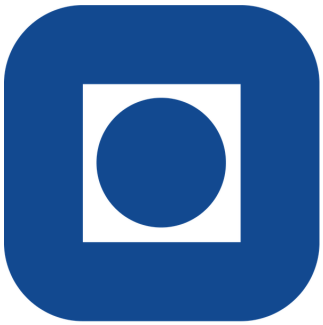
An Extensive Analysis of Email Phishing

Properties, Detection, and Successful Phishing

Sigrid Anne Hafsahl Karset
494330

2022-05-27



NTNU

Department of Information Security and Communication Technology
Gjøvik, Norway

Abstract

Title: An Extensive Analysis of Email Phishing
Date: 2022.05.27
Authors: Sigrid Anne Hafsahl Karset
Supervisor: Patrick Bours

Keywords: Phishing, Email Phishing, Phishing Detection, Information Security
Pages: 52
Studypoints: 7.5 ECTS

Abstract

Phishing as a form of Social Engineering has a prominent presence in today's threat landscape, and can result in devastating losses if successful. This paper presents an analysis of the phishing emails that have been observed in Q1 of 2022. The analysis sheds light on trends that can be seen with these phishing emails in relation to properties such as Content, Target, Method, and Impersonation. In addition to defining phishing email properties, the paper analyzes phishing features that have been found in previous literature. These features are compared to the phishing emails collected, and it is shown that many of these features are present in this paper's corpus, while others have no presence at all. Furthermore, the paper showcases instances of successful phishing emails, pointing out the characteristics that can be seen in a successful phish.

Contents

Abstract . . . . . **ii**
Contents . . . . . **iii**
Figures . . . . . **v**
Tables . . . . . **vi**
1 Introduction **1**
1.1 Background 1
1.2 Project Definition 1
1.3 Scope 2
1.4 Structure 2
2 Theory - Phishing **4**
2.1 Early Phishing and Evolution 4
2.2 Phishing Life Cycle 6
2.3 Categories of Phishing 7
3 Related and Relevant Work **8**
4 Methodology **12**
4.1 Information Collection 12
4.1.1 Literature Collection 12
4.1.2 Data Collection 13
4.2 Data Analysis 13
5 Data Collection **15**
5.1 Phishing Emails 15
5.2 Successful Phish 17
6 Findings **18**
6.1 Content 18
6.2 Target 23
6.3 Method 25
6.4 Impersonation 26
6.5 Time Distribution 27
7 Prior Research **30**
7.1 Features 30
7.2 What People Fall For 34
7.2.1 Excerpt 34
7.2.2 Aspects of a successful phish 37
7.2.3 Summary 39

8 Discussion **40**
8.1 Top Categories 40
8.2 HTML Attachments 41
8.3 Targeted Brands 43
8.4 Additional Detection Features 44
8.5 Validity of Data 45
9 Conclusion **46**
9.1 Challenges 47
9.2 Future Work 48
9.2.1 Investigate top categories 48
9.2.2 Repeated analyses 48
9.2.3 Automated collection tool 48
9.2.4 Additional domains 48
9.2.5 Test detection features 49
Bibliography **50**
A Content Categories **53**

Figures

6.2 CEO Scam - Gift Card - Example 18
6.1 Overview - Content 19
6.3 Password Expires - Example 20
6.4 Document Shared - Example 21
6.5 Invoice - Example 22
6.6 Post Payment - Example 22
6.7 Target to Content 24
6.8 Target to Method 26
6.9 Distribution of Organizations 27
6.10 CEO Scam - Gift Card Heat Map 27
6.11 Password Expires Heat Map 28
6.12 Update Bank Info Heat Map 28
6.13 Refund Heat Map 28

7.1 Fell for - Password Expires 34
7.2 Fell for - Click to Release Mail 35
7.3 Fell for - Authenticate to Continue Receiving Mails 35
7.4 Fell for - Invoice 36
7.5 Fell for - CEO Scam - Gift Card 36
7.6 Fell for - Post Payment 37

8.1 HTML Linked - Source Code 41
8.2 HTML Linked - Landing Page 42
8.3 HTML Local - Source Code Snippet 42
8.4 HTML Local - Landing Page 42

Tables

6.1 Overview - Target 23
6.2 Overview - Method 25
6.3 Overview - Impersonation 26

7.1 Features 31
7.2 Dots in URL 32
7.3 URL Length 32
7.4 Domains in URL 32
7.5 No. of links in Mail 33
7.6 Domains Linked to 33

Chapter 1

Introduction

1.1 Background

The digitization of the business world has lead to more and more of their respective processes being fully performed and carried out on digital mediums. Although many tasks are being digitized and automated, there is still a need for human intervention and decision making. The human factor in information technology is viewed as a weak spot, and it is stated that 88% of today's data breaches are caused by human errors [1]. As humans are psychologically driven, they can be lured and manipulated to perform tasks with undesirable consequences. Such manipulation is often referred to as Social Engineering and utilizes the digitization of data and processes to both collect information and perform malicious activities. One of the most prominent attacks in this Social Engineering category is Phishing, which uses the possibility of digital communication in order to launch its attack. These types of attacks tries to lure out sensitive information, such as passwords and credit card details, or trick the victim into downloading malicious software.

Due to the human factor in information technology, phishing attacks targeting this weakness have a heavy presence in today's society, both on a private and business basis. These attacks can target everything from monetary values to basic information, and can have disastrous consequences. Because of this, it is highly important to be aware of such attacks, including how they look and operate, in order to prevent successful attacks.

1.2 Project Definition

This paper will collect and analyze an array of various phishing emails in order to create an understanding as to how these types of attacks operate and presents themselves, as well as if there are any specific types of phishing emails that people tend to fall for. To fully understand the phishing phenomenon, a breakdown of the

attack's evolution and concepts will be presented. In inclusion of this, prior literature regarding the subject will be analyzed with the intention of viewing how the findings in these papers compares to the observations made in the collected phishing emails. A focus point with the prior literature will be features depicted as indications of a phishing mail, and showing how these features are present in the project's phishing corpus.

Three research questions are identified for this paper:
- How is the email phishing landscape today?
- How does prior research relate to present observations?
- What phishing emails do people fall for?

1.3 Scope

This project will primarily focus on phishing emails received by business entities as opposed to emails received on private accounts, due to the method of collection. This means that the emails collected and analyzed will be emails sent to an account registered to an employee or group within an organization. As some employees utilizes their business emails for personal purposes, the collected phishing emails will have entries reflecting privately targeted phishing emails as well. These organizations from which the emails are collected are mostly Norwegian based, with some entities based in other countries. Therefore, many of the examples of phishing emails analyzed and visualised will be in Norwegian.

As the project is limited to one semester, that is, 5 months, the collection window for the phishing emails is narrowed down to the months of January, February, and March. The phishing corpus of this project will therefore consist of emails representing the first quarter of 2022.

1.4 Structure

The paper is divided into nine chapters and one appendix.

Chapter 1 - Introduction
The introduction of the project includes the project's background, definition, research questions, and scope.

Chapter 2 - Theory - Phishing
The Theory chapter gives an overview of phishing, including what it is, its evolution, the phishing life cycle, and categories of phishing.

Chapter 3 - Related and Relevant Work
The chapter presents prior scholarly literature and research in the area of phish-

ing. This chapter is used as the basis for the analysis conducted in Chapter 7.

Chapter 4 - Methodology
The chapter describes how the project was conducted in the areas of the Literature Collection and Analysis, Data Collection, and Data Analysis.

Chapter 5 - Data Collection
The chapter details the data collection including the specific tools used and their description, the properties collected from the phishing emails, and specifications toward what constitutes as successful phishings.

Chapter 6 - Findings
The Findings chapter depicts the results from the collected phishing emails. The results are both presented with statistical and textual descriptions.

Chapter 7 - Prior Research
This chapter uses the features from the literature collected in Chapter 3 and compares them to the observations made from the project's phishing corpus. This chapter is divided into two parts consisting of 1. Features of phishing emails, and 2. features of successful phishing emails.

Chapter 8 - Discussion
This chapter discusses in more detail some of the findings of Chapters 6 and 7.

Chapter 9 - Conclusion
The conclusion of the project provides answers to the research questions, as well as notes on challenges and future work.

Appendix
Additional material relevant to the paper, but not necessary in order to understand the paper.

Chapter 2

Theory - Phishing

Before going into detail about how phishing can be perceived in today's landscape, it is important to understand the fundamentals of the phenomenon. This chapter will present an overview of phishing including what it is, its evolution, the phishing life cycle, and the various categories of phishing.

As mentioned, phishing is a type of social engineering attack where the main goal is to lure out sensitive information, such as passwords and credit card details, or trick the victim into downloading malicious software. These attacks often gain the target's attention by posing as a familiar entity, like a colleague or an organization, or making a desirable offer such as promising large amounts of money.

The result of such an attack might be devastating, should one fall victim for it. A password in the wrong hands can give a malicious actor access to the users account and all that lays within. If an organizational account is compromised, the actor can impersonate said person to gain access to sensitive information or communication channels to deploy further phishing attacks. High monetary losses can occur through credit card scams or payment of fraudulent invoices. While loss and leakage of sensitive information can be a result of malicious code deployment.

2.1 Early Phishing and Evolution

Although it is hard to determine what was the first instance of the phishing phenomenon, many sources, including [2] [3] [4], state the AOL scams from 1995 as one of the earliest usages of the term "Phishing". AOL (America Online) was a popular internet provider in the earlier days, where their platform provided chat rooms and direct messaging for their customers. Malicious actors took advantage of these functions by posing as employees to target new users into giving up their passwords. The actors would create accounts with official names such as "BillingDept" and send a direct message to a target user asking them to verify their account by providing their username and password. An example of such message

was as follows:

"*Hi, this is AOL customer service. Due to a problem with our records, we need you to reply to this message with your current password in order to avoid being disconnected.*" [3].

Due to general low awareness and experience with the internet, many users provided their login credentials giving the attackers access to their account. From here, the attackers would use this access to provide themselves with free internet access.

Although this phishing attack took place almost 30 years ago, the basis of the attack has not changed much (as will be seen in more detail in the following chapters). The general idea is still the same; trick a user into doing a desirable act for you by pretending to be someone you're not. What has changed however, is the technology used to conduct these attacks.

Electronic mail had been around since before the internet itself [5], but some development in the early to mid 90's paved the way for a new generation of phishing attacks. In 1992 the "attachment" was born [6]. Before this, one could mostly only send text. Now, they could attach files such as small applications or scripts. Another advancement on the email front was "non-tied" e-mail services [6]. These were email services that were not dependent on a specific ISP or network, and utilized the World Wide Web to connect its users. Phishers could now reach a wast magnitude of people using one simple service. One of the more known phishing incidents utilizing these two opportunities, was the "Love Bug" that emerged in 2000 [7]. The Love Bug was a phishing campaign masquerading as a love letter, where users were enticed to open an attachment, thinking it was the love letter, but in reality it was a worm. The worm would then send it self to all of the user's contacts, causing it to spread rapidly.

The popularization of social networking sites would be the next evolutionary step in the world of phishing. The popularization and shift in social media as we know it today occurred in the early 2000's with MySpace and mid to late 2000's with Facebook [8]. These platforms made for an easy way for phishers to gather information about their targets, and use it to conduct specifically crafted phishing attacks towards them. "Spear phishing" as it is called, is meant to gain our trust by coming from a seemingly known source and/or discussing a personally relevant subject, increasing the chances of a successful phishing attempt [9]. The more information one can collect about a specific person, the more in-depth the phishing attempt can be. This made the popularization of such social media platforms an integral part in the evolution of phishing.

Lastly, the digitization of organizations showed a shift in the outcome a successful phishing attack could have. Up until now, phishing was mainly targeted towards

individual people, trying to get their credit card information or infecting their personal computer. Through the digitization of businesses, phishers could now gain access to entire corporate networks through one simple successful password phishing attempt, such as in [10]. Instead of getting access to one persons credit card, they could now lure their way into hundreds of thousand of cash, like the Goolge-Facebook scam [11]. And instead of rendering a personal computer inaccessible, they could put a stop to an entire business' operation, as shown in the Ukraine power grid hack [12].

Phishing is no longer an activity utilized only by teenagers for small monetary gains as with the AOL attacks. It has evolved to become a technique widely used by both individuals and governmental cyber espionage and threat groups, with the potential to cause great harm not only to individual victims, but also to organizations and governmental entities as well.

2.2 Phishing Life Cycle

Although phishing can be utilized in different ways, by different actors, and with varying goals, literature suggest [13][14][15] that most phishing attacks follow a fairly equal life cycle consisting of a planning phase, execution phase, infiltration phase, collection phase, and a clean-up phase.

1. *Planning Phase*
   The planning phase can involve identifying the targets (A person, an organization, a list of email addresses, and the like), collect necessary information such as organizational role or commonly used platforms, set up a phishing site, create fraudulent invoices, develop malicious code, and construct a phishing message.

2. *Execution Phase*
   The execution phase involves launching the phishing attack, such as sending the phishing mail, calling the target, or intercept traffic.

3. *Infiltration Phase*
   The infiltration phase involves having the target perform the desirable action. This can be the target downloading your malware, typing in their password on the phishing site, or paying the false invoice.

4. *Collection Phase*
   The collection phase consist of the attacker retrieving the desirable information, such as extracting the money, logging the targets keystrokes, or collecting its credentials.

5. *Clean-up Phase*
   Lastly, the clean-up phase consists of all the actions performed to hide and conceal the evidence of the attack. This may be deleting the phishing web site or removing the malicious software.

2.3 Categories of Phishing

As can been derived from the evolution of phishing, there exists today many approaches when launching a phishing attack. Chiew et al. in their paper [13], define three categories of phishing: Phishing, Smishing, and Vishing, including vectors (applications and services) in which these attacks can be carried out over.

Phishing, as we are familiar with from before, incorporates the attacks carried out over the Internet. Vectors such as Email, eFax, Instant Messaging, Social Network, Websites, and WiFi is defined for this category.

Smishing stands for SMS phishing, and relates to the attacks carried out over the phone messaging service and mobile messaging apps.

Vishing means Voice phishing, and relates to the attacks carried out over phone calls.

