing links to obtain users' attention [42], while in those cases an unpopular TLD would be at the end of the domains (such as 'www.team3.info/paypal.com'). In addition, in terms of the domain name itself, although a new domain with TLD cannot be completely the same as registered ones, it will actually pass the review with only a little difference. Therefore, phishers tend to register some domain names that are indeed different from legitimate organizations' but are very similar, and those differences are hard or even impossible to distinguish by human eyes. Some frequent example can be the substitution between v and y (pavpal.com vs. paypal.com), disguising m with rn (virginrmedia.com vs. virginmedia.com) and so on [10]. Moreover, some visually-similar characters in non-English languages (from Unicode) would be used instead of English letters (ASCII), such as Russian 'б' against English 'b', or French à against English a [26]. As most URLs are only composed of English or native languages of non-English speaking countries, mixed languages in a link can be a strong indicator of phishing [52].

It is also a common practice to construct a URL that looks trustworthy with special symbols. For example, many automatic filtering algorithms [26, 28] consider the '@' symbol as a phishing feature, because phishers often use it to 'separate' one field in the URL to let users regard it as two sections, such as 'www@paypal.com' or 'paypal.com@bot.org', which is very rare in trusted URLs. In addition, other symbols like '_', '-', and '=' may also appear in malicious URLs [15], but they may also be included in a few legitimate domain names, so these symbols could not be a strong independent signal for phishing judgment.

In addition to various methods to make fraudulent URLs look legitimate, many phishers use non-text hostnames to confuse users on their domains [45]. This is considered as a strong phishing evidence because reliable organizations usually display their registered domain with brand names to show their identities [29]. Possible forms in this kind of disguises involve IP address (such as 'http://45.135.186.81') or IP address encoded as hex code (such as 'http://34352e3133352e3138362e3831.com') [43], which aim to hide the real domain name. Additionally, there could be a hostname encoded as a specific Punycode (Punycode is Unicode represented by ASCII), which is probably to abuse the browsers' transcoding mechanism to make them display the URL as a Unicode string. For example, 'xn-80ak6aa92e.com' used to be displayed as 'apple.com' in the address bar of the Firefox browser [27], while every character in the

'apple' field is actually not in English (ASCII), but from other languages (Unicode).

Sometimes some visually safe URLs may also direct the victims to a fraudulent website because they are embedded with redirects. Redirection is considered as an advanced method in phishing [58], because specific security vulnerabilities such as Covert Redirect, can be exploited by attackers to build a URL of a real site address, but redirect to a pop-up window with a fake domain name, to obtain users' information. Redirection is not an exclusive phishing feature. Proper organizations sometimes also apply redirection to show some login or authentication windows. However, redirects of trusted groups are usually to a subdomain under their own domain [1], while destinations of malicious redirects are likely some fraudulent URLs that mostly different from the original address. Therefore, for a URL with redirects, it is necessary to check the final stop instead of the original one.

2.4 Email Contents

In addition to specific behaviors in email headers and URLs in the body, phishing emails also have certain predictable characteristics in their textual content. First of all, most phishing emails address the recipients as a general title rather than specific names [25], such as "Dear Paypal User" or "XX account owner", because attackers probably do not have their detail and the general title allows phishing messages to be sent to thousands of people at the same time while maintaining a low cost. However, legitimate organizations frequently use the user's real name as the title in some important or formal emails [63], because they usually keep their customer information carefully (see Figure 2.3). Although with the development of phishing methods, more accurate attack modes such as spear phishing with real user information have emerged [1], but these methods require advance investigation of important targets, which is considered difficult for general phishing attacks.

As for the email message itself, it is very common to have mandatory requirements or inject some urgency to pressure users to react [28]. Usually a specific illustrative email will not be a scam, as the recipient cannot interact with the information. Therefore, topics such as finance, accounts, advertising which are related to money and interests have become popular subjects in phishing emails [42]. Fraud message in finance or user account may contain some intimation or indications of serious consequences, such as 'please log in/pay within 1 working day, otherwise you will be responsible for the consequences'. For false advertisements, it should be emphasized the rare benefit

and time limitation to take advantage of the avarice of human nature, such as 'the discount is only once, and it is going in two hours'. Generally, important notices from legitimate organizations would be issued for the first time before the deadline with a space for several days [63], and the languages should be objective without any warnings, such as 'Dear Jack, please update your account before August 20'.

Some scam emails may be also found to have poor grammar or even spelling errors [53]. However, as the increase of users' security awareness, such attacks are considered to be more ineffective. And these issues have been improved in many modern phishing emails [5].



Figure 2.3: Example recipient title from legal emails

Chapter 3

Related Work

Since this study is about human-centered phishing prevention, and the background of phishing emails have been mentioned in the previous chapter, this chapter will introduce some existing solutions in this field. For human education and user support tools as two categories, the development status will be explained and some potential gaps will be proposed. The subsequent design process is principally based on the analysis of existing work.

3.1 Human Education

Anti-phishing training is traditionally regarded as an important approach of combating phishing attacks, and it is also a method widely used in organizations around the world [14, 4]. Users are aware of their own situations and contexts, such as their banks, energy suppliers, account information, or recently purchased goods, which make them most qualified as the entity in the final decision on phishing emails that are not automatically identified. Psychologists have shown that when facing stress, humans tend to solve problems according to their familiar patterns [56], which in a sense establishes the importance of user education, that is, to help them become familiar with methods of detecting phishing. Experts are less susceptible to phishing attacks than ordinary users [7], because they understand some phishing behaviors and attack modules, so they could possibly make a safe decision by checking the sender, URL domain name, or with the help of some professional tools, which is also considered to be the outcome of human education.

Generally speaking, based on the discovered phishing features, training for judging phishing emails includes recognition of malicious link and phishing clue based on

the email header and body, and search engine normality checks [45]. Different types of training have different characteristics and levels of effectiveness. Specific training modes include: Upfront training such as independent training based on network materials, offline training with human trainers, and embedded training, including interactive programs/games, or repetitive training in simulated phishing contexts [6, 34].

Upfront training sessions are usually conducted before users are attacked by potential phishers. During the training, some security concepts should be explained, such as common phishing attack models, phishing features, and judgment strategies based on behaviors and features, etc. Some organizations hire human experts to carry out security courses for their members, which is similar to university routines [38]. A lower-cost method is to construct materials of security advice and distribute them to users through web pages or other electronic media for self-learning. This is also the method that most organizations are practicing. However, regardless of the method, as upfront education is a preventive training, its effectiveness greatly depends on users' learning motivation and ability. Research [31] reveals that a large number of users would directly ignore this kind of anti-phishing training due to subjective factors such as boredom, while users who study hard are still easy to forget many concepts such as some URL domain features. In addition, the continuous development of phishing attacks will gradually invalidate part of the training, which further increases the challenge of upfront training [30].

Educational researchers believe that human training will be more effective if the training materials can be combined with the background of the real world, work or laboratory environments [31]. Embedded training in human-centered anti-phishing aims to help users associate themselves with educational information, making it a part of routine activities, and consolidate the information by repeating. Graphical, short, simple, and diversified training materials combined with daily life stories have good potential to capture users' attention and to keep updated. Embedded training usually requires simulation of a real phishing context to strengthen the users' memory. For example [35], sending simulated phishing emails to users to urge them to enter a 'malicious' website, log in and provide personal information. The trainer intervenes and presents the training materials when the user clicks on the link. In addition, interactive games or programs are often implemented in embedded training. For example, the PhishGuru [36] training system automates the training process described above, while Anti-phishing Phil [55], an online game, train users by explaining how to identify phishing websites, find clues in the web browser, as well as how to use search engines

to compare the suspicious websites with legitimate ones. Embedded training is considered to have the following advantages [35]: (1) It enables system administrators or training companies to conduct continuously training when new phishing methods appear; (2) It saves users' time to attend formal training sessions (as embedded training is part of the primary task); (3) It creates a stronger motivation for users, because the training materials will only be presented after they actually become victims. However, even if it might be more advantageous than upfront training, the effectiveness of embedding training do not reach a reasonable level yet. About one-fifth of the participants entered private information in the fraudulent website with the training by PhishGuru, and 31% of Anti-phishing Phil users still can not judge the security of strange websites [31]. This issue might be caused by insufficient customization of the training materials, because some users may find it difficult to apply general training materials to specific phishing emails. In addition, the frequent update and repetition of embedded training indicates that it is a costly method [35], whether the organization is hiring a training team or developing a training system.

In summary, training, as a traditional human-centered method of preventing phishing, focuses on cultivating users' ability to identify phishing independently by making them understand phishing-related knowledge. However, its effectiveness is limited due to various factors such as insufficient learning motivation/ability, and low level of customization. In addition, the high cost makes reasonable human education a very challenging task.

