University of Twente

Master Thesis

18 June 2021

The impact of the Corona-pandemic on the business model of cybercrime

Jip Laan (s2202972)

Business Administration
Financial Management

Thesis supervisor: Dr. A. Abhishta

Wordcount: 20296

Abstract

The goal of this thesis was to understand what the impact of the Corona-pandemic has been on cybercrime. The thesis focused on a subset of cybercrime, namely phishing. It was hypothesized that cybercriminals made use of the Corona-pandemic to increase their revenue gathered from cybercrime. It was tested whether they adapted to the Corona-pandemic by adjusting their phishing-campaigns accordingly. This was done by making use of a term-frequency analysis. Unique mentions of selected keywords were plotted and analyzed how they evolved over time. In particular we tested whether there was a significant change before the Corona-pandemic and after/during the Corona-pandemic. The following categories were analyzed in this manner: “Corona-related mentions”, “Medical and protection equipment”, “Financials”, “Order and delivery scams” and “Dating, beauty and care”. We found that all the categories increased during the Corona-pandemic except for the category “Dating, beauty and care”. This category had decreased. We also found that the intensity of phishing emails increased during the Corona-pandemic. The results suggest that cybercriminals stepped up their game by both increasing the amount of phishing emails and adapted their phishing campaigns to match what was happening in the world during the Corona-pandemic.

Contents

Chapter 1 Introduction 3
Chapter 1.1 Current knowledge and research gap 4
Chapter 1.2 Goal 5
Chapter 1.3 Method, research design and data 6
Chapter 1.4 Reading guide 8
Chapter 2 Literature review 9
Chapter 2.1 Method 9
Chapter 2.2 Types of cybercrime 10
Chapter 2.3: Cybercrime as a business model 12
Chapter 2.4: DDoS, ransomware, and phishing. 14
Chapter 2.5: The impact of Corona-pandemic on intensity on cyberattacks. 16
Chapter 2.6: The impact of the Corona-pandemic on activity on dark web markets. 18
Chapter 2.7: Impact of the Corona-pandemic on business models. 19
Conclusion 20
Chapter 3 Methodology 21
Chapter 3.1 Data collection 21
Chapter 3.2 Approach 21
Cleaning and transformation of the data 23
Term frequency analysis 23
Statistical analysis 25
Chapter 4 Results 27
Corona related mentions 28
Medical and protection equipment 29
Financials 31
Order and delivery scams 33
Dating, beauty, and care 34
Chapter 5 Discussion and conclusion 37
Chapter 5.1 Discussion 37
Chapter 5.2 Main Conclusions 40
1. What is cybercrime? 40
2. What is the business model of cybercrime? 41
3. How can we empirically test the impact of the Corona-pandemic on phishing? 41
4. What are the implications of the Corona-pandemic on phishing? 42
Chapter 5.3 Limitations 43
Chapter 5.4 Future work 43
Literature 45
Appendix 47

Chapter 1 Introduction

An innovative approach has come from businesses during a crisis. A crisis can force businesses to apply their creative innovation. For example, supermarkets were invented during the great depression. Big crises can also give rise to innovation in crime. Cybercrime has become a serious threat for companies and organizations. Especially those organizations that rely heavily on digital and information infrastructure (Huang, Siegel & Madnick, 2018). Companies depend more on IT each year. Inventory managements systems, data management systems and online workspaces to name a few examples. These systems are susceptible to cyberattacks, preventing them to function as intended. This is evident in the questionnaire conducted by An & Kim (2020) where 78% of the respondents experienced a cyber-attack in the last 5 years. 31% on personal systems and 47% through work. This data was gathered however before the Corona pandemic. Cybercrime has especially seen a growth during the Corona pandemic (Ahmed, 2020). Since the corona pandemic began, the advice for companies mainly have been to operate from home as much as possible. This has resulted in much greater use of online communication and thus greater opportunity for cybercriminals.