Based on these categories and their underlying vectors, various technical approaches can be utilized in order to conduct a phishing attack. Technical approaches such as Man-in-the-Middle, Spear Phishing, Clickjacking, Browser vulnerabilities, and Search Engine Optimisation are amongst some of the approaches highlighted in the aforementioned paper. These technical approaches can be used in combination with each other to create a functional phishing attack. For example, Spear Phishing and browser vulnerabilities, or one can utilize a man-in-the-middle approach to intercept emails, such as invoices, and change its details before relaying it to the intended recipient.

This paper will primarily encompass the internet based vectors, including Email, eFax, and Websites, and their applicable technical approaches.

Chapter 3

Related and Relevant Work

Papers of previous studies have been analyzed for similar research and concepts that could be used when conducting this study. Some of the papers were fully relevant for this project, while others contained small aspects that were of interest. Following, a description of the papers of relevance will be presented.

A survey of phishing attack techniques, defence mechanisms and open research challenges [14]

In Jain & Gupta's paper, the researchers conducted an analysis of previous work related to phishing with a focus on phishing attack methods and defence techniques. The paper presents the phishing life cycle, its history, main motivations behind such attack, distribution methods, and taxonomy, while also detailing protection methods. Although this paper has many similarities with this project study, the main points that are relevant for this study are their parts on motivation, targeted brands, phishing emails, and phishing website features.

Motivation
The paper states three main motivational points of a phisher, which are as follows; Financial benefits, Impersonating identity, and Gaining fame.

Targeted Brands
Apple IDs were the most targeted brand (25%), while Microsoft Outlook (17 %) and Google Drive (13 %) came second and third. The seven next brands were USAA (12%), Paypal (11%), Adobe Account (6%), DropBox (5 %), BlackBoard (5 %), LinkedIn (4 %), and CapitaOne (2 %)

Phishing Emails
Emails, as a vector for conducting phishing attacks, are described as the largest vector with 91 % of phishing attacks having its starting point with an email. Furthermore, it is described that the main reasons for falling for such emails were, listed from the most prominent: curiosity, fear, urgency, award incentive, enter-

tainment, and opportunity.

Phishing Website Features
As this project will mainly focus on the phishing emails, not all of the phishing website features listed in this paper are relevant. The paper lists the following URL features as interesting when identifying a phishing link: IP address as URL, Number of sub domains, "@" in the URL, "-" in hostname, URL length, Position of Top-level domain, Suspicious words in URL*, Number of domains in URL, HTTPS protocol, Brand name in URL, Redirect from URL.

*The list of Suspicious words are not stated in the paper

Phishing Attacks Survey: Types, Vectors, and Technical Approaches [16]

Alabdan's paper details the various phishing approaches with a focus on their characteristics, as well as identifying phishing prevention techniques. Relevant for this project is aspects of a successful phishing attack and manipulation of emotions.

Aspects of a Successful Phishing Attack
Authority (Complying to an authority figure), Commitment (Needing to "defend" their stance), Liking (Doing something for someone they like), Contrast (Making an unreasonable option desirable compared to another), Reciprocity (Acting positively to a desirable action), Scarcity (Increased desirability due to limited offer), and Social Proof (Do what every one else is doing) are pointed out as aspects of the human psychology that a successful phishing attack relies on.

Manipulation of Emotions
Ten emotions are listed as emotions that is desirable to manipulate in order to prevent the target from acting rationally. These are: Greed, Fear, Anger, Patriotism, Friendship, Sense of Duty, Sense of Belonging, Sense of Authority, Philanthropy, and Vanity.

Classification of Phishing Email Using Random Forest Machine Learning Technique [17]

Akinyelu & Adewumi utilized machine learning in order to create an automated detection and classifier of phishing emails. The paper details 10 features that were used in the detection algorithm to determine the likelihood of a mail being phishing. The reported accuracy of their model was 99.7 %.

The following features were used in this model to indicate the likelihood of a mail being phishing:

- URLs containing IP addresses
- Disparities between "href" attribute and LINK Text. (LINK text is what the

user sees in the text of the email. If the LINK text i heavily different from the actual URL, it may indicate a phishing link)
- The words "Link", "Click", "Login", "Update" and "Here" in the LINK text.
- Number of dots in the URL. (More than three dots is considered suspicious)
- HTML Email (As phishing emails often rely on links, an HTML formatted mail is more likely to be phishing)
- Presence of JavaScript (JavaScript can be used to hide content from the user)
- Number of links (More links, more likely to be phishing)
- Number of domains linked to.
- Discrepancy of sender domain and linked domain (Difference in the sender domain and linked domain may be an indication of phishing)
- Frequently used words (Words in the word groups (1) Update; Confirm, (2) User; Customer; Client, (3) Suspend; Restrict; Hold, (4) Verify; Account; Notif, (5) Login; Username; Password; Click; Log, (6) SSN; Social Security; Secur; Inconcinien)

Detection method of phishing email based on persuasion principle [18]

Li et al. describe in their paper a phishing email detection method based on the persuasion principle. The method utilizes various features to determine whether a mail is phishing. The features were divided into two categories consisting of basic features and persuasion features. Similarly to the work by Akinyelu & Adewumi, this paper identifies six of the same features, including: IPs in URL, HTML format, Disparities between "href" and LINK text, No. of dots in URL, and Discrepancy between sender and linked domain. The paper include "Click" and "Here" as common LINK text words, however add "Debit" to the list as well. Two additional new features are mentioned consisting of: Fwd in subject, and Site redirection.

Persuasion features identified:

- Authority
- Scarcity
- Reciprocation

Phishing Email Detection Based on Hybrid Features [19]

In their paper, Yang et al. propose a phishing email detection method based on hybrid features. The features consist of data from the email header, structure, URL information, script functions, and psychological features. Their method gave a 95 % accuracy in featured tests.

The features identified draw similarities from the features identified in both Akinyelu & Adewumi and Li et al's papers. In total, seven of the features identified were depicted in the aforementioned papers. This included: HTML format, IPs in

URL, No. of links, No. of domains linked to, Discrepancy between "href" and LINK text, and Presence of JavaScript. While nine new features were identified:

- Inclusion of blacklisted words (Account, Access, Bank, Client, Confirm, Credit, Debit, Information, Log, Notification, Password, Pay, Recently, Risk, Security, Service, User, Urgent)
- Difference between sender and reply address
- Difference between sender and message ID domain
- Presence of image links
- Presence of abnormal ports in URL
- Number of "%", "@", and "." in URL
- Physiological features (Keyword categories: Negative, anxiety, indignation, sadness, comprehension, hesitation, certainty, repression, trust)

Falling for Phishing: An Empirical Investigation into People's Email Response Behaviors [20]

Jayatilaka et al. conducted an empirical study to research why people fall for phishing emails. In the study, eleven factors were identified having an effect on people's responses to certain emails. These factors were: (1) Sender Legitimacy, (2) Perception of links, (3) Need for validation, (4) Familiarity of the email title and body, (5) Professionalism of the email title and body, (6) Emotional attachment with emails, (7) Perceived likelihood of receiving an email, (8) Length and granularity of information, (9) Previous phishing experience, (10) Sense of security from auxiliary security content, (11) Individual habits.

Creative Persuasion: A Study on Adversarial Behaviors and Strategies in Phishing Attacks [21]

In Rajivan & Gonzalez's paper, a study was conducted regarding end-user and adversary behavior in regards to phishing emails. The study was two-fold, where the first part focused on the adversary and what strategies were employed to create phishing emails, while the second part focused on the end users and their examination of the emails including determining if it was a phishing mail or not.

The results of the study showed that the following strategies proved to be successful when conducting a phishing attack: (1) Send Notifications, (2) Use an authoritative tone, (3) Pretend to be a friend, (4) Express shared interests, (5) Communicate failure.

Further, inefficient strategies were also identified. These included: (1) Offering deals, (2) Selling illegal materials, (3) Using a positive tone. While strategies such as Offering to help or communicating a deadline were inconclusive in regards to the efficiency if the strategy.

Chapter 4

Methodology

4.1 Information Collection

4.1.1 Literature Collection

As a part of the information collection, relevant literature was gathered and analyzed for related work and commonly listed concepts. The main purpose of this collection phase was to gain an understanding of the areas within phishing that have been studied, including the main focus points, methods and findings. By doing this, we are also able to see the areas that are less covered and in need of further research.

To collect as credible literature as possible, scientific search engines such as Google's search engine for academic literature, Google Scholar[1], and NTNU's digital university library, Oria[2], were used. The literature found in these engines was checked in the Norwegian Register for Scientific Journals, Series and Publishers (Kanalregister)[3] to determine the trustworthiness and quality of their publishers. Further information collection for this project, outside of the literature analysis, utilized Google's search engine as well. When Google's search engine was used, an emphasis on source criticism was made. This includes evaluating the source and comparing them to other sources.

The literature that was found to be interesting and relevant for this project can be found in Chapters 2 (Theory) and 3 (Relevant and Related Work).

[1] Goolge Scholar: `https://scholar.google.com/`

[2] Oria: `https://bibsys-almaprimo.hosted.exlibrisgroup.com/primo-explore/search?vid=NTNU_UB`

[3] Kanalregister: `https://kanalregister.hkdir.no/publiseringskanaler/Forside.action?request_locale=en`

4.1.2 Data Collection

The data collection phase involved defining the properties that would be collected and then gathering this information from the phishing emails that can be seen in today's landscape. The properties were both based on prior literature, as well as taken directly from this prior literature. The data collection was achieved through the utilization of a mail reporting system where users could report suspicious mail, as well as a ticketing system were reports could be made. Registered tickets and analysis of traffic towards phishing sites were utilized to determine what people fall for.

The collection and subsequent analysis was done manually. An alternative method would be to develop a tool that could automate these processes so that more phishing emails could be analyzed. An automated tool would lead to a more in-depth view of the phishing landscape, but it was determined not to be feasible due to time constraints and limited resources for the project.

The collection of phishing mails brings up a challenge towards the validity and reliability of the results. The first identified shortcoming is that this collection depends on the users themselves reporting in suspicious mail. Therefore, the results of the collection will only be based on the reported mails, as we cannot access each and every person's inbox. Due to this, it was important to collect a large quantity of reported emails over a longer period of time.

Another flaw that can be identified, is spam filters. Spam filters will prevent certain phishing mails from reaching the recipient. Therefore, these emails will not have the opportunity to be reported and analyzed in this study.

Lastly, the determination of what people fall for is highly dependent on either the user reporting the incident or accessing / downloading phishing content on networks where the logs can be accessed. Due to this, no crisp numbers will be presented for this part of the analysis.

4.2 Data Analysis

The analysis of the collected data utilized the information gathered from the Data Collection phase. The analysis phase was divided into two parts; one analyzing the data based on the properties defined in the Data Collection, and one analyzing the data based on features found in prior literature. The predefined properties consisted of Content, Target, Method, and Impersonation. For each of these properties, their underlying categories were defined. These properties were analyzed both on their own and in relation to each other.

For the analysis of the prior literature features, only a subset of the phishing mails

were analyzed. This was due to the nature of the identified features being aimed mostly towards phishing mails utilizing the method of URL phishing. This is discussed more in detail in Chapter 6.

It was decided to angle the prior literature features part to focus on how well these features apply to the selected phishing corpus of the project. Another potential angle was to run a test to see how good the individual paper's detection features/algorithms were, and order them accordingly. This was decided against as the papers vary in degree of main focus and are based on corpora partially different from what this study will analyze.

Chapter 5

Data Collection

5.1 Phishing Emails

The collection of data utilized a mail reporting system called MailRisk[1], as well as a ticketing system in order to collect relevant phishing emails. The mail reporting system allows users to mark received emails as suspicious and send them in for analysis. When an email is reported, relevant information such as the message body, URLs, attachments, and meta-data is extracted and made available. These emails, based either on signatures or manual analysis, are classified into eight different categories. The categories are: Safe, Spam, Suspicious, Scam, Phishing, Virus, Targeted, and Dangerous. For the scope of this research, the five latter categories were utilized to collect the phishing emails. As for the ticketing system, the tickets classified as scam and phishing were used in this collection.

The collection of phishing emails spanned from February 3rd until March 31st. However, since the systems used for collection provide historic data, phishing emails from the period January 1st til March 31th were collected. This provided a collection span of three months and a total of 1502 phishing emails were collected.

Due to the scope of the task and a limited time frame, the collection had to be done manually in order to correctly gather the information from the phishing emails. Based on prior literature, it was decided to focus the collection on the following properties: Main content of the mail, Date, Links, Attachments, Target, Method, and Impersonation.

Content - The content of the mail concerns the main essence of the phishing mail. This involves what the mail is about and how it is presented including subject, keywords, and layout.

Date - The date is the timestamp of when the phishing mail was received. This

[1] MailRisk:https://securepractice.co/engage

was collected to investigate and keep track of surges from specific emails at specific time points.

Links - The links refer to the URLs and domains linked to in the mail. Information such as number of links, unique domains, IP-addresses, length, content, and obfuscation were collected.

Attachments - Attachments refers to the documents included in the mail. Information including type, content, and number of attachments was collected.

Target - Target is a category used to specifically state what the malicious actor wants to lure out of the victim. The various Target subcategories includes: Credentials (The victims login information such as username and password), Credit Card Details (Information on the credit card that can be utilized to withdraw money or subscribe to unwanted services), Money (Direct transfers or gift card purchases for the malicious actors), Infect (Deploy malicious code onto the victim's device), Personal Identifiable Information - PII (Information such as full name, address, P.O. box and Social Security Number), and Invoice information (Information about their invoices).

Method - The Method category specifies the next technique used to complete the phishing attack after the initial email is opened. This includes linkage to a site where the phishing is performed (URL/HTML), malicious documents (Attachment), Fraudulent invoices, or simply communication over mail.

Linkage to a phishing site can be done through a direct URL in the mail, but also through an HTML document. The HTML document may contain code to redirect the user to a specified web site or redirect user input from the local document to an external site. Because of this, it was chosen to keep HTML documents as a separate category from Attachments as they serve a purpose deviating from the average properties of other attachments. We will go more into detail about the HTML method in Chapter 8.

Impersonation - The Impersonation category specifies what or whom the malicious actor pretends the emails is from. This can be from the victim's organization such as a colleague or a department, an external person such as a business partner or a random entity, a known organization such as Microsoft, Netflix, or Telenor, or the victim them self.