3.2 User Support Tools

User support (or anti-phishing) tools refer to computer programs that can support users to make phishing judgments. The core concept of such supports is the combination of human judgment and automatic detection technology. Many existing works [69] recommend to involve automatic features in supporting users to make decisions because some important phishing tricks cannot be recognized by humans [4], such as the multi-language problem introduced previously. Theoretically, anti-phishing tools could be more reliable than user education, as they have ability to provide precise and professional information in a short period of time by combining different categories of phishing features, such as blacklists (known malicious websites), whitelists (known safe websites), rule-based ones and community ratings [69]. The existing anti-phishing tools are mostly designed for URLs or websites analysis, which include browser plug-


ins, local software, and online websites, etc [67].

Browser plug-ins are mostly implemented as an extension of the browser. This kind of tools give warnings when users visit or are intend to visit risky websites. For example, Netcraft Anti-Phishing Toolbar, provides protection against phishing and malicious JavaScript, which communicates with the reported database of the Netcraft website to obtain and display blacklist results [69]. When entering a website on the blacklist, a pop-up window will advise the user to cancel the visit. Additionally, around the address bar it can also show some extra information of risk ratings, domain registration date and location, ranking and host information. Including these sorts of information is considered to be helpful by existing study [69]. Take domain location as an example, a company based in China should not be hosted in areas such as Africa or South America. However, page rank might be a confusing feature to users, as it is based on the popularity rather than any security rules [61].

Some browser plug-ins include specific review mechanisms based on local databases, which are actually similar to the principles of local tools. Google Safe Browsing [9] is also a blacklist-based browser security plug-in, which provides two options for detection. Users can choose 'Downloaded Suspicious Site List' or 'Ask Google for every site I visit'. With the first detection mode, the browser will download or update the local blacklist every time before opening a new window. Whenever a user visits a page, Google Safe Browsing compares its URL with the local blacklist. The second mode will let each visited address to be forwarded to Google's server, and then return the analysis result. When a page is considered to be deceptive, the toolbar will interrupt the user's current activity. A design worth learning from is that this tool makes appropriate recommendations rather than only giving warnings, such as stopping visiting the website, or reporting false detection results.

The online detection website is a newer support method, which are usually developed to identify malicious URLs. One advantage of this type of tool is that users are free to choose the features by viewing the analysis reports of different websites without any installation. However, compared to browser plug-ins and local software, support websites require users to actively search for URLs with a higher level of motivation. Different websites are capable for different features. For example, WhereGoes [62] is developed specifically for checking redirection. By querying a URL, the user will be informed whether it is embedded with redirects, the number of redirects and the landing address. EmailVeritas focuses on identifying whether the IP address has been tampered with and providing some WHOIS information indicating the URL attribu-

tion. It's worth mentioning that this tool gives a clear judgment on the URL, such as 'We did not find this link to be malicious', which is not considered a good design because for those URLs that are not automatically recognized, users are appropriate for the final judgement, which has been discussed above. In contrast, Urlvoid [59] provides a more comprehensive report, and does not include a decision. As shown in the figure 3.1, the user can see the domain information of the URL, including the registration date, specific location, IP address, and the date when it was last detected. In addition, it will also verify the link with 44 blacklists, including some popular ones such as PhishTank and Quttera.

Report Summary	
Website Address	Leboncoinpaiementpro.fr
Last Analysis	26 days ago Rescan
Blacklist Status	2/44
Domain Registration	2021-07-24 29 days ago
Domain Information	WHOIS Lookup DNS Records Ping
IP Address	77.81.120.31 Find Websites IPVoid Whois
Reverse DNS	Unknown
ASN	AS200514 KnownSRV Ltd.
Server Location	 (NL) Netherlands
Latitude\Longitude	52.3564 / 4.8802 Google Map
City	Amsterdam
Region	North Holland

Blacklist Report	
-	
-	
-	

Figure 3.1: An example report from Urlvoid

Anti-phishing tools are reasonable in concept, which support the users by providing extra information efficiently and leave the judgement to users themselves. Additionally they are technically valid as computer programs can accurately analyze suspicious web pages and links in multiple dimensions [54], so that the advice can be more reliable and customized than that in human education. However, the biggest problem of these tools lies in their excessive professionalism. Previous works have argued that human support tools are commonly limited by demands for prior knowledge, such as the meaning of the URL fields or the meaning of encryption [69, 3]. Most anti-phishing tools focus on the technology to parse links/webpages, but ignore the comprehensibility of technical

terms. For example, none of the tools mentioned above tries to explain the features they cover, which makes the supporting information not helpful for many ordinary users because they cannot understand it.

With the goal of improving users' comprehensibility of security concepts in human-centered phishing prevention, Mossano et al. compared several anti-phishing recommendation webpages of different organizations around the world from banks to universities [44], who pointed out an important obstacle for users to understand security advice by lack of examples, such as a specific URL with its components marked, which could be a guideline for this project. Another study focused on designing a human-friendly URL report [3]. The researchers tried to interpret some automatic URL features to users through the design language of traffic lights with reasonable layouts and explanatory languages, who obtained a good outcome as their evaluation has shown.

Drawing on the ideas of existing work, this solution combines some phishing features that can be automatically parsed and the users' final judgment. In the design, make the information provided easy for users to understand will be an important interest. In addition, this project also adds features related to the header and body of the email, to expand the information of existing user support tools, most of which are only for the URL or website analysis, that is considered as another limitation [44, 3]

Chapter 4

Design Process

At this stage, a user interface of the automatic feedback for potential phishing report has been produced, aiming to support users to make safe decisions when facing suspicious emails in a timely manner. Phishing features are the key to detect malicious emails, thus during the study I investigate and select several phishing features based on existing works, covering the email header, email body and the URLs in the body, to make sure they are reasonable and sufficient to indicate phishing emails. Additionally, comprehension and usability are considered carefully, such as research on user-friendly layout, languages, and color palettes. This chapter is about the design process of this UI prototype, which is based on the idea of product design, including user study, user demand analysis, feature selection and layout design. 6 participants from University of Edinburgh were invited into the user study and also an evaluation after the initial interface was built, covering design and user comprehensibility aspects. Finally, the a improved version is accomplished based on their opinions. Since the entire research is carried out in the context of the UoE, currently all features and usage scenarios are considered based on the situations inside the University.

4.1 User Study & Demand Analysis

The purpose and behavior of phishing emails are already investigated through theoretical review. However, as limited work are actually found on such an autoresponder, the users' views, expectations, or emotions when using this service are not clear yet, although some situations can be predicted, for example, ordinary users may need quick responses and understandable feedback. In this project, the significance of user study is to understand the potential user groups, so as to make the design more in line with

their actual needs for anti-phishing suggestions and create a high-level user experience.

User study is basically done by interviews with people inside the UoE and some existing research about phishing reports. Six students were invited into semi-structure interviews [51], which could obtain rich information without digressing. Participants are all non-experts. Except for a CS student who has some basic knowledge of URL, they are from design, humanities, and education departments, who are not familiar with cybersecurity. The interviews are expected to collect different users' understanding and attitudes towards phishing emails, as well as their thoughts and needs when reporting emails. Several key questions are generated as follows, with potential follow-up questions during conversations.

- Do you know what phishing is? Do you know what is Email Spoofing/Email Header/URL Domain...?
- (Show an example phishing email) Do you think this email is safe? How would you judge it?
- Do you report suspicious emails? Why?
- If you report an email, what could be the reasons? What kind of feedback do you want to receive in terms of function and language?

Interviews indicate that participants generally do not have good understanding of phishing. Although 2 of them can roughly describe the purpose of phishing email and the concept of URL, no one has a deep knowledge of terms such as email spoofing, header, or domain. In addition, when judging a suspicious email, they commonly do not have a systematic thinking and method, who can only make some guesses based on intuition. However, five participants thought of comparing the email content with their own situation, which is considered to be a discovery that can promote the success of this work. In terms of phishing reports [67], none of participants often reports suspicious emails, which matches the findings of some existing work. Overall, all participants believed that even if they report an email, existing services would only give perfunctory responses, which would not help identify phishing. 3 participants confidently believe that they will not be attacked even without any help, and 2 are worried that their misreporting will affect the work of IT staff, and even be ridiculed. In the end, participants generally reckoned that if an email is reported, they hope to receive a binary result (phishing or not) directly. After explaining that they are the

most suitable one for decision, they set out some other expectations, which will be introduced below.

User interviews and background research led to 3 personas for needs analysis, describing different users. In addition to academic employees and students who are considered to be the primary users, Althobaiti et al. also stated that such a machine assistance may also be benefit to non-professional IT help desk employees, as they can help the inquirer more quickly and effectively with some automatic analysis [3]. Therefore, personas include a student, a tutor, and a help desk worker. It is worth explaining that although few people actually report suspicious emails, this research focuses on the situations when someone have already submitted reports. The following analysis of users is also based on this.