Cybercrime is no longer committed only by highly skilled programmers (De Groot, 2019; Huang et al., 2018). Online crime has seen a transformation where cybercriminals can use Crimeware-as-a-service (CaaS) without the need of advanced skills to carry out a cyber-attack. An example is RaaS (ransomware as a service). Where one can order ransomware-tools and support for money (Alhawi, Baldwin, & Dehghantanha, 2018). This is valuable for cyber criminals as ransomware and other types of CaaS can be bought from them and, as mentioned before, used by criminals without having great technical skill (Alhawi et al., 2018). Furthermore, this allows the developers of these tools to be more anonymous as they do not carry out the cyberattacks themselves. Entire underground economies have developed surrounding this type of cybercrime and has become a multibillion-dollar industry (EUCPN, 2015). In this way, cybercrime has become a lucrative business by which criminals can make a living. Cybercrime could be considered to have an actual business model that generates value. The main difference with a regular business is that these types of activities are illegal. To give an example: cybercriminals that have control over botnets, can leverage this infrastructure to offer DDoS attacks or spamming as a service (Putman, Abhishta & Nieuwenhuis, 2018). Other cybercriminals can carry out DDoS attacks this way, without having control over the botnets themselves. An example of an attempted attack in 2020 is on Tesla (Business Insider, 2020). Cybercriminals tried to bribe an employee of Tesla to install malicious software on the Tesla network. The employee was promised one million dollars if he employed the ransomware attack. Instead of accepting the bribe, the employee reported the attempted attack. An example of a serious ransomware attack that was successful was at the University of Maastricht. The university was forced to pay 30 Bitcoins (at the time of payment these were worth around 200,000 euro). Similar attacks were employed on banks in 2018 in the Netherlands on Rabobank and ABN Amro. As a result, the banks were not able to operate for several hours. These are however examples of cyberattacks on large companies.

Due to the Corona-pandemic, governments are urging people to work from home. Therefore, increasing the number of people working from home relying more on online means of conducting business. This results in increased potential targets for cybercriminals. It is no surprise that since the Corona-pandemic attacks on individuals also have increased dramatically (NOS, 2020). Instead of using brute force, an example is the use the human element in trying to steal money called social engineering. Such an example is the recent emerging phenomenon of Whatsapp fraud where the victim is tricked into sending money to cybercriminals. The associated tools and script are for sale on darknet marketplaces and provide support in carrying out these attacks (Huang et al., 2018; Alhawi et al., 2018). The

example of the attempted ransomware attack on Tesla also partly used social engineering in trying to infect the Tesla systems.

Chapter 1.1 Current knowledge and research gap

Due to the ever-increasing threat of cybercrime, there has been a growing interest in understanding cybercrime. Especially during the Corona-pandemic cybercrime has increased dramatically (Ahmed, 2020). DDoS attacks, ransomware attacks and phishing have gained momentum due to the Corona pandemic (Interpol, 2020). The Corona-pandemic has increased the opportunity for cybercriminals. In April 2020 Interpol (2020) wrote that they expect to see a rapid growth in cybercrime due to the sudden economic and social changes. Criminals want to take advantage of the shift towards working online from home and thus transfer part of their criminal activities to the cyber. It is expected that, among other things, criminals will put more effort into online scamming and take more advantage of Crimeware-as-a-service (CaaS) due to the low cost but high potential profit (Interpol, 2020). In 2020, the Netherlands government made 1 million Euro available for companies in the Netherlands to invest in cyber-security. It is essential to understand how cybercriminals operate to invest the subsidy as efficiently as possible (Rijksoverheid, 2020). Researchers devoted their attention to trying to understand the way cybercrime is organized and conducted. Cybercrime has become an actual business model (Armin, Thompson, Ariu, Giacinto, Roli & Kijewski, 2020; Huang et al., 2017;2018; An & Kim, 2018). Transitioning from product-oriented towards service-oriented (An & Kim, 2018). A business model can be a way to analyze how a business carries out transactions and creates value (Amit & Zott, 2001). Business models of cybercrime have been studied intensively. In extending the knowledge about the business model of cybercriminals, the rise of professional cybercrime business models may become more visible for government and organizations. If cybercrime is more visible it can be fought more effectively (An & Kim, 2018). In particular, understand how the Corona-pandemic has impacted the cybercrime business model specifically emergence of services.