As these emails vary in terms of presentation and content, some boundaries and guidelines has to be set in order to reliably categorize the impersonation. When determining if the email appears to be from within the organization, the sender domain, signature name, and display name are key indicators to be viewed. If the domain is or looks like the recipient's domain, the mail will fall under this

Impersonation category. Like wise, if the display name or signature is that of an employee within said organization or a generic department such as HR and Support, the same category will apply. All entities that cannot be directly linked to an internal party or a known organization, will fall into the External category. For the mail to be categorized as a known organization, the mail has to either contain the organization in its display name, have a look-alike sender address of the firm, or contain content such as the organization's logo, name in signature, or directly mentioning the organization it self. Lastly, the Self category encompasses all emails where the display name is that of the recipient, the address looks like or is (spoofed) the recipient, or the message is signed as the recipient.

5.2 Successful Phish

To determine the successful phishing emails, the collection was dependent on the users them self reporting in the incidents. This causes a challenge to the validity of the results, as not everyone will report such incidents. We also have the possibility of analyzing the network traffic towards the sites where the URL/HTML method has been utilized. However, this again presents challenges for the validity of the results. Firstly, we are only able to see traffic towards the sites if the user is on the networks within our scope. We are not able to detect such traffic if the user is on their home network or uses their mobile network. Secondly, we will only be able to see if the site was accessed, not if any actual information was provided by the user. In it self, clicking on the link does not equal a successful phish as long as their password (for instance) was not put in on the site. Because of these challenges, this study will not provide crisp numbers in regard to what people fall for.

The successful phishings will in this research encompass phishing attacks where the Target has been achieved. As with phishing emails that utilize the URL method with the target of credentials, the user has to actually write and submit their credentials on the phishing site for it to be considered successful. Only clicking on the link and accessing the site, is not considered a successful phish. The same applies to malicious documents as well. If the malicious code lies within a macro, the user must have enabled macros in order for the attempt to be considered successful. Only opening the document (although potentially dangerous in it self) will not be considered a successful phish if it requires the enabling of further properties.

Chapter 6

Findings

The findings are based on the collected data discussed in Chapter 5. In total, 1502 phishing emails were collected and data from these were extracted for analysis. The following chapter will showcase these various phishing emails with corresponding statistics based on the aforementioned categories.

6.1 Content

The content of the mail provides details about the essence of the mail and is easily identified by the subject and body of the mail. From the 1502 phishing emails analyzed, 33 different content categories were identified. These are emails that utilize similar appeals in order to conduct a successful phishing attack. Figure 6.1 displays these content categories in a descending order.

The top five content categories are explained more in detail in the following paragraphs, while a full description of all the categories can be found in Appendix A.

CEO Scam - Gift Cards (198)
This category includes phishing attempts in which the sender pretends to be a higher-ranking official, such as a manager or a director, and asks the user to buy gift cards for them and send them over. These are often very simple emails with little to none content, except for the subject line. Figure 6.2 shows an example of this.

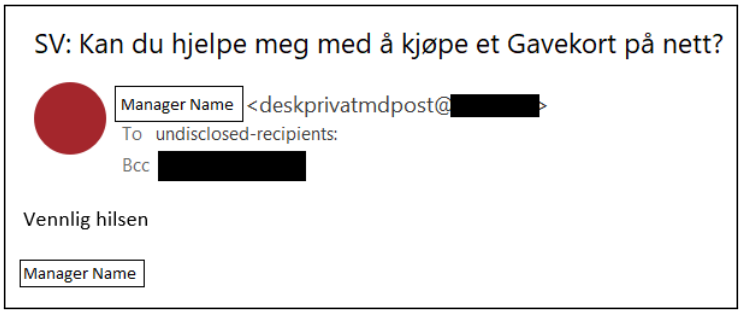


Figure 6.2: CEO Scam - Gift Card - Example

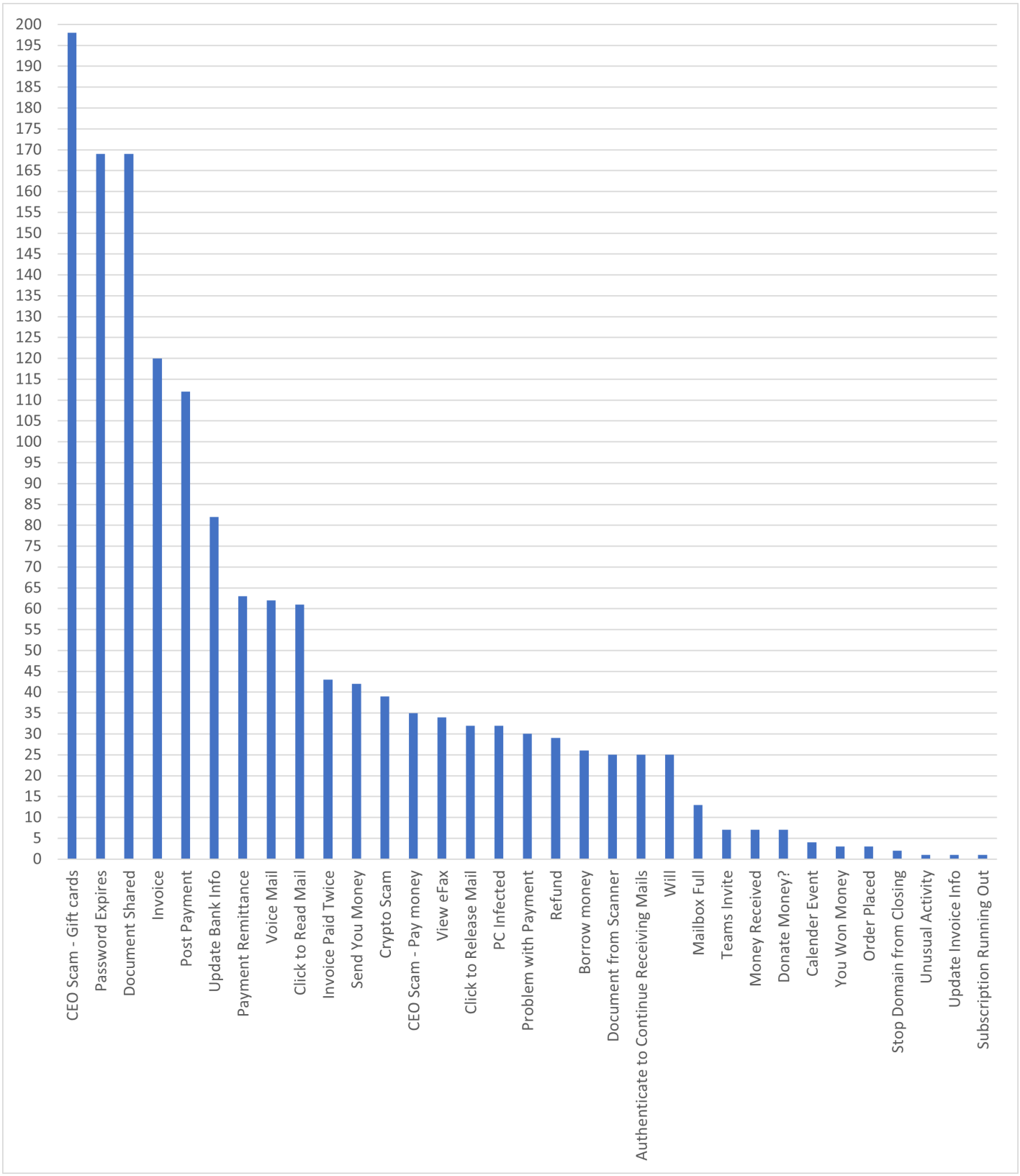


Figure 6.1: Overview - Content

Password Expires (169)
This category encompasses all phishing emails that notify the user, alerting them that their password has/is about to expire and that they can update or keep their old password by clicking a link and providing their "old" credentials. Figure 6.3 is an example of this type of mail.

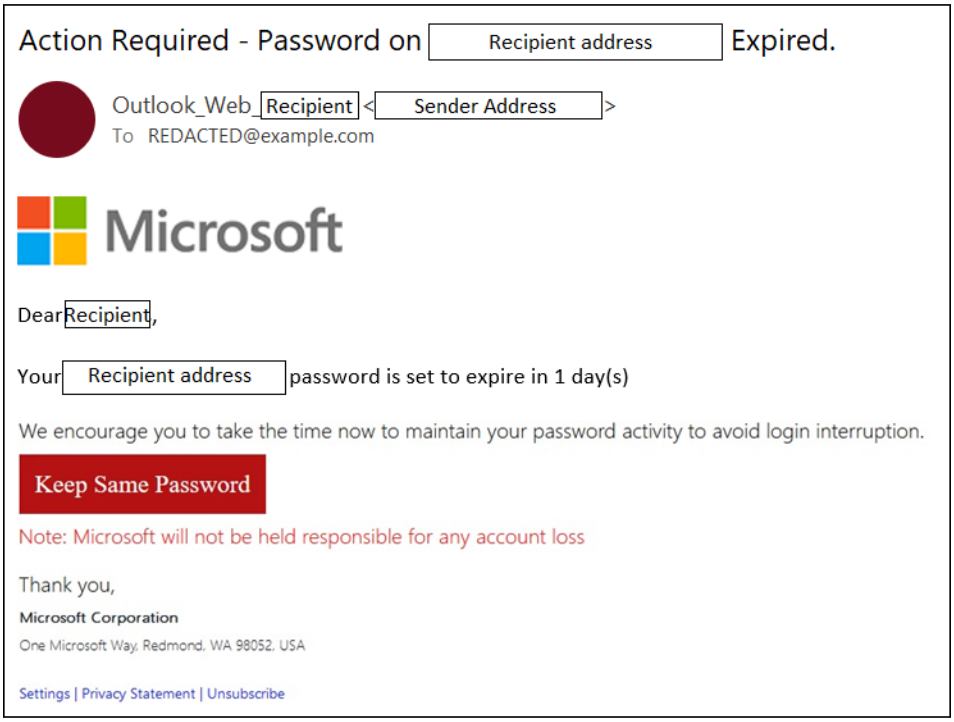


Figure 6.3: Password Expires - Example

Document Shared (169)
The "Document Shared" category are the emails where the user is tempted to click a link or open an attachment in order to view the contents of a document that has been shared with them, such as shown in Figure 6.4.

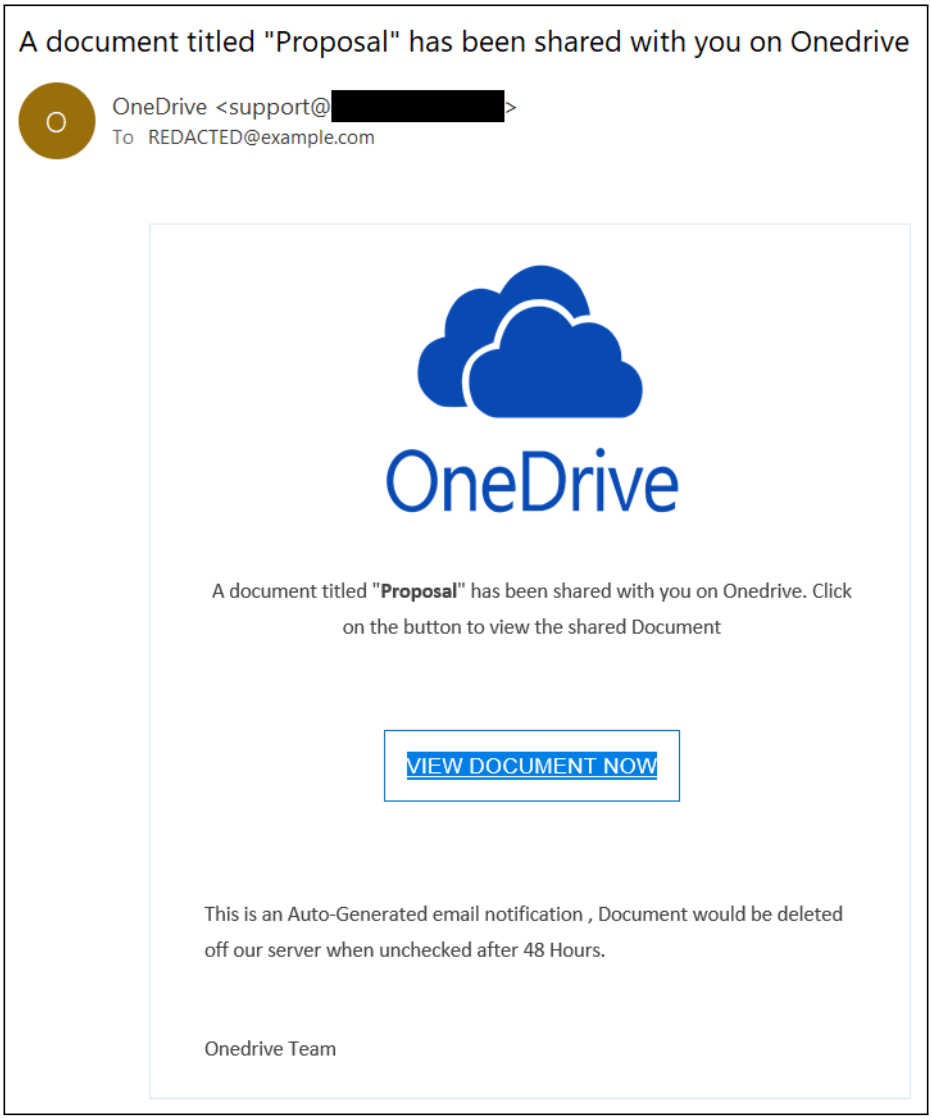


Figure 6.4: Document Shared - Example

Invoice (120)
The "Invoice" category covers a broad array of different methods and targets; however, the main factor that binds them is that they lure the user with a supposed invoice. This invoice could be behind a fake authentication page where the user is prompted for a password, or in an attached document. This attached document can again contain either malicious code or a fraudulent invoice. An example of this is seen in Figure 6.5.