A user persona shows a type of user's background, attitude towards phishing, and core expectations for email report responses. Experts would not be often in need of security help because they know how to judge phishing or use some professional tools. Therefore, as shown in the figure 4.1, non-expert users are divided into two groups according to their background and personality. Users represented by Kristin have little knowledge of phishing or technical terms. They may be not confident, cautious and worried in preventing phishing as they are more likely to have been victims than others. From interview participants, such users may need to learn how to identify phishing through information that is easy to understand. In addition, they may also need to reduce the pressure from the attacker and increase confidence in reporting and judging malicious emails. Additionally, a participant believed that it is still necessary for an access to human service, which is also considered an effective demand for users who cannot understand automatic feedback. Persona of Thomas represents users who have some knowledge of related concepts but are not professional. They may be more confident in judging strange emails because they have the ability to detect some phishing attacks. The motivation for such users to report emails may be to help strengthen the University's automatic filters, or they actually receive ones that are challenging to judge. In the interview, this kind of participants expected to be appreciated for their submitting, and some more professional information for those highly disguised emails.

The third persona describes help desk workers like William, whose important part of work is feedback on phishing reports. Although they are not professional either, they may have more resources in anti-phishing than end users, such as some tools or easier access to phishing experts. These people may care about the speed of the feedback, because the anxious inquirers need very timely help, but it can be time-costly

to manually analyze an email and write some responses. In addition, they may also need such a support to automatically generate customized feedback, so as to provide better help to those who might still need human assistance.

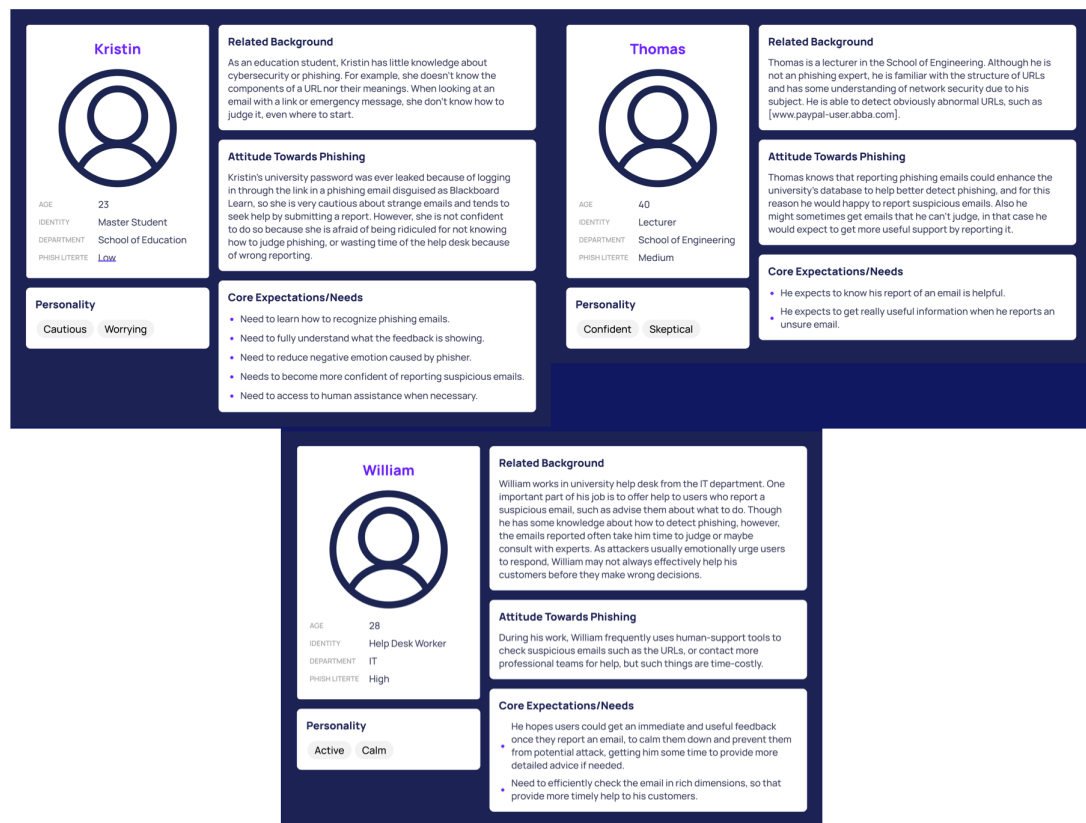


Figure 4.1: User Personas

4.2 Feature Selection

Analyzing the reported email is the basic of detecting phishing [16], which in realistic context can be reliably achieved by automatic functions. But computer programs need to be told a rule in advance to conduct the analysis, which is a set of phishing features. And this work can only be done by designers or researchers manually. Thus, in this section the detailed selection of the phishing features to be used in the security feedback will be introduce. All the features will be listed and explained the consideration for including them. It is already introduced that the features are divided into three categories, which will not be repeated here. Each of these features is included to detect specific phishing behaviors mentioned above, and most of them have been used in existing solutions, including automatic filters and human supporting works. In addition

to effectively indicating phishing behaviors, another criteria for selecting features is the potential to be explained to ordinary users, thus some ones that are very technical or likely to be confused are excluded [3], such as the domain popularity. As the study running, some features were removed or added for reasons, and all the thoughts or choices introduced here are for the final design. Some differences between the initial and the final versions will be explained in the following sections.

In the design, the different detection results of each feature are predicted and grouped into four danger levels, to provide intuitive warnings. Those levels are: 1. Facts: Objective facts in the email. As descriptive information, it does not indicate safety or danger directly. 2. Clean: This feature is not found signs of phishing, and most legitimate emails have the same behaviors. 3. Possible Danger: No clear evidence of phishing, but the behavior is considered to appear in many phishing emails, or only some legitimate ones. 4. Dangerous: This behavior appears in the vast majority of phishing emails or a very small number of legitimate emails. The idea of labeling information with different levels is feasible in the study of Althobati et al [3]. It is worth mentioning that they use 'No Issue', 'Possible Issue', and 'Known Issue' as indicators, while I consider using phrases related to 'Danger' can better alert users. Another point needs to be clear is that although the concept of the present solution is to let the user make the final decision, some particular features involved in the feedback may independently make the judgement, because they can basically indicate that the email is harmful if an abnormal behavior is detected. I refer these features as 'Key Features' which will be specially marked below.

4.2.1 Header Feature

Phishers are likely to play a lot of tricks on the email header, because this could convince the victims that they have received messages from legitimate organizations, which is very important for a successful scam. The final feedback prototype includes four different features for this part of the email, with their names and the criteria for judging the level of risk shown in table 4.1. In fact, for domain names protected by DMARC authentication, only three features will be actually displayed, which will be described below. It can be noticed that the feature of Domain Type has been removed from the original design, because based on supervisor's suggestions and some literature [48], this information is considered to be technically difficult to parse, which is the reason for its removal. For example, a large database of the domain names may be re-

quired for verification. But, it is undeniable that this feature could be a helpful security term if it can be realized, because it shows whether the email sender is an individual or an organization, so that facilitate the comparison to the context. For example, if the email says it is from Paypal, but actually the sender is found to be an individual, then it is probably a fake one.

For the email header, the first feature displayed should be the sender's address, which aims to introduce users the source of the email to help them make a part of the judgement based on the context. As the only indicator of the email origin, the domain will be identified for emphasis. To avoid misleading users, this feature is designed as an objective fact, because there will be no automatic security analysis on the domain name itself. If properly explained, the contextual check based on domain name should help detect most of the visible disguise in the email header covered above, such as confusing users by putting a legal organization name in the username or subdomain field [60]. In fact, it is indeed possible to interpret the domain clearly to common people. For any reported email, users of this feedback will be able to see a sentence as: 'Only the domain name indicates where the email from, which should match its declared sender', which not only clarifies the meaning of the domain, but also guides users to actively conduct contextual comparisons.

Among the retained features, authentication and domain check are regarded as two Key Features. According to the previous background, email authentication is one of the most effective means to detect email spoofing, which is an invisible trick for most users. If the domain of an organization is protected by a certain authentication strategy, then the mail sessions that fail to pass the authentication can indicate an illegal use of the domain name. Although currently the mainstream DMARC method allows organizations to automatically reject emails with failed authentication, the survey shows that not all organizations apply this setting, which means that there may still be spoofed emails appearing in users' mailboxes. In terms of user comprehension, the results of authentication could be easy to explain to non-experts. For example, some simple sentences such as 'the sender address is stolen from a legal organization' or 'the sender address is not stolen' can already be an intuitive and reasonable demonstration of this feature. Therefore, according to this design, the DMARC result in the email header should be checked for this feature. If it fails the authentication, the email is deemed dangerous, and it needs to be clearly indicated in the feedback that this is a confirmed malicious one. The result will be judged as clean if the email passes the DMARC authentication, as it means that the domain is at least legal in use. In addition, considering

that the number of organizations with DMARC protection is still very limited, email domains that are not found a DMARC deployment will be classified as possible danger, because it is uncertain whether they have been spoofed. In this case, the feature of Sender Type will be relied on to detect email spoofing. Although there are cases where the organization only deploys SPF or DKIM without DMARC [47], in order to avoid confusing ordinary users by many technical terms, it is decided to only check for DMARC for this feature.