Huang et al. (2017;2018) wrote an extensive literature review on the business model of cybercrime as a service that reviews activities of cyber criminals. In the literature review they analyzed components of the cyberattack business via a value chain model which models how cybercrime as a business model creates value. This gives an understanding of different examples of CaaS, one of the primary business models behind cybercrime (An & Kim, 2018). Such as vulnerability analysis as a service (VaaS), ransomware as a service (RaaS), and Botnet as a service (BnaaS), to name a few. Huang et al. (2018) made distinctions between “existing”, “evolving” and “emerging” services. Some services, such as “Botnet as a service” are well established and for sale on dark marketplaces. Other services, such as “Hacker recruiting as a service” is evolving. An example of emerging services is: “Domain knowledge as a service”. Evolving services are expected to evolve into new services. Emerging services are not yet observed as services but are expected to.

The value chain model in Huang et al. (2018) gives directions on how primary activities are carried out and how the support activities of the cybercrime business model interact, to give rise to CaaS, but lacks the relationship to practice. An & Kim (2018) argues this is due to the lack of adequate data analysis approach in studying cybercrime. An & Kim (2018) tried to solve this by proposing fitting types of data-analysis for the underground economy of cybercrime. In this paper market trends in CaaS, cybercrime market dynamics and target organizations were analyzed. Their results obtained through their methods indicate the most common mentioned sector is the technology sector, followed by the content sector and finance sector. Results about the market trends in CaaS indicate that between 2008 and 2017 the most trending class in as a service was RaaS. This agrees with practice where there is large increase of ransomware attacks (Kiru & Aman, 2019).

Due to the corona pandemic business models have been affected. The impact on the

business models of legitimate businesses has been studied by Ritter & Pedersen (2020). It was found that conduct of business has moved towards online space, where physical meetings have become less frequent and virtual interaction has become the norm. This is less relevant for cybercrime as this was conducted mainly online already. However, the shift from physical to virtual for legitimate businesses has posed a great opportunity for cybercriminals. The researchers proposed six different types of business models and how they each cope with a crisis. From business models that operate better when there is a crisis or increased stress to business models that can only survive with external help (Ritter & Pedersen, 2020). From the analysis of the impact of the Corona-pandemic on the business model of cybercrime it can be derived what type of business model cybercrime is.

To my knowledge, the impact of the corona pandemic on the business model of cybercrime has not been studied. Thompson et. al. (2020) argue that Cybercrime is a subject that is still in its infancy and much can be learned from other disciplines such as business and economics. An & Kim (2018) argue much is still unknown about the business model of cybercrime. Since cybercrime is constantly changing and improving, the most recent knowledge is needed to be implemented when combating cybercrime. Organizations such as the Cambridge Cybercrime Centre and Anti-Phishing Working Group (APWG) are constantly collecting the most recent data on cybercrime to reach to this of using the most recent knowledge to combat cybercrime. This can also be derived from Huang et al. (2018) and Interpol (2020) as there are constant trend changes in CaaS and cybercrime in general. Conclusions drawn from studies assessing the business model of cybercrime were mainly classifications and extensions of existing literature and thus miss links to practice. Due to missing links to practice and lack of reliable data, (An & Kim, 2018) proposed analytical frameworks to analyze different aspects of the cybercrime economy. However, the results regarding trends and dynamics of the cybercrime-market derived from this study were from before the Corona-pandemic. Since then, business have been impacted and most likely cybercrime has changed with it. An & Kim (2018) mentioned that money spent in prevention and monitoring of cybercrime decreases the likelihood of serious more serious consequences of cybercrime. With a broader understanding of how the Corona-pandemic has impacted the cybercrime business model, cybercrime can be understood more clearly and combatted more effectively.

Chapter 1.2 Goal

The goal of this thesis is to understand what the impact of the Corona-pandemic on cybercrime has been. The business model of cybercrime is a broad subject, in this thesis we zoom in phishing, a subset of cybercrime. From here we start our approach to find evidence of the expectation of Interpol (2020) that cybercriminals are making use of Corona-related in cyberattacks. How did cybercriminals react to the pandemic? And if they reacted: what has changed during the pandemic?

To address these issues, we formalize the following research questions. Our main research question is: What is the impact of the Corona-pandemic on cybercrime? To answer this question, we divide it into sub-questions:

1. What is cybercrime?
2. What is the business model of cybercrime?
3. How can we empirically test the impact of the Corona-pandemic on phishing***?***
4. What are the implications of the Corona-pandemic on phishing?

Chapter 1.3 Method, research design and data

In this section we explain our approach to answering each of the research questions. We will make use of a combination of approaches. Specifically, we rely on a retrospective study research design combined with the empiric cycle to give structure to the method of answering the research questions.