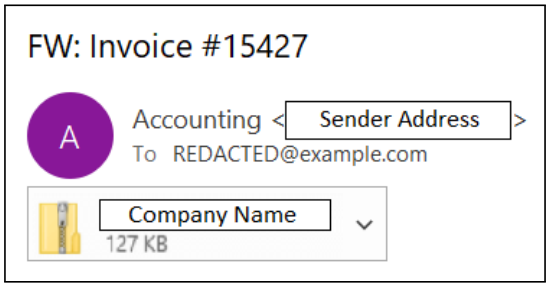


Figure 6.5: Invoice - Example

Post Payment (112)
The emails under this category tries to lure the user into paying a fee in order for their package to be delivered. Figure 6.6 shows an example of this type of mail.

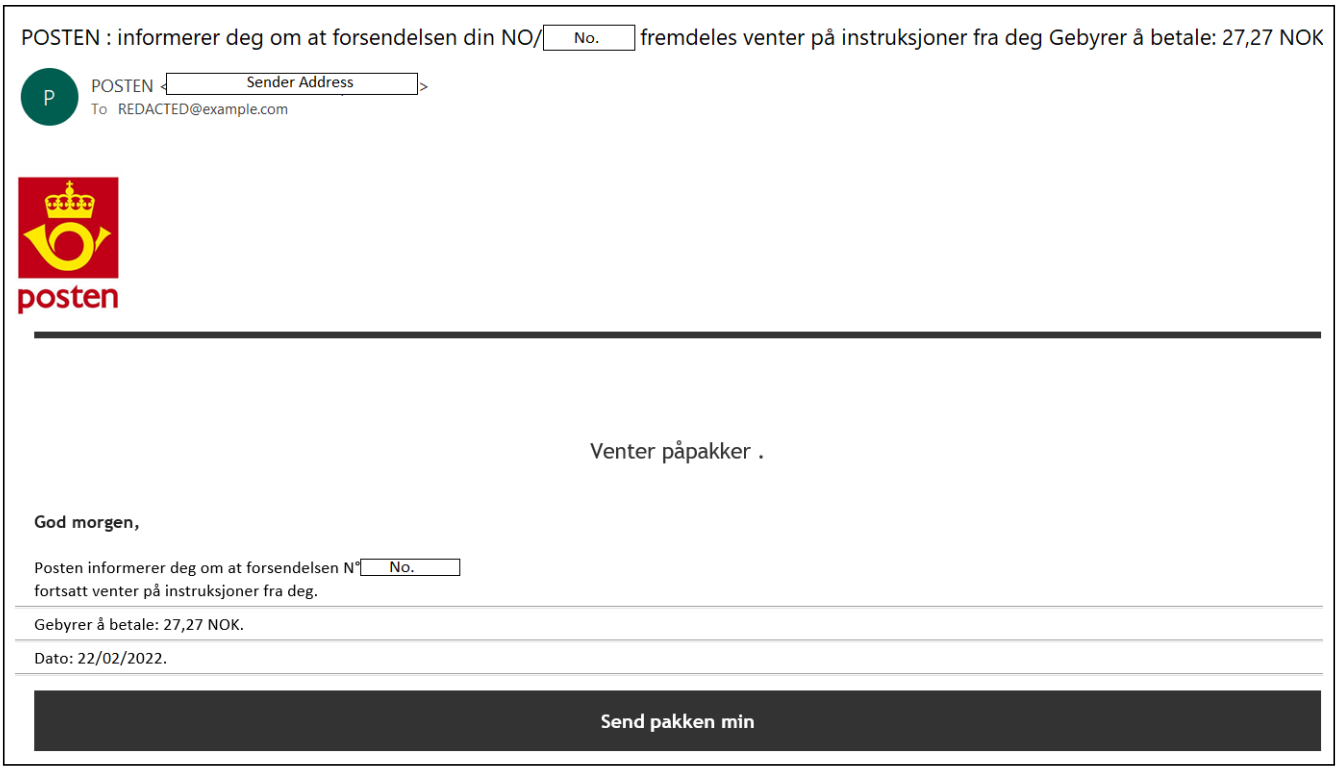


Figure 6.6: Post Payment - Example

As can be seen in the chart, the top most active phishing category is the CEO Scam involving Gift Cards. This category includes 198 emails and accounts for approximately 13% of all the phishing emails in the given time period. Following a bit further behind with 169 emails each, the "Password Expires" and "Document Shared" categories appear, accounting for 11.25% each. The "Invoice" and "Post

Payment" are next on the chart with "Update Bank Info" in suit, before the chart starts to even more out for the remaining categories.

The top five categories account for more than half of the collected emails, with a percentage of 51.1. This shows a clear trend of what the phishers use in terms of content when launching their attacks.

6.2 Target

Based on prior literature as well as an initial view of the current phishing emails, the six target categories Credentials, Money, Credit Card Details, Infect, PII, and Invoice Information were identified. The target categories are based on what the phishers tries to acquire from a successful phishing attack. From the 1502 phishing emails, 740 targeted credentials, 438 targeted money, 298 targeted credit card details, 21 tried to infect the user's device, 4 targeted PII, and one phishing mail tried to acquire the invoice information of the user's organization.

Target	Count
Credentials	740
Money	438
Credit Card Details	298
Infect	21
PII	4
Invoice Information	1

Table 6.1: Overview - Target

Again, we are able to see a clear trend in what is most actively targeted in this current phishing landscape. Credentials, with a count of 740, represent nearly half of the phishing emails that were collected (49.2%). Money is second with 29.1%, Credit Card Details are third with 19.8%, Infect fourth with 1.4%, and PII and Invoice Information with ∼0.26% and ∼0.07%.

It is an interesting viewpoint to look at the connections between the targets and the content categories. Figure 6.7 visualizes this connection through a sankey diagram.

Viewing the flows from the diagram, we can see that most of the content categories have a singular relationship with a target, such as all the "Password Expires" emails target the user's credentials, and all the "Update Bank Info" emails target credit card details. There are however a couple of content categories that vary a bit more in terms of targets, especially the "Invoice" category. The "Invoice" category have instances in four of the six target categories, including Credentials,

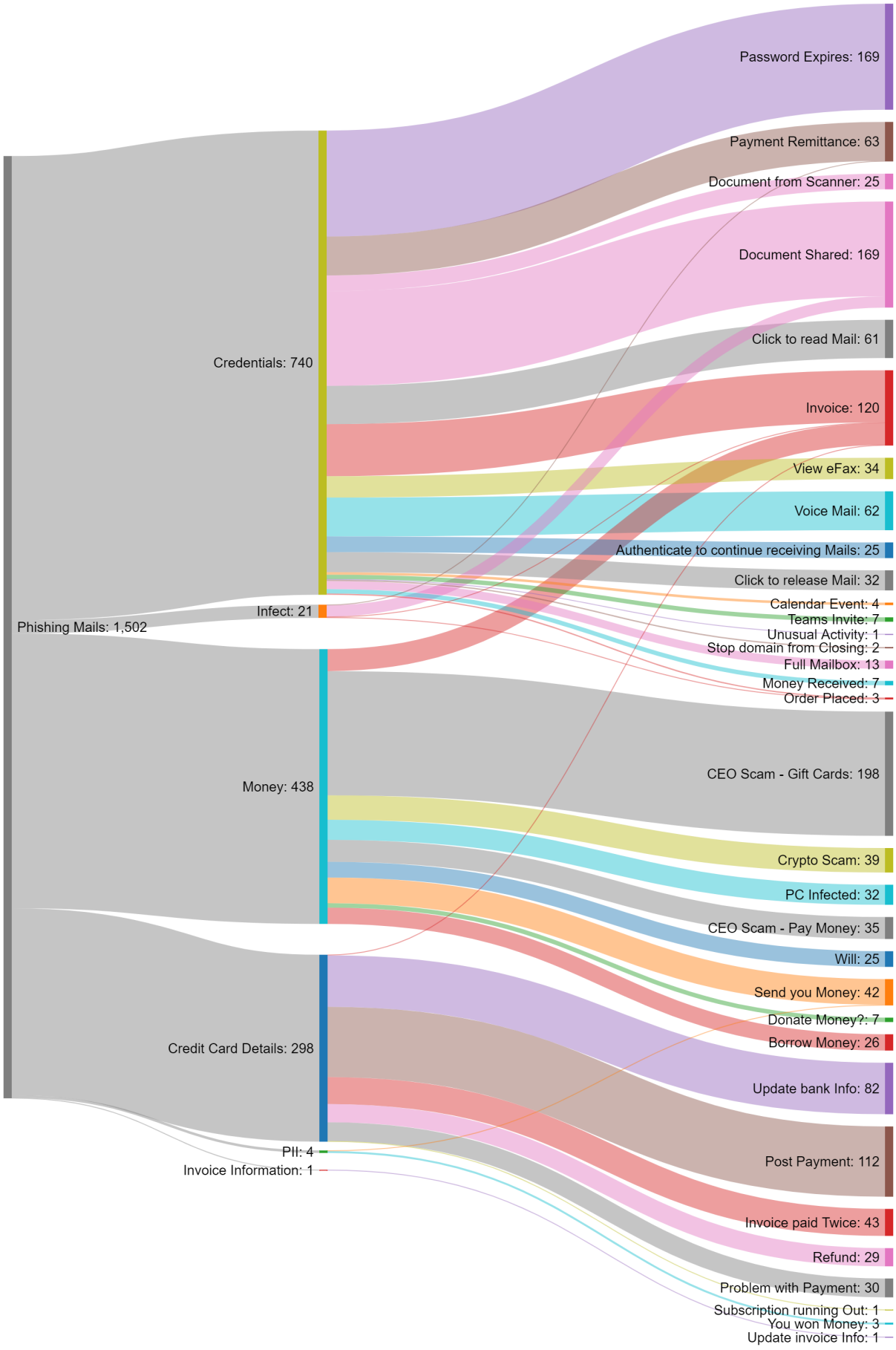


Figure 6.7: Target to Content

Infect, Money, and Credit Card Details. This shows a great variation in what the content category is usable for in the eyes of the phisher.

We are also able to depict that the Credential category is connected to the most variety of content categories, with 17 out of the 33 content categories.

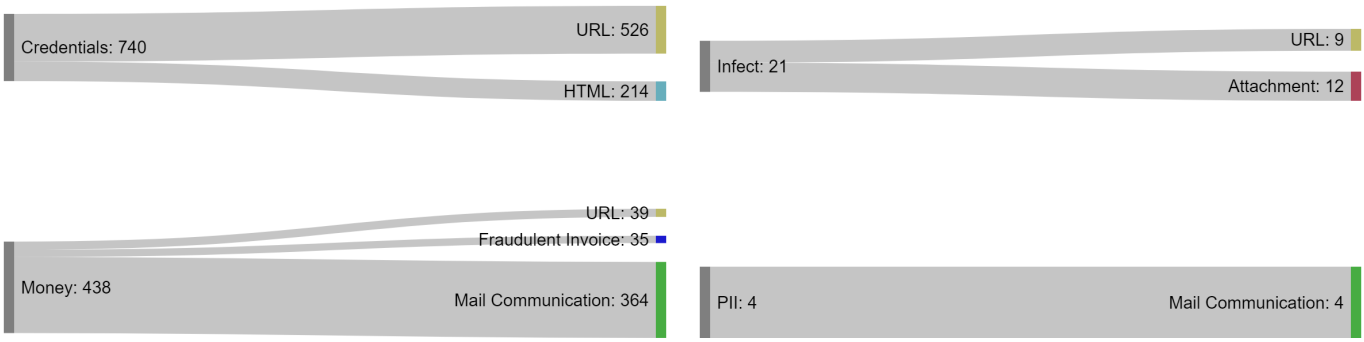
6.3 Method

During the collection, five different methods were identified out of the 1502 phishing emails. These five were: URL, HTML, (Malicious) Attachment, Fraudulent Invoice, and Mail Communication. Table 6.2 summarizes the distribution of the methods.

Method	Count	Percentage
URL	872	58%
Mail Communication	369	24.5%
HTML	214	14.2%
Fraudulent Invoice	35	2.33%
Attachment	12	0.80%

Table 6.2: Overview - Method

Further, we can look at the relationship between Method and Target. Figure 6.8 visualizes how the various methods are distributed for each and every target. Inspecting the figure, we can see that three of the Target categories, Credit Card Details, PII, and Invoice Information, all have a singular relationship towards their method. The phishing emails targeting Credit Card Details were only observed using the URL method, while emails targeting PII and Invoice Information only used Mail Communication. From this, it is also evident that each Target category, with the exception of "Infect", have a favorable method of choice. "Credentials" heavily uses URLs, "Money" prefers the usage of Mail Communication, and so forth.



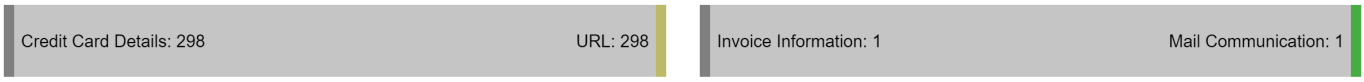


Figure 6.8: Target to Method

6.4 Impersonation

The Impersonation category focuses on whom or where the phishing email pretends to be from. As described earlier, this category consists of emails appearing to be from within the organization (Internal), from an external person or random external entity (External), A known business/organization (Organization), or from the person themself (Self). Table 6.3 summarizes the Impersonation distribution from the collected set of phishing emails.

Impersonation	Count	Percentage
Internal	644	42.8%
Organization	439	29.2%
External	416	27.7%
Self	3	0.2%

Table 6.3: Overview - Impersonation

What is interesting to look at in this case is the impersonation of known organizations. In total, there were 14 unique organizations impersonated in the various phishing emails. These included: Microsoft (143), Sparebank1 (82), PostNord (77), Telenor (46), Posten (27), Skatteetaten (22), Telia (19), FedEx (8), Helsenorge (7), Netflix (3), Hafslund (2), Spotify (1), PayPal (1), MacAfee (1). Figure 6.9 visualises this data in a corresponding pie chart.

As can be seen, Microsoft is a heavily impersonated brand in our email collection. However, the numbers presented are not 100% representative, since the Microsoft brand has been utilized much more. This is due to the fact that a fair amount of the phishing emails were presented as sent from Microsoft on behalf of the internal organization. Based on the criteria specified in Chapter 5 these could both fit under Internal and Organization. However, as they did contain references to the internal organization, they were categorized as such.