The definition of domain check here is to test whether the sender domain name displayed to the recipient is the same as that in the Message-ID. In terms of the functionality, it can also detect email spoofing as what the authentication feature can do. As the Message-ID in the email header is difficult to be tampered with, the domain in it basically identifies the true source of the email, which means that the sender address with a different domain name has a high probability of being spoofed. This is why domain check is also regarded as a Key Feature. Although its purpose is similar to authentication, domain name detection, however, is not a redundant feature, mainly because many organizations have not yet deployed DMARC strategies. In contrast, as authentication may be more user-friendly, it is not considered to be replaced by domain check. Therefore, this feature finally exists as an additional information for users to recognize email spoofing in this design. It is still the preferred strategy to check the authentication results for domain names protected by DMARC, and this feature will therefore be hidden to keep the interface simple. Only when DMARC is not detected, will domain check be used to judge email spoofing. In addition, although it is not as easy to explain to the user as the authentication result, most ordinary people should be able to understand the prompt as 'this email address has been tampered with by its sender'.

The last feature for the email header is called 'email is from', which is actually checking whether the reported email is from outside the University. This is also a built-in function of the University of Edinburgh's email system. The back-end server will tag every email that has not been certified by the University with 'This email was sent to you by someone outside the University'. This feature is not difficult for users to understand as it is not a technical term, but it is in fact slightly tricky in judging suspicious emails. Although UoE warns users not to easily click on the links in the emails outside the University, but an outside email is not necessarily dangerous, because a lot of messages from legal platforms will also be flagged, such as Miro, an online collaboration tool commonly used by Design Informatics students. Therefore,

it was decided to only use this feature as a secondary reference, and not to show extra analysis to the users. Of course, it will definitely helpful if a user receives an email from an outside attacker who masquerades as someone inside.

Feature	Fact	Clean	Possible Danger	Dangerous
Sender Domain	Domain	-	-	-
Authentication (K)	-	DMARC Passed	Not Found	DMARC Failed
Domain Check (K)	-	Same as M-ID	-	Different from M-ID
Sender Type(R)	-	Organization	Individual	-
Sender is From	-	Inside the Uni.	Outside the Uni.	-

Table 4.1: Phishing features for email header. (R):removed from the design. (A)added in the design. (K):key feature.

4.2.2 URL Feature

For the links in the reported email, 17 features have been added to assist the user (table 4.2). Each detected URL will be analyzed and shown to the customer. The included features are basically demonstrated to be effective and human-friendly by the work of Althobati et al. [3], which is the only solution that focuses on designing user-comprehensible URL reports for non-experts. I divided the features into attributes and tricks. Five attribute features will be presented in every feedback to display some basic information of the link, which can support users to conduct contextual analysis. The features about trick aim to detect other disguise in the URL that may not be noticed by humans through automatic recognition. In order to maintain a clean interface and reduce the technical terms presented, tricks will only be displayed when they are found, otherwise the user will see 'No other strange things have been found in this link'. For most URL features, boundary setting of risk levels by Althobati et al. is followed, which they have verified.

Among the basic attributes of the URL, the first one to be considered is the address redirection. As mentioned above, a phisher may display a visually legitimate link, but direct the user to another malicious website after clicking on it. Therefore, for links that are embedded redirects, the final destination should be checked rather than the one that is initially displayed with no significance. However, since redirects would also be applied by some legitimate organizations, the detection of redirects will not be considered dangerous directly. For links that only redirect once, the landing URLs

will be displayed as facts. However, more than two redirects in a chain will gradually be considered abnormal according to the number, because legitimate organizations usually only embed once redirect when operations such as login are required, while phishers may rely on multiple redirects to escape the security scan of the browsers [1]. It is worth considering that some users may not understand the term of redirection, so it could be important to explain this feature in simpler languages.

	Feature	Fact	Clean	Possible Danger	Dangerous
Attribute	Redirection	No/Landing URL	-	2-4	>4
	Domain	Domain	-	-	-
	Domain Age	Age \geq 6 Month	-	3-6 M	<3 M/Unknown
	Location	City/Country	-	-	Unknown
	Search Result (A)	-	Match	Partial Match	No Match
Trick	44 Blacklist (K)	-	-	-	Any Detected
	IP Address (K)	-	-	-	Detected
	No. of Subdomains	-	-	2-4	>4
	Has '@'	-	-	2-4	Detected
	Hex Code in Host	-	-	2-4	Detected
	Non-ASCII	-	-	Unicode	Mixed-language
	Out-of-position TLD	-	-	-	Detected
	Out-of-position Protocol	-	-	-	Detected
	Out-of-position 'www'	-	-	-	Detected
	Top Targeted in subdomain	-	-	-	Detected
	Similarity to Top Targeted	-	-	Any Detected	-
	Similarity to Alexa Top 10k	-	-	Any Detected	-

Table 4.2: Phishing features for URL. (R):removed from the design. (A)added in the design. (K):key feature.

After checking the redirection properly, the domain of the URL will also be emphasized, similar to the sender feature in the email header. Since the domain name in the link is also the only field that indicates its origin, this information is considered to be an important help for users who do not have URL knowledge. In addition, domain age and its registered address will also be automatically detected as two additional support. Knowing domain age can be useful. Since a phishing message is usually sent to thousands of users in parallel [45], malicious URLs in it are often reported and blocked. Therefore, domain with very short ages are likely to be held by phishers, because they may need to register new one frequently. On the other hand, domain names of legal organizations are usually older as they can be used steadily [8]. In particular, most famous companies may have registered their domain names several years ago. In terms of registered location of the domain, although it is less significant for automatic

security analysis, users can use it to make useful autonomous judgments [39]. For example, China Post is an obvious Chinese company, so an email that declares China Post should not contain a domain registered in Japan.

The last feature of the link attribute is called search result. According to the design, the domain name under investigation needs to be searched in Google and check the URL's matching degree with the addresses of the first 5 search results (for example, search for world for paypal.world.com). Technically, legitimate links should be matched reasonably, because modern browsers rank search results based on searched terms and page popularity [39]. And since most harmful domain names are meaningless or unpopular [33], they would be rare to appear in the top results. In addition, this feature could have good potential to be understood by unprofessional people, because searching for unknown information in the browser conforms to the mental model of most Internet users [19].

In many cases, only depending on the above features may not be sufficient to make a secure judgment on unfamiliar URLs. The remaining 12 trick features are designed to reveal potential traps other than the contextual ones, which will not be displayed as clean to avoid misleading users. 44 blacklist was set as one of the Key Features in the final prototype, which indicates that every reported URL should be searched on 44 authoritative phishing blacklists, which are inspired from the Urlvoid website introduced above. A secure URL must not appear in any blacklist. On the contrary, an email with a blacklisted URL should be determined to be dangerous because it is a certified phishing [61]. Features except blacklist exist to check specific lexical behaviors, which are considered suitable in human-centered phishing behaviors. For example, the address '116.2.0.10.com' will be considered dangerous by both IP address and number of subdomains. In this case the user will see two warnings, 'This URL is abnormally presented as an IP address' and 'This URL has too many subdomains'. As explained earlier, addresses with '' or encoded as hex code are also risky, which should be marked in the feedback. The Non-ASCII feature will label links containing multiple languages as 'dangerous' to defend against attackers who replace some English characters with similar ones. Single-language but non-English addresses are considered possible dangerous because only a small number of legal organizations use non-English as their URL language (especially in the UK) [2]. In addition, three out-of-position features are used to identify links that place common fields in strange positions, such as 'www', 'https' or 'com'. Finally, for other spelling tricks, search and similarity detection will be performed through the two lists of PhishTank and Alexa. Links similar to phishing

URLs in PhishTank's Top Targeted list or Alexa's top 10,000 popular addresses may be malicious [3].

Compared with the work of Althobati et al. [3], this solution removes some of the URL features they apply, which mainly based on user comprehension and feature redundancy. For example, in this feedback Domain Popularity and Page Rank will not be displayed. Although either of these two features can be helpful, they are essentially considered redundant information of the search result, which actually summarizes the popularity and page rank. In addition, from the perspective of reading experience, ordinary users may be unfamiliar with the concept of popularity or rank, which is likely to incur additional learning costs. Therefore, the search result is retained for being better comprehensible.

Feature	Fact	Clean	Possible Danger	Dangerous
Recipient Title	-	-	Not Name	-
Phishing Keyword	-	-	Detected	-

Table 4.3: Phishing features for email body. (R):removed from the design. (A)added in the design. (K):key feature.