Questions 1 and 2. Research questions 1 and 2 will be answered by the literature review. There have been various studies done to understand what cybercrime is, what is meant by the business model of cybercrime and how it operates. The literature study explains what types of cybercrime exists. Who commits cybercrime and what are their motives? I also address the intensity of cyber-attacks during the Corona-pandemic. Has there been a change in the intensity of attacks before and after the beginning of the Corona-pandemic? The literature study gives handles on what we mean by the business model of cybercrime and what the impact of the Corona-pandemic has been on cyber-attacks.

Question 3. This question will be answered by analyzing secondary quantitative data retrieved from the Anti-Phishing Working Group. In this thesis the choice of a retrospective study design was made. Secondary phishing data from before the Corona-pandemic and data after the initial start of the Corona-pandemic will be gathered. The method used to analyze the data was inspired by the paper of An & Kim (2018). An & Kim (2018) proposed data-analyses suited for the discovery of trends in cybercrime. One of these analyses is keywords analysis or term frequency analysis. Put simply, term frequency analysis can uncover trends in cybercrime by counting the occurrence of keywords. This is further explained in chapter 3, the methodology.

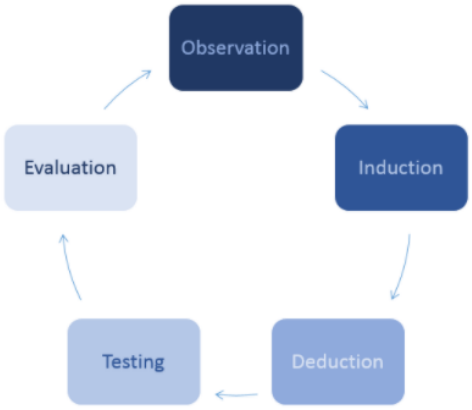
Question 4. The final question will be answered by the conclusions derived from the data-analysis done on the dataset. What are the implications of the Corona-pandemic on the method of phishing by cybercriminals and what can we learn from this? What are the implications for the way that we combat cybercrime? How can we translate the results in policies to combat phishing or cybercrime in general more efficiently?

Research design and data

To help answer our research questions we are making use of the Empirical Cycle developed by (De Groot, 1961). It follows a clear structure on how to conduct empirical research. The cycle starts with collecting and organizing empirical facts. In this case this is the observation of the impact of the Corona-pandemic on various aspects of society *(Observation).* This can be found in chapter 1 and 2, where there is an introduction of the topic and in chapter 2 where the literature is found. Here existing knowledge about cybercrime is presented. Then hypotheses are being formulated test *(Induction).* In the case of this thesis this is the prediction of Interpol (2020) to see an increase of cybercriminals trying to take advantage of the Corona-pandemic. This is found in chapter 2. The consequences of the hypotheses are being translated into testable predictions *(Deduction).* This is also presented in chapter 2 as a conclusion of the literature review. What remains unanswered here? The hypothesis will then be tested against empirical material, the data gathered from the Anti-Phishing Working Group *(Testing).* This part of the empirical cycle is found in chapter 3 and 4. In chapter 3 the methodology is presented, how is the data tested and held against evidence? In the results, in chapter 4, the outcomes of the tests are presented in tables and graphs. The final step in the empirical cycle is the evaluation of the outcome of the testing. Drawing conclusions from

analyses *(Evaluation)*. This is found in chapter 5, where the discussion, main conclusions, and limitations are found.

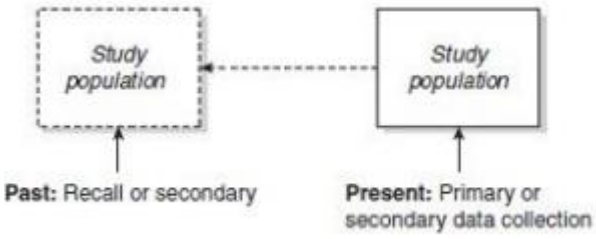
Figure 1. Visual representation of the empirical cycle as presented by De Groot (1961).