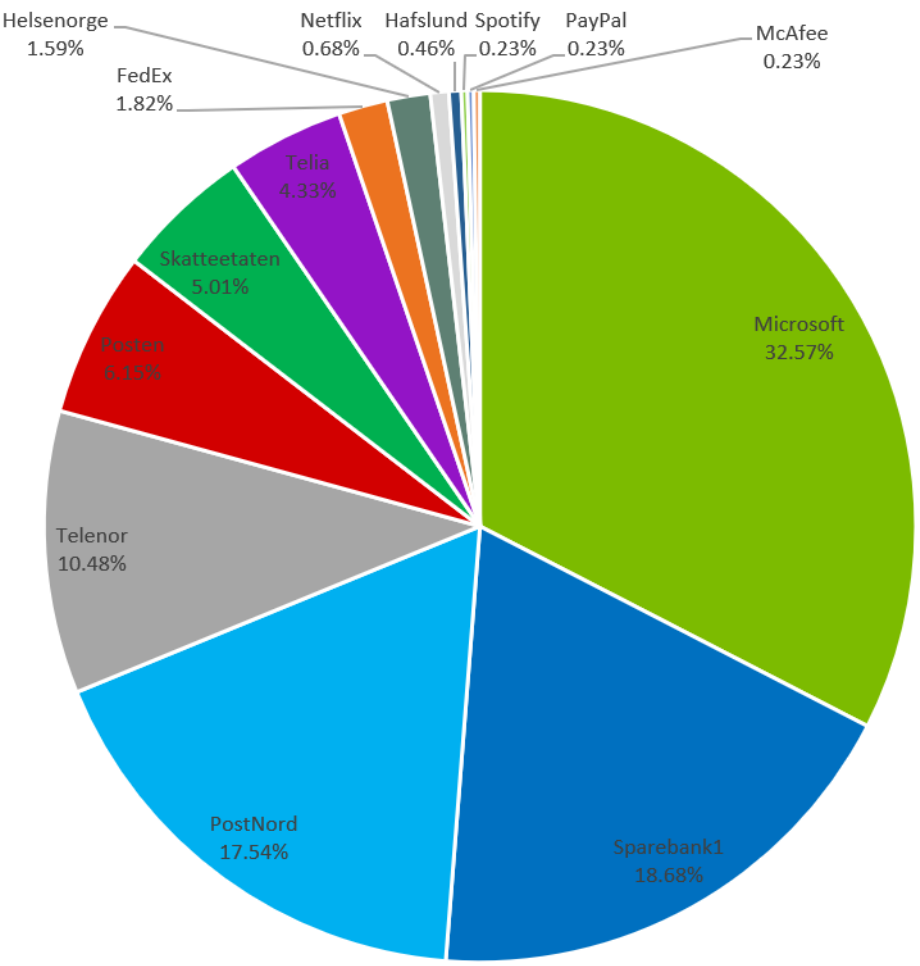


Figure 6.9: Distribution of Organizations

6.5 Time Distribution

A property noted down during the collection phase was the time stamps of the phishing emails. By analysing this, we are able to see how the various content categories are distributed throughout our collection window. Figures 6.10, 6.11, 6.12, and 6.13 display a heat map for the content categories CEO Scam - Gift Card, Password Expires, Update Bank Info, and Refund.

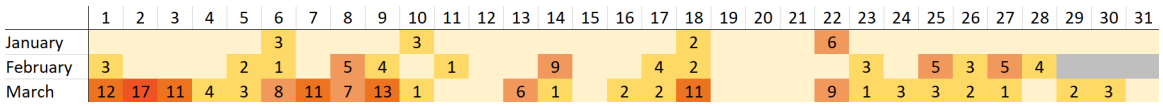


Figure 6.10: CEO Scam - Gift Card Heat Map

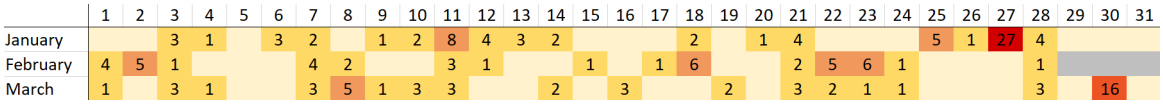


Figure 6.11: Password Expires Heat Map

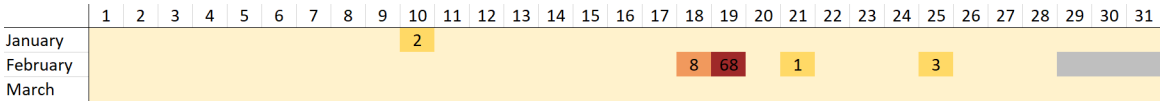


Figure 6.12: Update Bank Info Heat Map

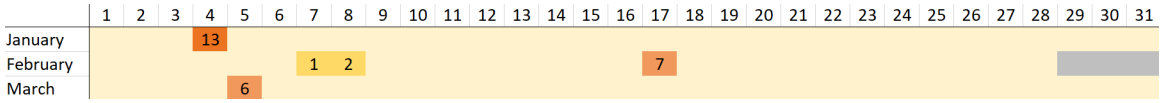


Figure 6.13: Refund Heat Map

The heat map depicting the CEO Scam - Gift Card content category showcases a surge in events towards the last third of the timeline. There were a total of 133 instances in the month of March, which accounts for 67% of all the occurrences. January saw only 14 events, which equals a percentage of 7.

Figure 6.11, which depicts the occurrences of the Password Expires, has the most even distribution for each of the months, with 73, 43, 53 instances in each month, respectively. This distribution is most similar to the other top content categories that are not depicted in any of these heat maps (Document Shared, Invoice, Post Payment).

Figures 6.12 and 6.13 showcases examples of content categories that have a low distribution of events, but gain numbers by event surges. This is especially evident in Update Bank Info's heat map, where 82.9% of the events occurred during the same day.

An interesting perspective to consider is the distribution of emails throughout the week. From the heat maps, we are able to sort the occurrences based on entries per day:

- Monday: 55
- Tuesday: 91
- Wednesday: 80
- Thursday: 74
- Friday: 62
- Saturday: 94
- Sunday: 22

Based on this distribution, we are able to see that most occurrences happen during the mid-week work days Tuesday, Wednesday, and Thursday. There is, as can be seen, an exception to this with Saturday which has the highest count in the presented heat maps. However, 68 out of these 94 are tied to one surge of events in one of the content categories.

Chapter 7

Prior Research

7.1 Features

When viewing the properties of phishing emails identified in prior work (Chapter 3), one can see that they almost exclusively base their detection on the fact that the email has linked content in them. Only a total of seven features, out of 26, were not based on the presence of a link. As can be seen from the Method statistics, URLs account for most of the phishing emails form the collected mails, however, still just 58%. This suggests that the method of identification will not be useful for the remaining 42%. Because of this, the following section will only use the phishing emails included in the URL Method category.

It is important to note that this section won't be a test of how good the detection methods and algorithms from the prior literature are. This section will solely use their identified features to see how they apply to this project's corpus and if there are any potential useful trends or features present that were not depicted in said literature.

Due to limited automatization and mostly manual analysis, the corpus is sized down for this segment. A total of 200 URL phishing emails were taken apart and compared against the features identified in previous research.

The prior research containing features for detecting a phishing mail included Jain & Gupta's "A survey of phishing attack techniques, defence mechanisms and open research challenges" [14], Akinyelu & Adewumi's "Classification of phishing Email Using Random Forest Machine Learning Technique" [17], Li et al's "Detection method of phishing email based on persuasion principle" [18], and Yang et al's "Phishing Email Detection Based in Hybrid Features" [19].

Table 7.1 showcases the count of matches from the collected features that could be answered with yes/no. Column 3-6 states whether the given literature had depicted the following features (Green = Yes, Red = No). For the features that contained lists of words, only one count was given per email if a match occurred.

		Literature			
Features	Count	Jain & Gupta	Akinyelu & Adewumi	Li	Yang
IP in URL	0				
Abnormal port in URL	0				
"@" in URL	29				
"%" in URL	21				
"-" in hostname	36				
HTTPS	139				
Brand name in URL	17				
Redirect URL	109				
href - LINK text difference	197				
LINK text words (1)*	35				
LINK text words (2)**	24				
HTML format	200				
JavaScript	0				
Image links	118				
Word in content (1)***	165				
Word in content (2)****	173				
Fwd in subject	18				
Sender - linked domain difference	200				
Sender - reply address difference	43				
Sender - Message ID difference	51				

Table 7.1: Features

*LINK text words (1) includes: Link, Click, Login, Update, Here

** LINK text words (2) includes: Click, Here, Debit

*** Word in content (1) includes the word groups: Update, Confirm, User, Customer, Client, Suspend, Restrict, Hold, Verify, Account, Notif, Login, Username, Password, Click, Log, SSN, Social Security, Secur, Inconvinien

**** Word in content (2) includes: Account, Access, Bank, Client, Confirm, Credit, Debit, Information, Log, Notification, Password, Pay, Recently, Risk, Security, Service, User, Urgent.

IP in URL, Abnormal port in URL, and JavaScript had all zero occurrences in the analyzed corpus. A few of the links did contain ports, these however were normalized web ports such as 443 and 80. JavaScript had no entries as well, this is mainly due to the fact that emails with such code are filtered out before they can reach the recipient. JavaScript is present in a portion of the HTML attachments analyzed, but is not included in the score as it is determined to be out of scope for this analysis.

During he counting of entries for the word list for both LINK text and content, there were identified words from said lists that were more common than the others

from the same list. From the Link Text lists the words "Click" and "Here" accounted for most of the entries often appearing together, while "Link", "Update", and "Login" were almost non existent. From the content word lists, words/word groups such as "Password", "Notif", "Confirm", "Update", "Click", "Bank", "Pay" and "Account" had a prominet presence, while "Log", "SSN", "Social Security", "Recently" and "Risk" were words that had close to zero occurrences.

The reminding six features, No. of dots in URL, URL length, Top Level Domain, No. of domains in URL, No. of links in mail, and No. of domains in mail, could not be answered with a yes/no and is therefore presented separately. It is important to note that for the No. of dots in URL, URL length, and No. of domains in URL features, only the URL leading to the phishing site was measured.

No of dots in URL - The findings from this collection shows that the most common number of dots in the analyzed phishing URLs was 2, with a count of 76/200. Following at second was 3 with 58/200, 1 with 49/200, and lastly, 4 with 17/200. No URLs had more than three subdomains.

No. of dots in URL	Count
2	76
3	58
1	49
4	17

Table 7.2: Dots in URL

URL length & No. of domains in URL - The URL length of the emails fluctuated a lot, with the shortest being 24 characters and the longest being 1116 characters. 78 of the URLs were shorter than 100 characters. In these URLs, there were never more than two domains present. 163 of these only contained one domain.

URL Length	Count
0 - 99	78
100 - 199	64
200 - 299	24
300 - 399	12
400 - 499	9
500 - 599	4
600 - 699	4
700 - 799	2
800 - 899	0
900 - 999	0
>= 1000	3

Table 7.3: URL Length

Domains in URL	Count
1	163
2	37

Table 7.4: Domains in URL

No. of URLs and Domains Linked to - As for the No. of URLs, 89 contained only one link, 49 contained two links, 23 contained 3 links, 21 contained 4, 8 contained 5, 5 contained six and 5 contained 7, 8, 11, 12, 15 links each. While the most common amount of linked domains present in the analyzed emails were 1 with a count of 134. Following, 52 had 2 domains, 11 had 3, and 3 had 4 domains linked.

No. of links in Mail	Count
1	89
2	49
3	23
4	21
5	8
6	5
7	1
8	1
11	1
12	1
15	1

Table 7.5: No. of links in Mail

Domains in URL	Count
1	132
2	52
3	11
4	3

Table 7.6: Domains Linked to

Top level domains - There were in total 20 unique domains used in the analyzed phishing URLs. The following domains and their associated count were identified: .com (98), .net (22), .br (15), .id (10), .app (9), .in (8), .lu (5), .np (4), .de (4), .co (4), .no (4), .co.uk (3), .co.ke (3), .cl (3), .pl (2), .mx (2), .dk (2), .ng (1), .au (1).

Based on these metrics, some statements can be made about the features in relation to this project's phishing corpus.

- All the emails from this collection were HTML formatted, and none of the phishing domains were that of the sender domain.
- Few of the emails stated the whole URL in the text as most utilized the Link Text function of HTML.
- Quite few of the phishing pages utilizes HTTPS.
- The words/word groups from the word lists used in content detection had a high count, however the count was dominated by a few selected word/ word groups including: Password, Notif, Confirm, Update, Click, Bank, Pay, and Account.
- The words from the word lists used in LINK text detection had a significantly lower occurrence rate. Again being dominated by a few words including

Click and Here.
- URL redirects and image links were observed in a bit over half of the phishing emails.
- Special characters in URLs/hostname, as well as brand name in URL and Fwd in subject, did not occur often in the analyzed phishing emails.
- Most of the phishing URLs had one subdomain, however quite a few had two and zero as well.
- Most phishing URLs have less than 200 characters and contains only one domain.
- Most phishing emails have 1 to 2 URLs, and 1 domain linked to.
- The top level domain of .com is most utilized in phishing URLs.

7.2 What People Fall For

Based on user reports and network traffic analysis, several phishing emails with which people have fallen for have been identified. These emails cover a wide variety of the identified content categories, as well as the methods, targets, and impersonation categories. The following section shows an excerpt of the collected emails that people have fallen for. The Excerpt contains types of emails that have been observed fallen for multiple times.