Phishing Keywords		
ACCOUNT	ACCESS	BANK
CREDIT	CLICK	IDENTITY
INCONVENIENCE	INFORMATION	LIMITED
LOG	MINUTES	PASSWORD
RECENTLY	RISK	SOCIAL
SECURITY	SERVICE	SUSPENDED

Table 4.4: 18 high-frequency words in phishing emails

4.2.3 Body Feature

Based on the current research status, the body of the phishing email is primarily to create some stressful context through specific languages, without technical disguise. Therefore, for this part, only two features are considered as warnings (table 4.3), neither of which is risky in confusing users. Firstly, according to the above introduction,

most phishing emails and some legitimate emails would refer users as general titles due to cost and other factors. Therefore, emails that are not found the name of the inquirer in the body will be classified as possible danger. In addition, some high-frequency phishing attack vocabulary will also be retrieved in the text through a keyword list (table 4.4), and emails with listed words will also be considered potentially dangerous. This feature with the keyword list has been applied in broad study on automatic phishing detection [17].

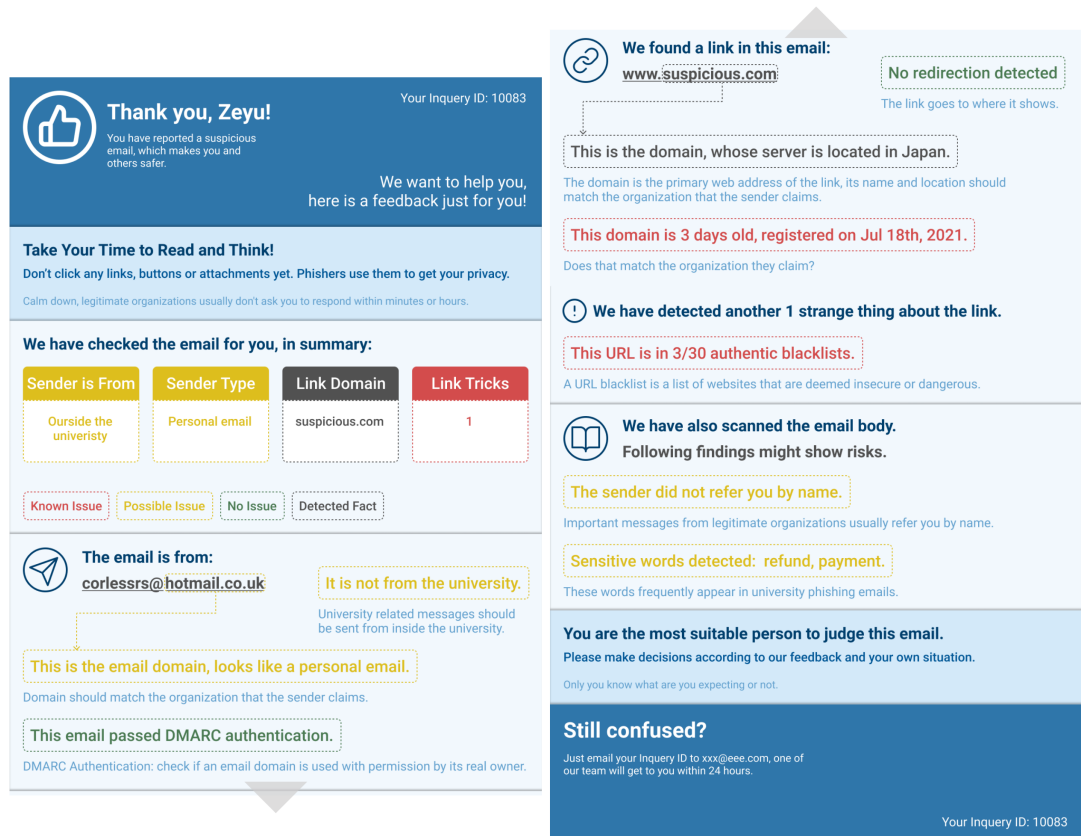


Figure 4.2: Initial design of the interface (the right part follows the left part).

4.3 Initial Design of the Autoresponder

Figure 4.2 presents the initial design of the feedback UI. In this version, the Key Feature that can independently identify phishing is not yet set, where all the features are displayed as auxiliary information. Eleven different colors are selected in the feedback (figure 4.3), among which the four levels of feature are represented by gray, green, yellow, and red. Facts are represented by dark gray, which is considered to have no emotional preference. The colors of traffic lights correspond to the urgency in most

people's mental models, so they are widely used to indicate different risks in security design [3, 26]. In addition, the blue color is set for the background and different levels of text, as it is considered to have a cooling and astringent effect to manage stress [66].



Figure 4.3: Color platters in the initial design.

In terms of the layout, the top and bottom of the feedback are some non-customized greetings, introductions, guidance, etc. For example, in order to prevent users from making anxious decisions and triggering some suspicions of the reported email, they are reminded of the common behavior of legitimate organizations. In addition, it explains why it is appropriate for users to make own decisions, to enhance their confidence in self-judging emails. Considering that some users may still be unable to make judgments based on automated feedback, the last module provides a way to obtain human services.

The customized email report based on the phishing features is placed in the middle. A summary module that provides four pieces of information is at the top, which in turn shows where the email is sent from, the type of sender, the link domain, and the number of tricks detected in the link. Such design is to give users a comprehensive view of the reported email. The following three areas, header, link, and body, present features of different categories, with classic icons to help identifying and locating. In order to ensure user comprehension, technical terms are avoided, such as URL replaced by link. And each feature information is described in one sentence, with a explanation below in blue, instead of only providing feature names and results like Urlvoid. In the feature interpretation, some security strategies are included as much as possible, such as guiding users to compare the domain name and registered location with the

sender shown in the email. Except for 'sender is from' and redirection, most of the information is arranged to the left because it is aimed to emphasize these two features. In addition, sender domain and URL domain are also emphasized by the dotted frame and arrow.

4.4 Evaluating the Initial Design

It can be very possible that the initial prototype is not satisfactory to users because the user study were conducted in the early stage, and their opinions were not continuously obtained during the design. Therefore, the 6 participants were invited again to evaluate the first design. In this project, the purpose of user evaluation is to understand potential users' attitude and understanding of feedback contents, as well as their feelings about the UI, as a criteria for judging the success and the reference for subsequent iterations. Evaluation is still conducted by semi-interviews, but unlike user study ones, evaluative interviews are task-driven, rather than question-centered. Two examples of phishing emails are researched, and corresponding feedback are made. The emails with feedback were shown to each participant, who was asked to judge the emails based on the feedback. In addition, participants were shown the anti-phishing recommendation webpage of the UoE and an example result from Urlvoid to provide a comparative reference for different anti-phishing methods. Finally, obtain their opinions through some questions and follow-up discussions.

4.4.1 Feedback Content

For contents in the feedback, the following key question were set:

- According to the feedback, which parts of the two emails do you think are dangerous? How do you judge?
- Can you understand the customized information? Is there anything you don't understand?
- Overall, do you feel that this service is helping you? Will such a response encourage you to report suspicious emails in the future?
- Among the three anti-phishing methods shown, which one do you prefer? Which one do you dislike the most?

In general, participants showed positive feedback on the information provided by the prototype. By reading the automated response, 5 and 4 participants each had a strong awareness of the dangers of the two emails, while the rest also mentioned that they would not respond to those emails. Participants basically agreed that this feedback formed them some strategies of judging phishing through the header-url-body model. By asking questions on each feature, the 6 participants clearly stated that most of the feature descriptions are interesting and understandable. Half of them emphasized the usefulness of the guidance for contextual judgment in the feature explanation. However, the terms DMARC authentication and redirection still caused some confusion which were questioned by two participants. In addition, two participants had questions like 'is the URL in the blacklist necessarily harmful?' After getting an affirmative answer, they suggested that such determined phishing emails can be directly notified, so as to increase the efficiency for judging.

Additionally, all participants reckoned that compared to their original knowledge and expectations of anti-phishing services, they felt more attentive and helpful on this solutions, and 3 of whom claimed that the feedback would arouse their interest in reporting emails. Two participant were happy about the greeting words at the top and believed it could help build self-confidence, because they did not predict reporting suspicious emails is helpful. Three participants considered that the instruction to let users calm down can be effective as they did not have the knowledge on what tone legal organizations usually use in important notices. Another student admired the access to human support provided at the end. Compared with UoE's phishing recommendation webpage and Urlvoid, participants showed a strong preference for this feedback (all ranked first). Five participants ranked UoE's website as the last place, who generally thought it was boring and did not provide effective information. The remaining one disliked Urlvoid the least because terms such as reverse DNS and ASN are too professional to understand without any comments.