The phishing dataset retrieved from the Anti-Phishing Working Group is secondary and thus there is a risk for lower validity of the data. The choice for secondary data regarding this topic comes from the great difficulty in gathering data regarding cybercrime. Criminals generally want to stay anonymous and are hard to find. On top of that, even if data is gathered, it is debatable if this data is then reliable (Cambridge Cybercrime Centre, 2020). However, the Anti-phishing Working Group is a well-established organization that collects data on phishing. Members of the APWG come from over 2200 institutions from all over the world. Institutions ranging from universities, Europol, and governmental bodies such as the European Commission to name a few.

In this thesis, the data retrieved from the Anti-Phishing Working Group is being analyzed through an observational retrospective research design. An observational research design is applicable since there is no experiment being conducted and the data is gathered by observing (Song & Chung, 2010). Retrospective refers to the timing of the data gathered. The data is already collected, before the Corona-pandemic and after the initial start of the Corona-pandemic. Data gathered after the Corona-pandemic is being compared to the data before the Corona-pandemic and is analyzed for significant changes in the business model of cybercrime. From the secondary datasets both the past and present is being studied to being able to infer conclusions about the impact of the Corona-pandemic on the business model of cybercrime.

Figure 2. Structure of the retrospective research design (Song & Chung, 2010).



Chapter 1.4 Reading guide

This thesis is organized as follows:

Chapter 2. Through the means of the literature review the questions about what types of cybercrime exists, who commits cybercrimes, and what we mean by the business model of cybercrime are answered. Additionally, comparing what relevant studies have concluded about the impact of the Corona-pandemic on the intensity of cyber-attacks and the impact on Dark Web Markets.

Chapter 3. The literature review will form the basis for chapter 3, where the methodology is explained, predictions and hypotheses are being stated. How do we prepare the data to test how the Corona-pandemic has affected phishing?

Chapter 4. In this chapter the results will be presented to show and understand what the impact of the Corona-pandemic has been on the subset of cybercrime, phishing. The results will be presented in tables and by showing graphs of the intensity of phishing emails over time and how certain categories evolved over time.

Chapter 5. The discussion, conclusion and limitations will be presented in this chapter. What do the results of chapter 4 mean? How do we interpret the results and are the variations in the results the result of bias? Finally, what are the limitations of the results gathered and what can be improved? From these limitations, future research is suggested.

Chapter 2 Literature review

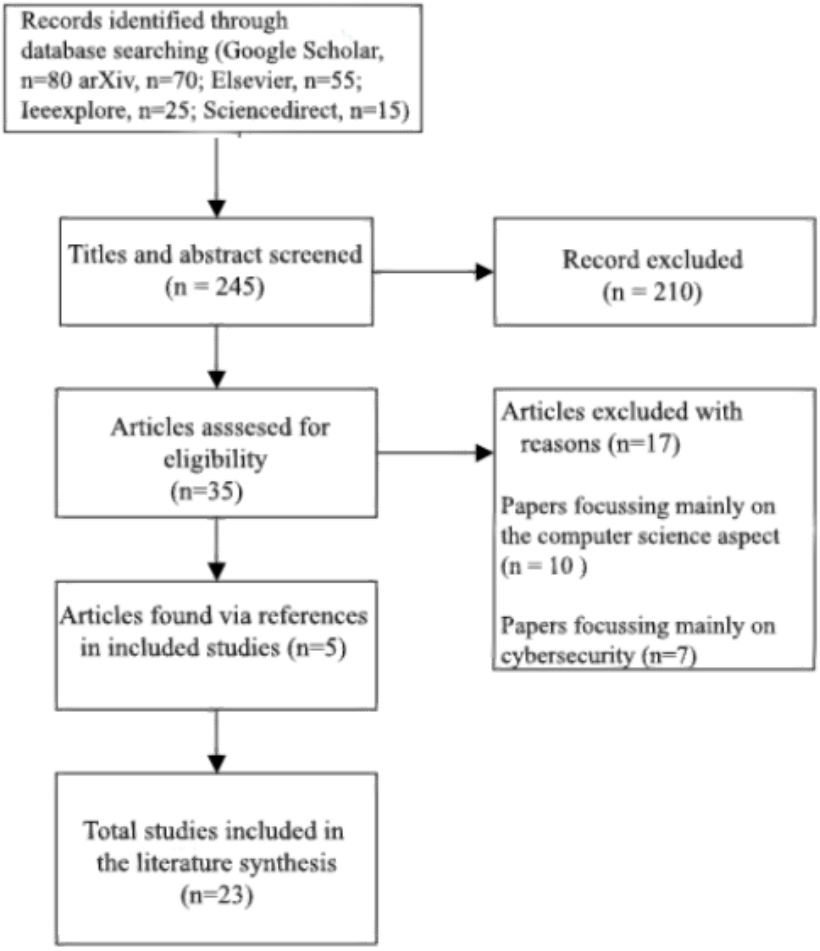
There has been much research on the topic of cybercrime, however relatively little research has been done concerning the business model of cybercrime. To my best knowledge, no research has been done to understand how the business model of cybercrime was affected by the Corona-pandemic. Thus, before we can address this issue in this thesis, a synthesis of literature regarding current knowledge can be read in this chapter. The goal of this chapter is to provide a firm grasp on current knowledge and to identify research gaps. The approach to the structure of the literature review follows the approach as presented by Webster (2002).