7.2.1 Excerpt

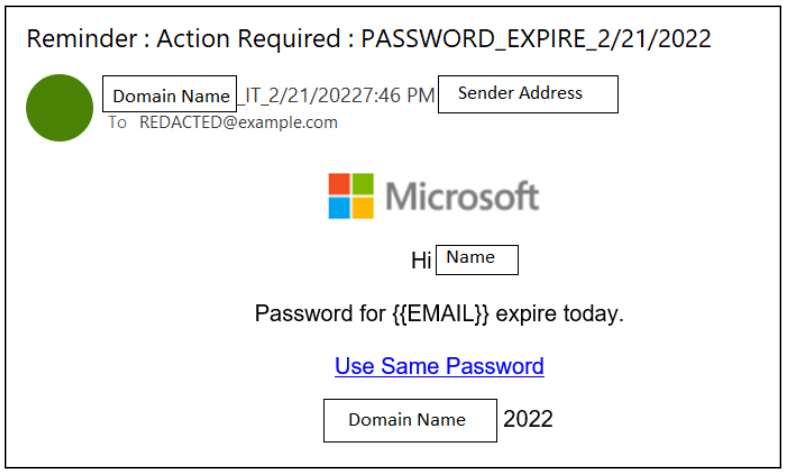


Figure 7.1: Fell for - Password Expires

Figure 7.1 shows an email from the Password Expires category that a user fell for. The mail mentioned the user's organisation's domain throughout the mail, as well as giving a mention to the user and their email address. The sender address had no relation to the domain, user, or content of the mail.

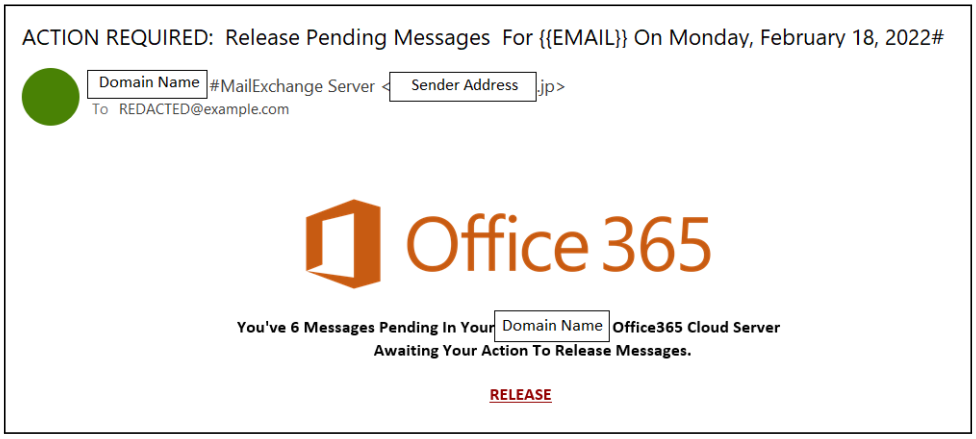


Figure 7.2: Fell for - Click to Release Mail

Figure 7.2 displays a mail in the Click to Release category. This mail linked to a fake Microsoft login site asking for the user's credentials, which were provided. This mail also mentions the user's organization's domain a few times in the mail. The sender address had no relation to the domain, the user or the content of the mail.

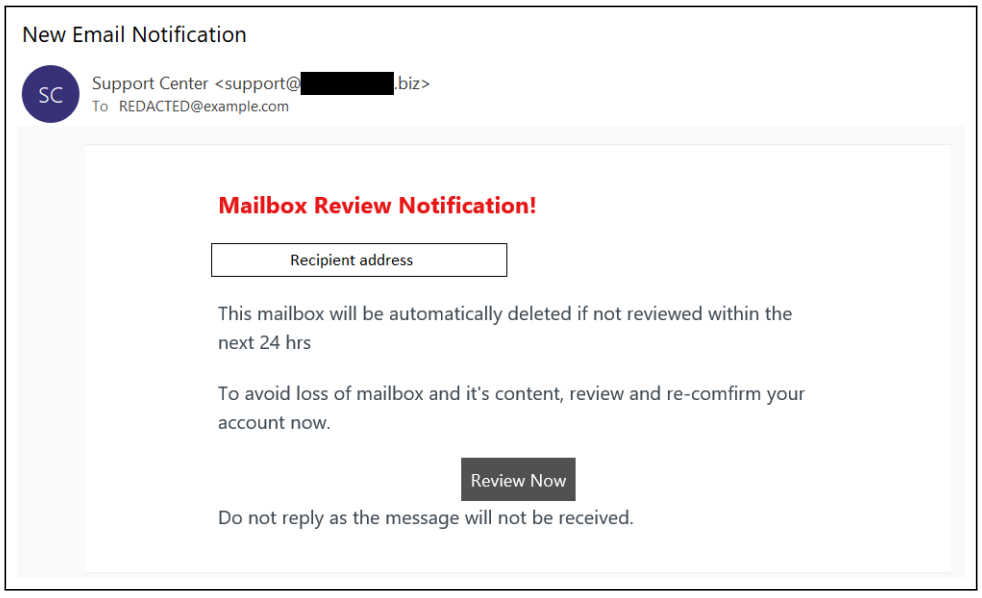


Figure 7.3: Fell for - Authenticate to Continue Receiving Mails

Figure 7.3 is an example of an email in the Authenticate to Continue Receiving Mails category that a person fell for. This email has a partially relevant sender address with "support" as the username. However, the domain of the sender had no relevance to the recipient. This mail utilizes words such as Notification, Con-

firm, and Account presented earlier as common phishing words from the previous literature.

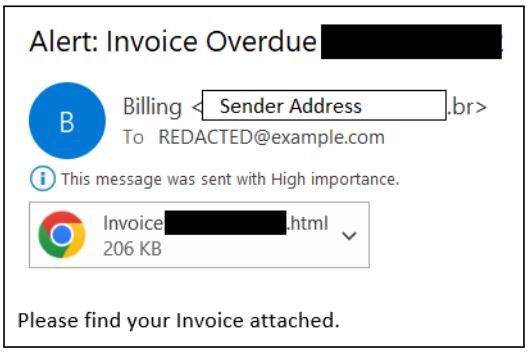


Figure 7.4: Fell for - Invoice

Figure 7.4 showcases an Invoice email that a user was tricked by. This is a case of the HTML method, where the HTML attachment led to a fake sign-in page asking the user for their credentials to view the invoice. This mail had little content expect its title and attached HTML document, with only a short sentence in its body. A notable feature in this case is that the mail was sent with a high-importance tag. The sender address has no relation to the domain, user, or content, and has no direct mention of the user's name/address.

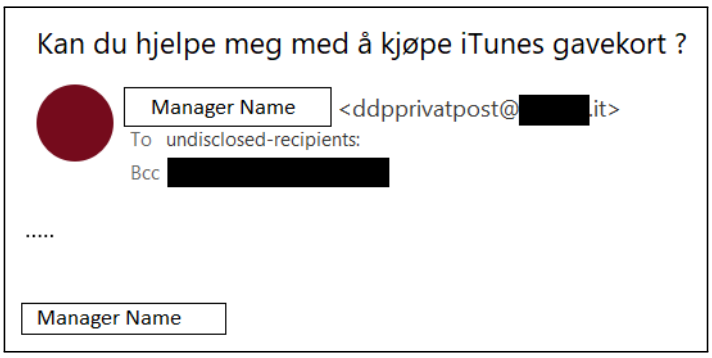


Figure 7.5: Fell for - CEO Scam - Gift Card

Figure 7.5 is one of the many examples of a CEO Gift Card Scam that a user has fallen for. This email, again, has little content except its title. This is a phishing that utilizes the Mail Communication method, which means that the user has to reply in order to receive further instructions. Not included in this picture is the following thread in which the actor specifies what they want, and the user buying and sending over pictures of the gift cards. The mail uses the user's manager's name while having a private email address.

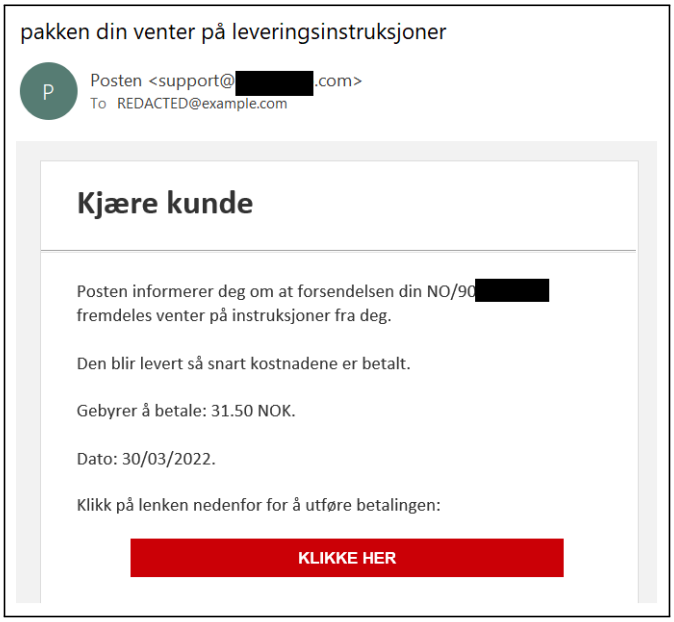


Figure 7.6: Fell for - Post Payment

Lastly, figure 7.6 showcases a Post Payment phishing mail where the recipient was lured into giving up their credit card information. The mail impersonates Posten and displays a seemingly well designed mail body.

7.2.2 Aspects of a successful phish

During the literature analysis, it was identified several papers that mentioned the various aspects of an email that made it a "good" phishing mail, including what emotions/incentives to manipulate. The persuasion tactics such as Authority, Scarcity, and Reciprocation were stated in multiple of the texts, as well as the emotions of fear, greed, and anger. Other technical factors such as Sender Legitimacy, Professionalism, Likelihood of Receiving such mail, and Previous Phishing Experience were some of the factors identified in a successful phishing attack. Based on the factors identified in the previous literature, this section will determine which of the factors are present in the successful phishing emails identified.

As some of these factors requires more thorough knowledge of the recipient, they will not be evaluated in this section.

Password Expires - The "Password Expires" mail is a notification mail that can be seen trying to create a sense of urgency with the recipient, stating that their password will expire today unless the recipient performs an action. As the domain name of the company is stated both in the display name and the closing line of

the mail, it might be perceived as a more legitimate sender, i.e sender legitimacy.

Ties to the "sense of duty" aspect can also be drawn, as an employee may feel responsible for their own account and, in turn, making sure that their password is up to date.

Click to Release Mail - This phishing mail plays into the curiosity incentive of the recipient. It lures the user into clicking the "RELEASE" link by appearing to be a notification mail from the company's exchange server. On a similar note as the "Password Expires" mail, the domain name of the company is present both in the display name and in the message body, leading to a possible perception of sender legitimacy.

Authenticate to Continue Receiving Mails - This phishing mail is again a notification mail, and on par with the "Password Expires" mail, it appeals to the sense of urgency and duty. However, this mail brings in the Fear element as well, threatening with mailbox deletion if action is not taken. If an account's password expires one can easily change it after, however, once a mailbox is deleted it can be difficult and/or time consuming to fully restore it.

Invoice - The "Invoice" mail aims to manipulate both the sense of urgency and curiosity. Sense of urgency comes from both the supposed invoice being overdue as well as the message being sent with high importance. Curiosity is created by the limited information provided in the mail, only referencing the phishing attachment.

CEO Scam - Gift Card - This phishing mail utilizes the full name of the recipient's manager, increasing the perception of sender legitimacy. Authority, and in turn, sense of duty are invoked due to the email appearing to be from an authoritative entity.

Post Payment - Lastly, the "Post Payment" phishing mail can be seen trying to manipulate either the Curiosity or the Perceived Likelihood of Receiving the Email incentives. With this particular case, the latter applied, however, both have been observed in the collection of phishing emails that people have fallen for. The Curiosity incentive applies when the recipient does not expect a package, but is interested in seeing what the package could be. This could further give rise to the Award Incentive and Reciprocity motivators due to having received a "gift" and the only action required is to pay a seemingly small monetary fee. For the other case, recipients are ticked into thinking that the mail received is related to an already expected package, increasing the chances of a successful phish.

7.2.3 Summary

As showcased with these excerpts, both the Urgency and Curiosity incentives have a presence in many of the collected emails, with a exception of the "CEO Scam - Gift Card" phish that had neither. Sender Legitimacy, as well, has a prominent presence in the emails by utilizing a familiar named person in the display name or the domain of the company in the message body. It was, however, also observed that the actual sender address did not need to be relevant at all for the phishing to be successful, showing that the display name and message body has been the focus of the recipients.

Another factor worth mentioning is the length and granularity of the information in the viewed phishing emails. The emails are short in length, and limited information is given in the mail body and subject. It can be viewed as just enough information is given to get the recipients attention and appeal to one or more of the aforementioned incentives/features, while avoiding to give more information than necessary that could raise suspicions.

Chapter 8

Discussion

8.1 Top Categories

Through the analysis of the collected phishing mails, various categories and their popularity were identified. A question yet to be answered is why these instances appear more than the others. Although answering this question requires insight into both the phishers themselves as well as the recipients, some potential factors can be deduced based on the collected data.

The top most occurring content category is the CEO Scam - Gift Card instance. There are two main factors that could be presented as desirable from the phishers point of view. Number one is that these phishing attempts do not include any links or attachments, making it so that the phishers don't have to set up and maintain phishing sites or maliciously coded documents, and the mail filters cannot base their detection on any link/attachment properties. Second, these attempts have an easy payout. Once the gift card is bought and the codes sent over, there is little to do to reverse the purchase from the victims point of view and the phish can be considered done.

The next two categories are Password Expired and Document Shared. Both of these deviate a fair bit from what could be considered desirable in the instance of the CEO Gift Card Scam. They lack the simplicity of the planning and clean-up phase, however, one could argue that the execution and infiltration phase is "easier" for the phishers as they do not have to communicate any further with the target after the initial mail is sent. This could also be viewed as the reason as to why the URL method accounts for more than half of the observed phishing emails. The reasoning as to why these two content categories are so prominent could be because they seem familiar/are expected. In a work environment, people do share digital documents in a work flow and IT often notifies the user once their password is about to run out. Familiarity can also be used to explain why Credentials is the top Target category. Typing in your password to access sites is again something that can be considered a part of the normal work flow and isn't always second guessed.

These are, as mentioned, assumptions made based on the observed data. As we have not interviewed or questioned the recipients or have insight into the actors sending these emails, no direct conclusions can be made. This is a subject that requires more research before anything can be said for certain.