4.4.2 UI Design

The evaluation of the UI does not depend on specific questions, but is based on asking participants about their intuitive feelings in terms of different design aspects, such as function, color or layout. As a summary, the UI design received more negative feedback than the content. First of all, participants commonly thought that the page was cluttered. Two interviewees believed that this was caused by the current layout, who

suggested that all features should be displayed in the same style and regular positions, instead of placing them in differently or adding frames/arrows for emphasizing. Half of the total specified that texts in too many different colors made reading difficult. In addition, the other two participants thought that the feature area in the middle was weakened by the lighter background color, which is actually the more important part. Regarding the functionality, 2 participants thought that the summary was not reasonable. One of them suggested that the exclusion of email body features may make such information less important or useful in users' mental model. Another person believed the summary may not be enough to warn of danger, as no red feature will be displayed if no link tricks are found, even if it is a dangerous email. Thus, it is likely to mislead users. Another problem lies in the presentation of individual features. Although the participants generally believed that the information was understandable, half of them expected to see the title/name of features. They reminded the low efficiency of the demonstration, because it was hard to know what is being described unless reading the complete sentence.

In addition to the above feedback, participants expressed approval for the choice of colors and the arrangement of different areas, thus it is considered for the subsequent versions to retain these two styles. Four interviewees felt good about the traffic light color system or the blue theme color. And everyone pointed out more or less that the regionalization of different information is an intuitive design.

4.5 Final Design of the Autoresponder

A final prototype with several improvements is made based on user evaluation and some suggestions from my supervisor (figure 4.4). Since the invited users generally considered the original design is caring and interesting, greeting, introduction and some general guidance are retained for their satisfaction. The improvements are mainly focused on solving the problems of interface/function design and user comprehension. The UI are refactored, where the changes are obvious. The first is to simplify the colors in the feedback from 11 to 7 (figure 4.5). Currently, in addition to the four necessary colors that indicate the danger level, only three other colors are used. While maintaining the blue theme, the background color was unified into a dark blue, with the other two light blue backgrounds discarded. The 4 areas for phishing features are added with translucent gray containers for emphasis. Two general suggestion fields are also centered with two white edges for users to identify. In addition, to enhance readability, all

texts are colored white, which are distinguished only by font size, thickness and transparency. Another action to reduce reading costs is the display of features. In the final design, all the information of each feature is placed in rounded rectangles in different colors (figure 4.6) to strengthen the indication of different risks. Feature descriptions are simplified and emphasized, and feature names are added to improve users' efficiency in browsing and searching. The interpretation of the feature is still reserved for ordinary users, which are now in lower transparency, reducing its interference with the main content. Moreover, in the initial version, the arrangement of different features is irregular, and length of an individual component is determined by its content, which may be an important factor for the clutter of the page. Therefore, with the improvement, all components are either full-width or half-width. Features about similar aspects are placed in the same row (such as Authentication and Domain Check), which not only helps users to organize information, but also shortens the entire report.

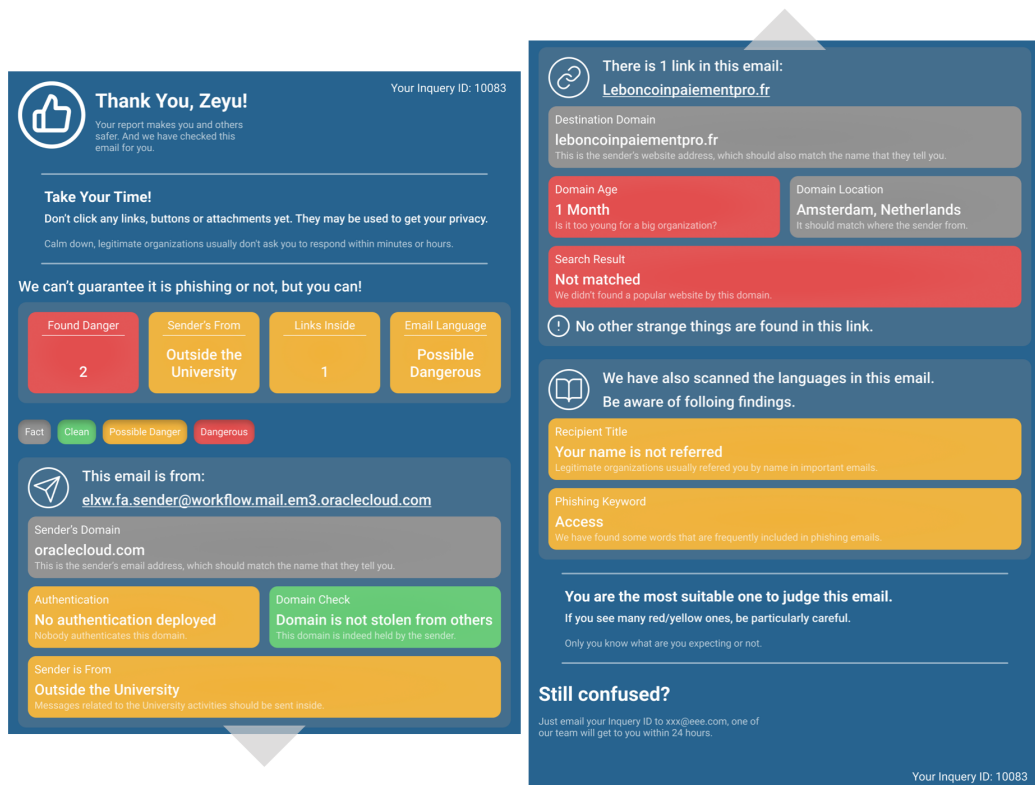


Figure 4.4: Final design with improvements.

In terms of function, 3 Key Features are set as described above to mark some confirmed phishing. The feature of search result is added to the URL to show the popularity of the webpages. In addition, the summary of features has also been redesigned.

The new function displays the total number of dangers, where the email is from, the number of links in the email, and the analysis of the email language, which finally covers the content in the following 3 areas and warns of the most dangerous information. My supervisor reminded that it is not ideal to display the link domain in this field, as it will be difficult for multiple links or a very long domain, which is now solved by showing the number of links. If no dangerous feature is found, the first component will be telling the number of possible dangers, and so on. If any Key Features are detected as dangerous, they will all be displayed in this area replacing the other four components, and a malicious email warning (figure 4.7) will be issued.

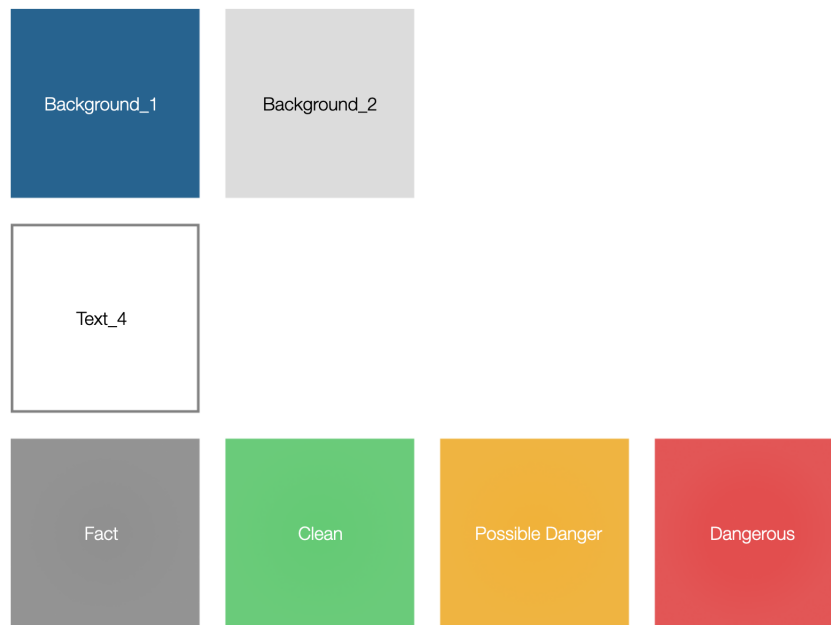


Figure 4.5: Color platters in the final design.



Figure 4.6: Single feature presentation.

To further improve user comprehension, the description of some features has been simplified, such as DMARC and redirection. Users will no longer see the term 'DMARC', but instead it is simply described by 'authentication' and 'stolen'. The redirection is now displayed as the more intuitive destination domain. Moreover, even if the redirection is not detected, it will not be considered clean to minimum the possibility of users being misled. Finally, some other languages have been slightly modified, such

as simplifying the language of the top bar, and adding 'If you see many red/yellow ones, be particularly careful' in the second general guidance as a clearer suggestion in judging the emails.

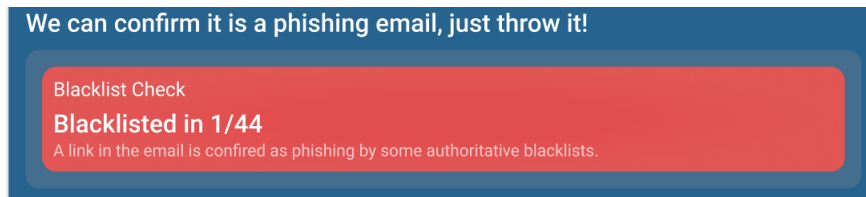


Figure 4.7: The summary module when a Key Feature is detected as dangerous.