Chapter 2.1 Method

This literature review limits itself to DDoS, ransomware, and phishing for two main reasons. These types of cybercrime were identified by Interpol (2020) to have increased the most during the Corona-pandemic. Also, these types of cybercrime are existing in the form of as-a-service according to Huang et al. (2018). Other types of cyberattacks were not fully available as-a-service form but rather evolving or emerging. The review was chosen to be structured in the following way: first the types and motives of cybercrime are identified. Then the business model of cybercrime is introduced. The impact of the Corona-pandemic on the intensity of cybercrime and the impact on dark web markets. Finally, the impact of the pandemic on legal/traditional business models has been assessed.

A flowchart is shown in figure 3, visually representing on how the literature was collected. It is based on the steps suggested by PRISMA to help to improve on the reporting of literature reviews (UNC, 2021).

Figure 3. Flowchart method of literature review



The literature was found searching for keywords related to cybercrime and the Corona-pandemic on various databases. arXiv, Elsevier, Google Scholar, Ieeexpore and Sciencedirect were used when searching for the literature. The following keywords were used in the databases: as a service, business models, impact on business models, business model of cybercrime, corona, covid, covid-19, cybercrime, cyberattack, crimeware, cybercrime as a service, crimeware as a service, CaaS, DDoS, pandemic, phishing, ransomware, routine active theory cybercrime, RAT. Many articles were available on the types of cybercrime. From this a total of 245 papers were scanned for eligibility. However, few articles on the business model of cybercrime were available, at least in comparison to papers on cybercrime. Therefore, these papers were mostly discarded. Leaving 35 papers appearing to be eligible. Another 17 papers were discarded due to focusing mainly on computer science and/or cybersecurity. Five additional relevant papers were identified reviewing references of included papers. For a total of 23 papers used in the literature review.

The papers regarding the intensity of cybercrime during the Corona-pandemic were retrieved mainly from arVix, were unpublished and thus not peer reviewed. This is due to the corona-pandemic on cybercrime has not been studied much and is relatively new. This has impact on the reliability of the papers used because these papers are not peer-reviewed. Other papers, often from organizations, did not state their method of research were used, like Deloitte and Interpol. However, these are reliable organizations and work with government and other notable institutions such as Interpol, and Cyber Fusion Centre.

Chapter 2.2 Types of cybercrime

Before looking into what the business model of cybercrime/crimeware as a service looks like it useful to take a broader view to look at the literature on what cybercrime is and why people engage in cybercrime.

Crime is defined by Poonia (2014) as an act that is forbidden by law and on such act is often a punishment imposed. Cybercrime is a crime where the punishable act is conducted with a computer and/or the target is a computer or a system of computers (Poonia, 2014). Cybercrime can also be understood as any crime where computers and networks played an integral role in committing the crime (Sabillon et al., 2016). Cybercrime is defined by the European commission as crimes that are done via the internet (EUCPN, 2015). In other words, there is no single definition for cybercrime.