8.2 HTML Attachments

The usage of HTML documents as a method to conduct phishing attacks was not mentioned in detail in any of the prior literature analyzed. It is however present in over 14% of the analysed phishing emails. A search on blog posts and news articles shows that the usage of HTML documents isn't something completely new, as indicated by a blog post from Webroot [22] and an article from Bleeping Computer [23]. However, they are mostly discussed when the HTML documents are used to download malicious code onto the computer, and not in relation to credential phishing. In the analyzed phishing emails, HTML documents were only used as a means to gain the user's credentials and not to further download malicious documents. This method seems to have become popular in the recent years [24] [25] [26].

The question is why is this method utilized as opposed to just linking to a website? This answer may vary depending on how the method is utilized. During the analysis of the emails two different HTML methods were observed. One method was simply linking directly to the phishing site in the attachment so that when the user opened the document, the phishing site was loaded. Figure 8.1 shows an example source code for this method, and Figure 8.2 shows how the landing page may look like.

Figure 8.1: HTML Linked - Source Code

The other method is to embed the document with a web page and have the user host the page on their own device rather than a page directly on the Internet. This page is then linked to the attackers site, so that when the user enters their credentials, they are sent to the attackers. Figure 8.3 presents a portion of the HTML source code used for this, and Figure 8.4 shows the hosted page on the target's machine.

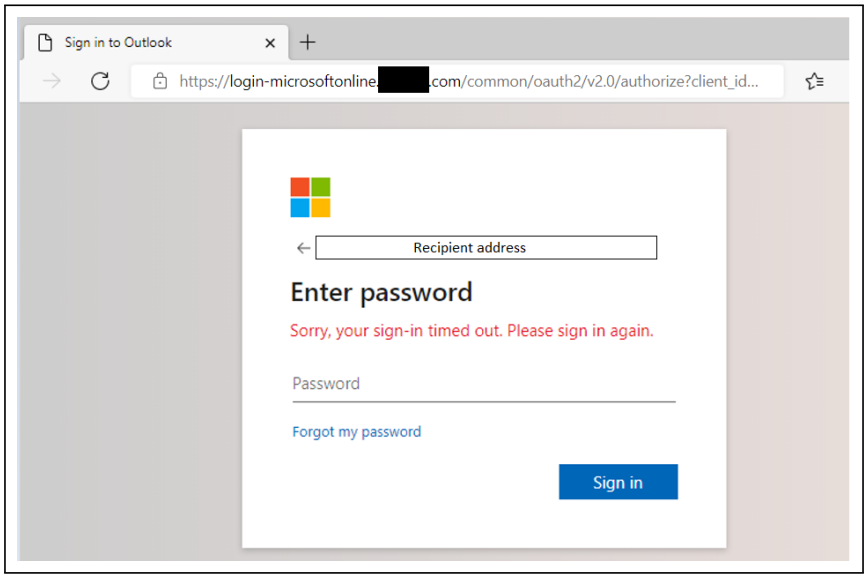


Figure 8.2: HTML Linked - Landing Page

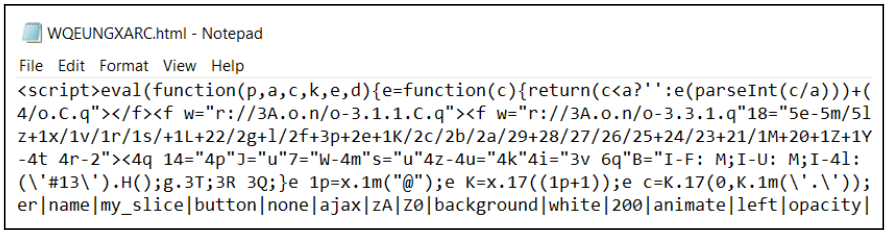


Figure 8.3: HTML Local - Source Code Snippet

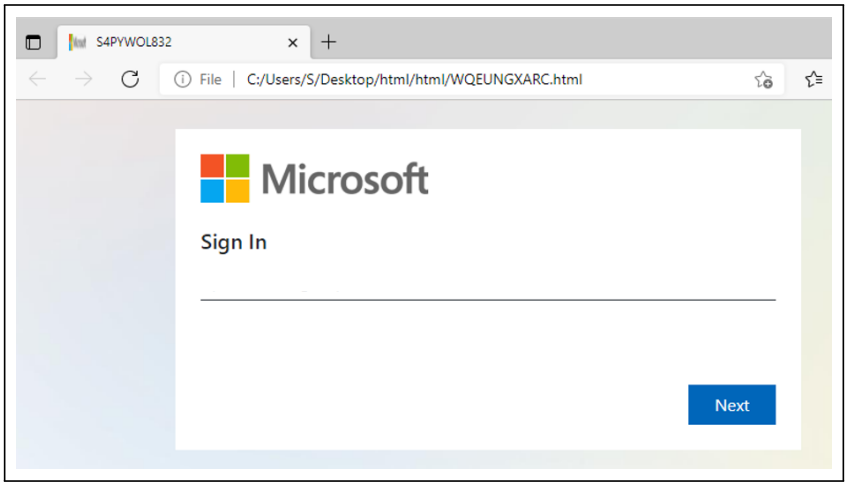


Figure 8.4: HTML Local - Landing Page

The reasoning for these approaches lies in their ability to further evade security measures detecting malicious emails. The first layer of evasion is putting the malicious content, such as the link, in an attachment, as opposed to directly in the mail body. The next layer of evasion may lie in the ability to obfuscate the content of the attachment by encoding it (not shown in the above examples). The second instance provides a third layer of evasion as well. As the site isn't hosted on the internet, engines cannot detect the site and flag it. They can of course flag the site which the page communicates with, but many automated detection methods rely on visual queues in order to categorize a site.

8.3 Targeted Brands

When analyzing pre-existing literature, one of the papers [14] displayed an overview of the top ten targeted brands in phishing emails. From this list, Apple was stated as the number one targeted brand, with Microsoft OWA and Google Drive second and third. As this paper referenced a 2017 report, it would be suitable to find a report from from 2021/2022 and see how they compare to the 2017 report, but most importantly the findings of this project (Chapter 6, Section 4).

Software and Cyber Security company CheckPoint released in January of 2022 their Q4 Brand Phishing Report for 2021 [27]. This report details the most recent top 10 targeted brands.

1. DHL (23%)
2. Microsoft (20%)
3. WhatsApp (11%)
4. Google (10%)
5. LinkedIn (8%)
6. Amazon (4%)
7. FedEx (3%)
8. Roblox (3%)
9. PayPal (2%)
10. Apple (2%)

As can be seen, some of the brands depicted in this list are present in the findings of this project, although it is only three brands. DHL, which was at the top of the Q4 report, was not present in any of the analyzed emails. Microsoft however, is high on all of the lists. This may show a consistency for the Microsoft brand in regards to phishing emails, as well as proving to be a valuable target for attackers over time and in different environments. Different environments could potentially be the reasoning for why the Q4 report and the project findings deviates in such a high degree. The phishing emails, as specified earlier, are collected from mostly users in Norwegian based companies. A majority of the users being Norwegian

can clearly be seen from the impersonated companies. Eight out of the 14 companies identified are either Norwegian owned or Scandinavian based companies. Because of this, we cannot make any direct conclusions towards the changes from CheckPoint's 2021 Q4 findings and the 2022 Q1 findings.

8.4 Additional Detection Features

Chapter 5 focused on features observed with phishing emails, as well as what categorizes as a "good" phishing mail. These features and aspects were based on elements and conclusions from prior literature. During this analysis, it was shown that many of the prior literature features were present in this paper's corpus, while others showed little to none relevance. As phishing is constantly evolving, so should the detection methods of them. Because of this, it is relevant to look at the features observed in the collected phishing emails that were not discussed in the prior literature, but was observed in a great deal of the collected emails.

The first feature that was observed throughout the phishing emails were the use of "Action Required" in the subject of the mail. The usage of this word combination was observed regularly in the phishing emails that targeted the password of the recipient. Subjects such as "Action Required: your password expires on February 2, 2022", "Action Required - Payroll Request", and "Password Action Required for [NAME]" were some examples of this usage. "Alert" was also a highly used word present in the phishing corpus. This was not included in the word lists from the prior literature, but showed a great presence in the analysed emails.

Another observation regarding the subject of these phishing emails was the usage of the date of sending in the subject line, such as "Comment: 1 mention 1/3/2022", "Payment copy sent on Monday, March 14, 2:27:56 AM", and "07 January, 2022 Your salary increment approved". Although this was present in quite a few of the analyzed phishing emails, the format of these may present itself as a challenge if they were to be used as detection features. As can be seen, they tend to use different placements and formats for the date presentation. It is, however, a feature worth considering.

Depicted in section 3 of Chapter 6 and in section 2 of this chapter, was the usage of HTML attachments to conduct the phishing attacks. Although URLs is the most utilized method of phishing, they are also highly used in harmless emails as well. When looking at at the "Safe" section in MailRisk (the tool used to collect many of the phishing emails), only one in the given period had an HTML attachment. This indicates that the presence of an HTML attachment should raise some alarms whether the mail is a phishing mail. Due to this, HTML attachments can be proposed as a potential detection feature.

The final detection feature derived from this analysis is "Mismatch between Display Name and Sender Address". This one is derived from the "CEO Scam" phishing emails and is a bit special as it cannot rely on just information from the phishing mail alone. The intention of this detection method is to, for every time a display name matches a list of employee names, the sender address is compared to that employee's registered address(es). If there is no match, the email is flagged. This, of course, requires synchronization between the detection filter and the entity storing employee information. The detection method could be susceptible to errors as multiple people can have the same name; however, the benefit can greatly outweigh its drawbacks.

8.5 Validity of Data

The data retrieved from the heat maps brings up an important challenge to the validity of the results depicted. This is especially evident in the heat map showcasing the distribution of events for the "Update Bank Info" content category. As shown in Chapter 6's Figure 6.1 - "Overview - Content", the "Update Bank Info" category is the 6th highest ranking content category, with 82 occurrences in total. The data shows that this particular strain of phishing emails has a high occurrence rate, and should potentially be a focus point during awareness trainings and when creating phishing signature features. However, section 5 of that same chapter presents data that challenges the perception of this information.

Section 6.5 presents heat maps showing the daily distribution of occurrences for various content categories, including the "Update Bank Info" category. From this distribution, we are able to see that close to 83% of its occurrences are tied to one single day, and that no events have been recorded the last month of the collection window. This information also heavily skews the week-day distribution presented. From this data alone, one could come to the conclusion that Saturdays sees the highest phishing content received, while in reality it is one of the more quieter days.

This additional data is able to change the initial perception of the firstly presented findings, and without it, wrong conclusions and decisions could be made. This is not to say that one should not consider the "Update Bank Info" phishing category in trainings, signatures and other managerial tasks, however, it shows that one should always consider seeing the data from more perspectives before coming to any closing conclusions and decisions. As in this case, the additional perspective may indicate that other categories deserve higher prioritization than the "Update Bank Info" category of phishing emails.

Chapter 9

Conclusion

In this paper, a collection and extended analysis of 1502 phishing emails has been carried out, including an examination of a subset of these phishing emails that people have fallen for. The collection of emails was a representation of the phishing emails observed in Q1 of 2022 in a domain dominated mainly by Norwegians. The analysis of the emails was two-fold, where the first part focused on the properties chosen and described specifically for this paper. This included the categories Content, Target, Method, and Impersonation, while the Date of Delivery property was collected as well. The second part revolved around prior literature regarding the phishing subject. This part based its analysis on previous literature on the subject, where identified phishing features from said literature were compared to the observations made in the collected phishing emails.

First and foremost, we are able to see that the phishing phenomenon that arose in the 90's has a prominent presence in today's threat landscape, and that the essence of the attack has not changed much since the first depicted incidents.

The analysis conducted showed that 33 different content categories could be identified from the phishing corpus, where the "CEO Scam - Gift Card" category was the most observed, and, in inclusion with the following four content categories, accounted for more than half of the observed emails. On a similar note, it was observed that Credentials was the most occurring target and URLs were method of choice in a large portion of the emails, both accounting for close to half and more than half of the analysed phishing emails respectively. Although no direct conclusions can be made as to exactly why these are the top categories, assumptions can be made that tie their popularity to both the simplicity of certain phases in the phishing life cycle and the familiarity aspects of the emails. What can be said is that these findings shows a clear trend of what the phishers utilize in today's phishing landscape.

Plotting the findings of these categories together, one is able to see that most of the content categories has a singular relationship with a target, although, there

are some deviations form this such as with the Invoice content category. It is also evident that close to all of the target categories, with the exception of "Infect", has a favorable method utilized.

The findings of the Impersonation and Date of Delivery properties showed challenges and viewpoints regarding the validity and applicability of the collected data. The Impersonation property, depicting organization impersonation, shows that this aspect of the phishing landscape is distinctly different form reported data, due to the user base being mostly Norwegian. While the Date of Delivery challenges the perception of the stats in the content categories, showing how surges of events can create a false impression of what should be expected and focused on.

Regarding the features of the previous research, it was determined that they focused mainly on phishing using URLs as the method of choice. Analyzing a subset of the URL method phishing emails, it is shown that quite a few of these features matched observations from the project's phishing corpus, while some had little to no presence at all. The features "HTML Format" and "Sender - linked domain difference" appeared in all of the analyzed emails, while "IP in URL", JavaScript, and "Abnormal port in URL" were not observed at all. There were identified features observed during the inspection that were not depicted in the analyzed literature, including subject phrases/words, HTML as attachment, and sender - address difference. These features should be tested and considered added to detection algorithms to better flag potentially phishing emails.

Lastly, the analyzed successful phishing emails showed that the incentives of Urgency and Curiosity had a presence in many of the successful phishing emails, while Sender Legitimacy can be determined to have had an impact as well, but only in regard to the Display Name and Message Body.

9.1 Challenges

Throughout the study, challenges have been identified that have had an impact on the overall result of the analysis.

The first challenge is tied to the collection and analysis of the phishing emails. Due to time limitations, there was no time to create a tool capable of collecting the desired properties, thus leading to the collection having to be done manually. This put a limitation on how many emails could be collected and later analyzed. Another challenge identified during the study was the focus point of the prior literature collected for the analysis. These proved to mainly focus on phishing emails that utilized URLs as the method of phishing, while this paper focused on a wider variety of methods. This resulted in a shift of focus in Section 1 of Chapter 5 - Prior Research, to focus solely on the emails using the URL method.