Chapter 5

Discussion

5.1 Contribution

It is undeniable that the perfect automatic detection of phishing emails is the most effective and efficient security method. However, the common flaws of automatic filters have led to the necessity of supporting human identification of phishing. In theory, email recipients themselves are indeed more suitable for making decisions than machines or other people, because they are clear about their own context. The auto-responder I designed combines the human autonomous decision and more reliable automated detection methods. The significance of this study is to improve some defects that reduce the effectiveness of the existing human-centered phishing prevention, including insufficient defense against high-level phishing, limited detection range, poor user comprehension, and unfriendly user experience. The artifact of the autoresponder is expected to reduce the risk of phishing attacks for users in the university, which can bring benefit for students, staff and IT help desk team.

As a design-oriented project, the functional design and UI construction of the responder are the main interest, rather than the implementation. First, I studied typical phishing behaviors through academic review, and selected several features that may be suitable for ordinary users in terms of the email header, attached links, and text accordingly. Since the current involved features have been basically used in existing auto-filters or user-support work, the autoresponder can be feasible to implement. As for the prototype design, in addition to feature selection, it also covers user study, demand analysis, UI design and some improvements based on evaluation. 6 non-experts from UoE were invited into the user study and the evaluation. Participants' feedback shows that most of the features are easy to understand for them, and are more helpful

compared to their own judgment or using some other supports. And they can basically feel the friendliness of the service. Thus, the evaluation basically indicate the achievement of the study goal. In addition, some negative comments and suggestions were collected, mainly about information layout and functionality, and there were also confusions about a few features. In the end, an iterated prototype is made, with several improvements inspired from the issues discovered in the evaluation.

5.2 Limitation

Due to time and resource constraints, there are some aspects of this work that still can be improved. First of all, current evaluation of the design may not be sufficient to deeply obtain users' opinions and suggestions. As only six non-experts have been invited for one round of evaluative interviews, it could only focus on user comprehension testing and obtaining the first impression of the UI. According to the original plan, the mockup should go through more iterations by focus group evaluation with different purposes. Compared with interviews, a major advantage of focus group is that it can be used to collect more targeted comments through different user groups or experts in different domains. For example, a group of design experts can be invited to give specific suggestions on the layout, or form a cybersecurity expert group for feedback one feature explanation. In addition, participants may also find extra demands through discussion, or continue to give feedback with evaluations running. However, the actual situation did not support more iterations, because it would take several days for discussion, analysis and modification of the design for each round. Besides, to get more effective comments, formal focus groups should include more users and experts from different backgrounds as participants, which is also difficult for me to access. Another evaluation that was not conducted as planned is a larger-scale verification of the design, which is supposed to collect 200 volunteer feedbacks through online questionnaire. The purpose of such an online survey is to test the satisfaction and effectiveness of the solution on wider extend of users, which was also omitted due to the time limitation.

In addition, although the feedback I designed should be able to provide customized suggestions based on features after being implemented, however, the current support for contextual judgement is still limited to some general advice, which means that extra rational thinking will be still needed to connect the detected feature with users' actual situations, which may be difficult for some panicked users or those who have weak

security mental models. For example, they may not consider a ‘Paypal’ website with a 3-month-old domain as dangerous. Therefore, this is considered a potential defect that may fail the feedback. It may be possible to solve this problem by strengthening the contextual warning, such as obtaining the source in the email through certain methods, allowing the machine to make more comparisons for users. However, since this idea has not been found in any existing work, I did not have enough time to investigate its technical feasibility and add it into the design.

5.3 Conclusion

In summary, I designed an automatic responder for phishing email reports, which is technical feasible to implement to return immediate feedback to the inquirer. According to the design, the responder can automatically parse the reported emails and notify the findings through specific colors and texts, ranging from clean to dangerous. This solution combines the concepts of some existing human-centered strategies, which cover a wider range of phishing features compared to most user-support tools, including email headers, internal URLs, and email body. Also it focuses on explaining those features to ordinary users. Through user evaluation on 6 non-experts participants, the feedback is demonstrated to be generally comprehensible, helpful and friendly. Finally, the design was improved in terms of function, layout and user understanding based on participants’ experience.

5.4 Feature Work

Future work first depends on further improving the feedback design. According to the potential defects mentioned in the Limitation section, it is suggested to add more customized warning about the context. It might be a good idea to consider some automatic methods to dig deeper into the reported emails, such as the sender’s self-proclaimed information. In this way, it can be possible to provide users with more effective support because more diverse automated analysis will become feasible. After that, verifying the selected features through quantitative analysis can be a reasonable step as some existing works do. Authoritative phishing databases such as PhishTank can be used for counting the frequency of features in phishing emails, and high-frequency features can be regarded as valid. In addition, experts in design, HCI, psychology, and cybersecurity can be invited for expert interviews or several focus groups to obtain more com-

prehensive and professional feedback to support subsequent improvements. Finally, an implementation can be considered based on the improved design, which is the key to making the autoresponder truly effective and being applied in realistic context.

Bibliography

- [1] Neda Abdelhamid, Aladdin Ayesh, and Fadi Thabtah. Phishing detection based associative classification data mining. *Expert Systems with Applications*, 41(13):5948–5959, 2014.
- [2] Ahmed Aleroud, Emad Abu-Shanab, Ahmad Al-Aiad, and Yazan Alshboul. An examination of susceptibility to spear phishing cyber attacks in non-english speaking communities. *Journal of Information Security and Applications*, 55:102614, 2020.
- [3] Kholoud Althobaiti, Nicole Meng, and Kami Vaniea. I don't need an expert! making url phishing features human comprehensible. In *The ACM CHI Conference on Human Factors in Computing Systems 2021*. Association for Computing Machinery (ACM), 2021.
- [4] Kholoud Althobaiti, Ghaidaa Rummani, and Kami Vaniea. A review of human-and computer-facing url phishing features. In *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 182–191. IEEE, 2019.
- [5] Giovanni Apruzzese, Michele Colajanni, Luca Ferretti, Alessandro Guido, and Mirco Marchetti. On the effectiveness of machine and deep learning for cyber security. In *2018 10th international conference on cyber Conflict (CyCon)*, pages 371–390. IEEE, 2018.
- [6] Nalin Asanka Gamagedara Arachchilage and Steve Love. A game design framework for avoiding phishing attacks. *Computers in Human Behavior*, 29(3):706–714, 2013.
- [7] 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.

- [8] Brandon Atkins, Wilson Huang, et al. A study of social engineering in online frauds. *Open Journal of Social Sciences*, 1(03):23, 2013.
- [9] Simon Bell and Peter Komisarczuk. An analysis of phishing blacklists: Google safe browsing, openphish, and phishtank. In *Proceedings of the Australasian Computer Science Week Multiconference*, pages 1–11, 2020.
- [10] Aaron Blum, Brad Wardman, Thamar Solorio, and Gary Warner. Lexical feature based phishing url detection using online learning. In *Proceedings of the 3rd ACM Workshop on Artificial Intelligence and Security*, pages 54–60, 2010.
- [11] Roderic Broadhurst, Katie Skinner, Nick Sifniotis, Bryan Matamoros-Macias, and Yuguang Ipsen. Phishing and cybercrime risks in a university student community. Available at SSRN 3176319, 2018.
- [12] Jake D Brutlag and Christopher Meek. Challenges of the email domain for text classification. In *ICML*, volume 2000, pages 103–110, 2000.
- [13] Marcus Butavicius, Kathryn Parsons, Malcolm Pattinson, and Agata McCormac. Breaching the human firewall: Social engineering in phishing and spear-phishing emails. *arXiv preprint arXiv:1606.00887*, 2016.
- [14] Gamze Canova, Melanie Volkamer, Clemens Bergmann, and Benjamin Reinheimer. Nophish app evaluation: lab and retention study. In *NDSS workshop on usable security*, 2015.
- [15] Happy Chapla, Riddhi Kotak, and Mittal Joiser. A machine learning approach for url based web phishing using fuzzy logic as classifier. In *2019 International Conference on Communication and Electronics Systems (ICCES)*, pages 383–388. IEEE, 2019.
- [16] Rachna Dhamija, J Doug Tygar, and Marti Hearst. Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 581–590, 2006.
- [17] Yan Ding, Nurbol Luktarhan, Keqin Li, and Wushour Slamun. A keyword-based combination approach for detecting phishing webpages. *computers & security*, 84:256–275, 2019.