The analysis of the features, specifically the ones regarding successful phishings, brought out a limitation of this study. As there was no communication with the victims during the study and little to no insight into the victims themselves, some of the features identified in the prior literature could not be used. This included features such as Previous Phishing Experience and Emotional Attachment. The analysis had to focus mainly on the data / information seen in the email and any information that had been reported by the victim prior to the study.

9.2 Future Work

The paper raises questions and challenges that require further research.

9.2.1 Investigate top categories

From the findings, it was shown that some of the categories for content, target, and method had a higher presence than the others in the same category. The discussion chapter came with arguments as to why these are the most seen categories based on interpretations and assumptions of the data. However, as this could not be concluded on without more insight into the recipients and phishers, further research should be done on the subject in order to determine why these categories have such a grand presence. Being able to communicate with the recipients can also further identify more aspects regarding why some people fall for certain phishing emails.

9.2.2 Repeated analyses

The scope of this research encompassed phishing emails observed in Q1 of 2022, however, to gain a broader view of the phishing landscape, similar studies should be conducted for the rest of the quarters. This will show how/if the trends shift throughout the years and if there are any observable patterns with regard to the selected properties.

9.2.3 Automated collection tool

If this study is to be continued for additional quarters, a tool should be considered made in order to ease the collection and analysis of phishing emails. With an automated tool such as this, more emails can be collected and analyzed, resulting in an even more precise interpretation of the current phishing landscape.

9.2.4 Additional domains

As this study mainly encompasses Norwegian based users, it can be interesting to conduct a similar analysis for users based in other regions to see if there are any significant differences between the various locations. The emails in this study are collected from business email accounts as well, which means that additional

research can be done on private accounts to again identify similarities and differences in the results.

9.2.5 Test detection features

The discussion chapter points out five additional detection features for phishing emails. As this study mostly analysed emails that were categorized as phishing, a test has to be made in order to determine whether these detection features are applicable, or if they give high accounts of false positives.

Bibliography

[1] Tessian Research, 'Psychology of human error,' *Tessian*, 2020, Accessed 21.05.2022. [Online]. Available: https://www.tessian.com/research/the-psychology-of-human-error/.

[2] B. B. Gupta, A. Tewari, A. K. Jain and D. P. Agrawal, 'Fighting against phishing attacks: State of the art and future challenges,' *Neural Comput & Applic*, vol. 28, pp. 3629–3654, 2017, Accessed 09.03.2022. [Online]. Available: https://doi.org/10.1007/s00521-016-2275-y.

[3] K. Rekouche, 'Early phishing,' *arXiv preprint arXiv:1106.4692*, 2011, Accessed 09.03.2022.

[4] B. B. Gupta, N. A. G. Arachchilage and K. E. Psannis, 'Defending against phishing attacks: Taxonomy of methods, current issues and future directions,' *Telecommun Syst*, vol. 67, pp. 247–267, 2018, Accessed 09.03.2022. [Online]. Available: https://doi.org/10.1007/s11235-017-0334-z.

[5] Wikipedia contributors, *Email*, Accessed 12.03.2022, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Email.

[6] S. Gibbs, 'How did email grow from messages between academics to a global epidemic?,' 2016, Accessed 12.03.2022. [Online]. Available: https://www.theguardian.com/technology/2016/mar/07/email-ray-tomlinson-history.

[7] Wikipedia contributors, 'ILOVEYOU,' 2022, Accessed 12.03.2022. [Online]. Available: https://en.wikipedia.org/wiki/ILOVEYOU.

[8] K. McIntyre, 'The evolution of social media from 1969 to 2013: A change in competition and a trend toward complementary, niche sites,' *The journal of Social Media Society*, vol. 3, no. 2, pp. 5–25, 2014, Accessed 12.03.2022. [Online]. Available: https://thejsms.org/index.php/JSMS/article/view/89.

[9] B. Parmar, 'Protecting against spear-phishing,' *Computer Fraud & Security*, vol. 2012, no. 1, pp. 8–11, 2012. [Online]. Available: https://doi.org/10.1016/S1361-3723(12)70007-6.

[10] 'An update on our security incident,' *Twitter*, 2020, Accessed 12.03.2022. [Online]. Available: https://blog.twitter.com/en_us/topics/company/2020/an-update-on-our-security-incident.

[11] S. Gibbs, 'Facebook and google were conned out of $100m in phishing scheme,' *The Guardian*, 2017, Accessed 12.03.2022. [Online]. Available: https://www.theguardian.com/technology/2017/apr/28/facebook-google-conned-100m-phishing-scheme.

[12] R. M. Lee, M. J. Assante and T. Conway, 'Analysis of the cyber attack on the ukrainian power grid. defense use case,' *SANS-ICS, E-ISAC*, 2016, Accessed 12.03.2022. [Online]. Available: https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf.

[13] K. L. Chiew, K. S. C. Yong and C. L. Tan, 'A survey of phishing attacks: Their types, vectors and technical approaches,' *Expert Systems with Applications*, vol. 106, pp. 1–20, 2018, Accessed 09.03.2022. [Online]. Available: https://doi.org/10.1016/j.eswa.2018.03.050.

[14] A. K. Jain and B. B. Gupta, 'A survey of phishing attack techniques, defence mechanisms and open research challenges,' *Enterprise Information Systems*, vol. 0, no. 0, pp. 1–39, 2021. [Online]. Available: https://doi.org/10.1080/17517575.2021.1896786.

[15] A. N. Shaikh, A. M. Shabut and M. Hossain, 'A literature review on phishing crime, prevention review and investigation of gaps,' in *2016 10th International Conference on Software, Knowledge, Information Management Applications (SKIMA)*, 2016, pp. 9–15. [Online]. Available: https://doi.org/10.1109/SKIMA.2016.7916190.

[16] R. Alabdan, 'Phishing attacks survey: Types, vectors, and technical approaches,' *Future Internet*, vol. 12, no. 10, 2020. [Online]. Available: https://doi.org/10.3390/fi12100168.

[17] A. A. Akinyelu and A. O. Adewumi, 'Classification of phishing email using random forest machine learning technique,' *Journal of Applied Mathematics*, vol. 2014, 2014. [Online]. Available: https://doi.org/10.1155/2014/425731.

[18] X. Li, D. Zhang and B. Wu, 'Detection method of phishing email based on persuasion principle,' in *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, vol. 1, 2020, pp. 571–574. [Online]. Available: 10.1109/ITNEC48623.2020.9084766.

[19] Z. Yang, C. Qiao, W. Kan and J. Qiu, 'Phishing email detection based on hybrid features,' *IOP Conference Series: Earth and Environmental Science*, vol. 252, p. 042 051, 2019. [Online]. Available: https://doi.org/10.1088/1755-1315/252/4/042051.

[20] A. Jayatilaka, N. G. Arachchilage and M. A. Babar, 'Falling for phishing: An empirical investigation into people's email response behaviors,' 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2108.04766.

[21] P. Rajivan and C. Gonzalez, 'Creative persuasion: A study on adversarial behaviors and strategies in phishing attacks,' *Frontiers in Psychology*, vol. 9, 2018. [Online]. Available: https://doi.org/10.3389/fpsyg.2018.00135.

[22] Blog Staff, 'Beware spam with html attachments,' *Webroot*, 2010, Accessed 17.04.2022. [Online]. Available: https://www.webroot.com/blog/2010/07/20/beware-spam-with-html-attachments/.

[23] B. Toulas, 'Microsoft warns of surge in html smuggling phishing attacks,' *BleepingComputer*, 2021, Accessed 17.04.2022. [Online]. Available: https://www.bleepingcomputer.com/news/security/microsoft-warns-of-surge-in-html-smuggling-phishing-attacks/.

[24] Y. Nathaniel, 'Html attachments: The latest phishing trend targeting office 365,' *Avanan*, 2019, Accessed 17.04.2022. [Online]. Available: https://www.avanan.com/blog/phishing-trend-targeting-office-365-uses-html-attachments.

[25] L. A. Remorin, 'How html attachments and phishing are used in bec,' *TrendMicro*, 2017, Accessed 17.04.2022. [Online]. Available: https://www.trendmicro.com/en_no/research/17/g/html-attachments-phishing-used-bec-attacks.html.

[26] A. Spadafora, 'This new 'linkless' phishing scam is even tricking tech experts,' *TechRadar*, 2020, Accessed 17.04.2022. [Online]. Available: https://www.techradar.com/news/this-new-linkless-phishing-scam-is-even-tricking-tech-experts.

[27] 'Dhl replaces microsoft as most imitated brand in phishing attempts in q4 2021,' *CheckPoint*, 2022, Accessed 17.04.2022. [Online]. Available: https://blog.checkpoint.com/2022/01/17/dhl-replaces-microsoft-as-most-imitated-brand-in-phishing-attempts-in-q4-2021/.

Appendix A

Content Categories

CEO Scam - Gift Cards
This category includes phishing attempts in which the sender pretends to be a higher-ranking official, such as a manager or a director, and asks the user to buy gift cards for them and send them over. These are often very simple emails with little to none content, except for the subject line. Figure 6.2 shows an example of this.

Password Expires
This category encompasses all phishing emails that notify the user, alerting them that their password has/is about to expire and that they can update or keep their old password by clicking a link and providing their "old" credentials. Figure 6.3 is an example of this type of mail.

Document Shared
The "Document Shared" category are the emails where the user is tempted to click a link or open an attachment in order to view the contents of a document that has been shared with them, such as shown in Figure 6.4.

Invoice
The "Invoice" category covers a broad array of different methods and targets; however, the main factor that binds them is that they lure the user with a supposed invoice. This invoice could be behind a fake authentication page where the user is prompted for a password, or in an attached document. This attached document can again contain either malicious code or a fraudulent invoice. An example of this is seen in Figure 6.5.

Post Payment
The emails under this category tries to lure the user into paying a fee in order for their package to be delivered. Figure 6.6 shows an example of this type of mail.

Update Bank Info

The Update Bank Info emails concerns mails that aims to steal credit card details by tricking the recipient into thinking they have to provide credit card info in order for their account not to be disabled.

Payment Remittance
These emails tricks the recipient into clicking a link/opening an attachment thinking that they have received a payment remittance.

Voice Mail
The Voice Mail phishing attempts masks their emails as notification mails, stating that the recipient has received an online voice mail, often through Teams or WhatsApp.

Click to Read Mail
This category encompasses all phishing emails that tries to lure the recipient into clicking a link/opening an attachment stating that they have unopened or pending emails.

Invoice Paid Twice
This category consists of emails that impersonates services such as Telenor, Telia, and the like, notifying the recipient that their latest invoice was paid twice and that they will receive reimbursement. To receive the money, the recipient is asked to provide credit card details.

Send You Money
The Send You Money category consists of all phishing emails that presents the recipient with an opportunity to receive large amounts of money. Often from "Rich Foreigners" that "Have too much money" or are feeling generous. In order to receive said money, the recipient either has to pay a smaller sum to prove their identity or provide PII.

Crypto Scam
These phishing emails tries to lure the recipient into buying crypto schemes from illegitimate sites that presents themselves as "Get you Rich Fast" schemes.

CEO Scam - Pay money
Similar to the CEO Scam - Gift card, these phishing mails impersonates a high-ranking officer trying to lure the recipients into paying money to the phishers accounts. These are often masked as legitimate business payments that needs to be performed urgently.

View eFax
Similar to the Click to Read Mail, these mails lures the recipient into clicking a link/attachment thinking they have received an eFax.

Click to Release Mail
This category consists of the phishing emails that states the recipient has to perform an action (providing their credentials) in order to release pending emails from quarantine.

PC Infected
These phishing emails tries to scare the user into paying the phishers money (often through Bitcoin) by stating that they have infected their computer and acquired images/videos of them watching adult content, and will release these if payment is not received.

Problem with Payment
This category consists of emails that notifies the recipient that a problem with their latest payment has occurred and that they have to pay again. These masks as legitimate services such as Netflix and Spotify.

Refund
The Refund category encompasses the phishing emails who states that the recipient is eligible for a refund. Most of these mails impersonates Skatteetaten saying that the user has over paid their taxes and may get back this over payment. To do so, the user has to provide their credit card details.

Borrow Money
This category of phishing emails comprise of mails coming from a "Rich Foreigner" needing to borrow money because their assets currently are unavailable. If they can borrow this money, they will pay more back in return once their funds are accessible.

Document from Scanner
The Document from Scanner category are phishing mails that notifies the recipient that their document has been scanned and is available for download.

Authenticate to Continue Receiving Mails
This category of phishing emails consists of emails stating that the user needs to perform a re-authentication in order to continue using their email account.

Will
These phishing emails notifies the recipient that they have received a monetary amount from a recently deceased person. This can a be a formerly unknown family member, for whom the recipient is the only living person that this person is related to. The deceased could also be a non-family member who generously willed their money to the recipient. In order to receive the will, the recipient has to pay a smaller sum to prove their identity.

Mailbox Full
This line of phishing mails are notification emails sent to the recipient stating that their mailbox is/will soon be full and that they have to perform an action (providing their password) in order to increase the mailbox size.

Teams Invite
These phishing emails masks themselves as Teams invites (groups/external teams) in order to lure the recipient into clicking a link and providing their credentials.

Money Received
This category encompasses all phishing emails who states that the recipient has received a monetary amount, and that the details are available in the linked site/attached file.

Donate Money?
These phishing mails impersonates people in need or an organization providing humanitarian aid, asking for donations.

Calendar Event
This category of phishing emails notifies the recipient that they have a new calendar event, an that they can review this by clicking the following link.

You Won Money
These phishing mails impersonates lottery companies notifying the recipient that they have won large amounts of money. In order to receive the money, they have to provide PII.

Order Placed
These emails notifies the user that their order has successfully been placed, ans that the details can be viewed in the attached document.

Stop Domain from Closing
This category comprise of emails that alerts the recipient that their domain is about to be closed and that they have to perform an action (providing their credentials) in order to keep their domain.

Unusual Activity
These emails alerts the recipient that there has been observed unusual activity from their account. To view this information, the recipient has to re-log into their account.

Update Invoice Info
This phishing mail impersonates an internal employee asking the recipient to

provide info on selected invoices.

Subscription Running Out
This phishing mail poses as a legitimate service and alerts the recipient that their subscription is about to run out. In order to keep their subscription, the user has to provide their credit card details.