- [18] Julie S Downs, Mandy B Holbrook, and Lorrie Faith Cranor. Decision strategies and susceptibility to phishing. In *Proceedings of the second symposium on Usable privacy and security*, pages 79–90, 2006.
- [19] Ian Fette, Norman Sadeh, and Anthony Tomasic. Learning to detect phishing emails. In *Proceedings of the 16th international conference on World Wide Web*, pages 649–656, 2007.
- [20] Tushaar Gangavarapu, CD Jaidhar, and Bhabesh Chanduka. Applicability of machine learning in spam and phishing email filtering: review and approaches. *Artificial Intelligence Review*, pages 1–63, 2020.
- [21] Surbhi Gupta, Abhishek Singhal, and Akanksha Kapoor. A literature survey on social engineering attacks: Phishing attack. In *2016 international conference on computing, communication and automation (ICCCA)*, pages 537–540. IEEE, 2016.
- [22] Isredza Rahmi A Hamid and Jemal Abawajy. Hybrid feature selection for phishing email detection. In *International Conference on Algorithms and Architectures for Parallel Processing*, pages 266–275. Springer, 2011.
- [23] Reza Hassanpour, Erdogan Dogdu, Roya Choupani, Onur Goker, and Nazli Nazli. Phishing e-mail detection by using deep learning algorithms. In *Proceedings of the ACMSE 2018 Conference*, pages 1–1, 2018.
- [24] Paul Hoffman et al. Smtip service extension for secure smtp over transport layer security. Technical report, RFC 3207, February, 2002.
- [25] Jason Hong. The state of phishing attacks. *Communications of the ACM*, 55(1):74–81, 2012.
- [26] Jiwon Hong, Taeri Kim, Jing Liu, Noseong Park, and Sang-Wook Kim. Phishing url detection with lexical features and blacklisted domains. In *Adaptive Autonomous Secure Cyber Systems*, pages 253–267. Springer, 2020.
- [27] Hang Hu, Steve TK Jan, Yang Wang, and Gang Wang. Assessing browser-level defense against idn-based phishing. In *30th {USENIX} Security Symposium ({USENIX} Security 21)*, 2021.

- [28] Ankit Kumar Jain and BB Gupta. Phish-safe: Url features-based phishing detection system using machine learning. In *Cyber Security*, pages 467–474. Springer, 2018.
- [29] Ankit Kumar Jain and Brij B Gupta. Two-level authentication approach to protect from phishing attacks in real time. *Journal of Ambient Intelligence and Humanized Computing*, 9(6):1783–1796, 2018.
- [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(1):1–41, 2020.
- [31] K Jansson and Rossouw von Solms. Phishing for phishing awareness. *Behaviour & information technology*, 32(6):584–593, 2013.
- [32] Kanako Konno, Naoya Kitagawa, and Nariyoshi Yamai. False positive detection in sender domain authentication by dmarc report analysis. In *Proceedings of the 2020 The 3rd International Conference on Information Science and System*, pages 38–42, 2020.
- [33] Murray Kucherawy and Elizabeth Zwicky. Domain-based message authentication, reporting, and conformance (dmarc). *ser. RFC7489*, 2015.
- [34] 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*, pages 1–12, 2009.
- [35] Ponnurangam Kumaraguru, Yong Rhee, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. Protecting people from phishing: the design and evaluation of an embedded training email system. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 905–914, 2007.
- [36] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. Testing phishguru in the real world. In *Proceedings of the Symposium on Usable Privacy and Security*, 2007.

- [37] Youngsun Kwak, Seyoung Lee, Amanda Damiano, and Arun Vishwanath. Why do users not report spear phishing emails? *Telematics and Informatics*, 48:101343, 2020.
- [38] Elmer Lastdrager, Inés Carvajal Gallardo, Pieter Hartel, and Marianne Junger. How effective is anti-phishing training for children? In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*, pages 229–239, 2017.
- [39] Eric Lin, Saul Greenberg, Eileah Trotter, David Ma, and John Aycock. Does domain highlighting help people identify phishing sites? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2075–2084, 2011.
- [40] Liping Ma, Bahadorrezda Ofoghi, Paul Watters, and Simon Brown. Detecting phishing emails using hybrid features. In *2009 Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing*, pages 493–497. IEEE, 2009.
- [41] Eva Marková, Tomáš Bajtoš, Pavol Sokol, and Terézia Mézešová. Classification of malicious emails. In *2019 IEEE 15th International Scientific Conference on Informatics*, pages 000279–000284. IEEE, 2019.
- [42] D Kevin McGrath and Minaxi Gupta. Behind phishing: An examination of phisher modi operandi. *LEET*, 8:4, 2008.
- [43] Craig M McRae and Rayford B Vaughn. Phighting the phisher: Using web bugs and honeytokens to investigate the source of phishing attacks. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, pages 270c–270c. IEEE, 2007.
- [44] 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 *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 130–139. IEEE, 2020.
- [45] Dyaa Eldeen Nasr Motawa and Ahamed El Shrief. A url with image-based feature extraction for preventing phishing attacks. *Journal of Information Security and Cybercrimes Research*, 2(1):116–127, 2019.

- [46] Tejas Nanaware, Prashant Mohite, and Rajendra Patil. Dmarcbox—corporate email security and analytics using dmarc. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, pages 1–5. IEEE, 2019.
- [47] Stephen J Nightingale and Stephen J Nightingale. *Email Authentication Mechanisms: DMARC, SPF and DKIM*. US Department of Commerce, National Institute of Standards and Technology, 2017.
- [48] Adam Oest, Yeganeh Safei, 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)*, pages 1–12. IEEE, 2018.
- [49] Kunal Pandove, Amandeep Jindal, and Rajinder Kumar. Email spoofing. *International Journal of Computer Applications*, 5(1):27–30, 2010.
- [50] Proofpoint. 2021 state of the phish. <https://www.proofpoint.com/sites/default/files/threat-reports/pfpt-us-tr-state-of-the-phish-2021.pdf>. Proofpoint, Annual Report. Online; accessed 3 May, 2021.
- [51] Silvia E Rabionet. How i learned to design and conduct semi-structured interviews: An ongoing and continuous journey. *The Qualitative Report*, 14(3):203–206, 2009.
- [52] Justinas Rastenis, Simona Ramanauskaitė, Ivan Suzdalev, Kornelija Tunaitytė, Justinas Janulevičius, and Antanas Čenys. Multi-language spam/phishing classification by email body text: Toward automated security incident investigation. *Electronics*, 10(6):668, 2021.
- [53] Umesh Kumar Sah and Narendra Parmar. An approach for malicious spam detection in email with comparison of different classifiers. *International Research Journal of Engineering and Technology (IRJET)*, 4(8):2238–2242, 2017.
- [54] Himani Sharma, Er Meenakshi, and Sandeep Kaur Bhatia. A comparative analysis and awareness survey of phishing detection tools. In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pages 1437–1442. IEEE, 2017.

- [55] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason 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*, pages 88–99, 2007.
- [56] Chiranjib Sur. Ensemble one-vs-all learning technique with emphatic & rehearsal training for phishing email classification using psychology. *Journal of Experimental & Theoretical Artificial Intelligence*, 30(6):733–762, 2018.
- [57] Talos. Email & spam data. https://talosintelligence.com/reputation_center/email_rep. Talos. Online; accessed 2 May, 2021.
- [58] Khoi-Nguyen Tran, Mamoun Alazab, Roderic Broadhurst, et al. Towards a feature rich model for predicting spam emails containing malicious attachments and urls. 2014.
- [59] Urlvoid. Anti-phishing solution. <https://www.urlvoid.com>. Urlvoid. Online; accessed 20 August, 2021.
- [60] Arun Vishwanath. Examining the distinct antecedents of e-mail habits and its influence on the outcomes of a phishing attack. *Journal of Computer-Mediated Communication*, 20(5):570–584, 2015.
- [61] Liu Wenyin, Gang Liu, Bite Qiu, and Xiaojun Quan. Antiphishing through phishing target discovery. *IEEE Internet Computing*, 16(2):52–61, 2011.
- [62] WhereGoes. Url redirect checker. <https://wherergoes.com>. WhereGoes. Online; accessed 20 August, 2021.
- [63] 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.
- [64] J Yao and W Mao. Smtip extension for internationalized email addresses. *RFC 5336 (Experimental)*, Internet Engineering Task Force, 2008.
- [65] Adwan Yasin and Abdelmunem Abuhasan. An intelligent classification model for phishing email detection. *arXiv preprint arXiv:1608.02196*, 2016.

- [66] Zhaowu Yu, Gaoyuan Yang, Shudi Zuo, Gertrud Jørgensen, Motoya Koga, and Henrik Vejre. Critical review on the cooling effect of urban blue-green space: A threshold-size perspective. *Urban forestry & urban greening*, 49:126630, 2020.
- [67] Hiba Zuhair Zeydan, Ali Selamat, and Mazleena Salleh. Survey of anti-phishing tools with detection capabilities. In *2014 International Symposium on Biometrics and Security Technologies (ISBAST)*, pages 214–219. IEEE, 2014.
- [68] Jian Zhang, Zhen-Hua Du, and WEI Liu. A behavior-based detection approach to mass-mailing host. In *2007 International Conference on Machine Learning and Cybernetics*, volume 4, pages 2140–2144. IEEE, 2007.
- [69] Yue Zhang, Serge Egelman, Lorrie Cranor, and Jason Hong. Phinding phish: Evaluating anti-phishing tools. 2007.