It is important to note that cyber-criminal activities are not only committed by criminals or hackers with bad intentions, as one might expect. A broad distinction between hackers can be made to establish a framework. Huang et al. (2018) made distinctions between a defensive side and offensive side. The defensive side concerns itself with cybersecurity and the offensive side with cybercrime. Certain activities are considered “double edged sword” activities as they are conducted by both the defensive and offensive side (Huang et al., 2018). An example of a double-edged sword activity is to try and find vulnerabilities within a system that can be exploited. The defensive side tries to repair these security issues before it can be exploited. The offensive side tries to use these vulnerabilities to their advantage to exploit them. It can be used to undertake an effective cyberattack. Another classification that is used is the distinction between white, grey, and black hat hackers. White hat hackers can be considered the defensive side and black hat hackers the offensive side. Grey hat hackers are hackers who fall between the defensive and offensive side. The grey hat hackers do not have permission to break into any system but do so anyway. However, the grey hat hackers do not have any malicious intent like the offensive side and want to improve security of the system

like the defensive side.

The main concern that can be read in various papers (Huang et al., 2018; EUCPN, 2015) is that the offensive side seems to have an edge over the defensive side. To gain perspective and understand where and why cyberattacks might happen, it is useful to dive into the main motives and goals of cybercriminals. EUCPN (2015) makes distinction between the motives: money, emotion, politics, and religion and just for fun. Li (2017) found 28 motives on why people can engage in cybercrime. The most important motives overlap with the ones described by EUCPN (2015) and are described below.

Money (EUCPN, 2015)/ For acquiring financial gains (Li, 2017)
Cybercriminals who fall in this category want to benefit financially from their crime. In other words, cybercriminals who are motivated by money (EUCPN, 2015). The financially motivated cybercriminal does this for instance because he or she seeks ease in making money (Poonia, 2014), (maintaining) comfort in lifestyle or repaying a debt (Li, 2017). This can thus mean the ransomware attacks targeted on large organizations. There are also hackers and malware tools for hire. An example is RaaS (ransomware as a service). Where one can order ransomware-kits for money (Alhawi, Baldwin, & Dehghantanha., 2018; Huang et al., 2018; Kim & An, 2018). This is very valuable for cyber criminals as ransomware can be bought from them and used without having great technical skill (Alhawi et al., 2018). More broad examples are hackers who steal data from a (rival) company to gain a competitive advantage and stealing credit card credentials to impersonate a victim (Li, 2017).

Emotion (EUCPN, 2015)/ Out of hatred (Li, 2017)
EUCPN (2015) describes cybercriminals being motived by emotion as the most destructive. They act out of anger, revenge, despair, or envy (Li, 2017; EUCPN, 2015). Discontented employees are a group can disrupt their employee's systems out of anger. One can also sabotage another person's system out of revenge or despair. For instance, an ex-spouse receives harassment e-mails. Organizations can attack their competitors out of envy by their wealth and overall success (Li, 2017). Terrorist organizations or anarchists can carry out cyberattack out of hatred, such as ransomware, that seriously disrupts or damages critical facilities (EUCPN, 2015).

Politics and religion (EUCPN, 2015)/ Mobilizing political movement (Li, 2017)
People are willing to commit cybercrimes in the name of politics or religion. This can mean cybercrimes that are being carried out or ordered by terrorist, as described before. But also, in the name of political movements, such as sabotaging a rival political movement. Li (2017) describes the worry for the threat of cyberwarfare. In a cyberwarfare, rivalling countries might attack other countries to attack critical infrastructure that is relying on computer systems and networks. Trautman & Ormerod (2018) argues that this has already happened, with cybercriminals from North-Korea, by infecting more than 200,000 computers in 150 different countries in 2014 with ransomware called WannaCry.

Just for fun (EUCPN, 2015)/ For recreation (Li, 2017)
This type of cybercrime is motivated out of excitement and entertainment (Li, 2017; EUCPN, 2015). Li (2017) compared it with gaming, as there are similar pleasures in hacking as in gaming. There is pleasure in experiencing being able to hack someone's password. There is often no malevolence or financial motivation and a motivation can be to improve their skills. Often these groups of cybercriminals are teenagers (EUCPN, 2015). Although the intent is not to cause damage, it can cost a lot of money for organizations if these attacks are severe. Examples can be ordering a DDOS attack on organizations to prove that they can or out of curiosity what will happen. (Li, 2017).

Routine Active Theory (RAT)

Li (2017) & EUCPN (2015) concluded that the main motive for engaging in cybercrime shifted from "just for fun" to "acquiring financial gain". Other papers in the field of cybercrime (An & Kim, 2018; Leukfeldt & Yar, 2016) provide explanation regarding this motivation in the Routine Active Theory (RAT). RAT is borrowed from criminology to explain the causes of cybercrime. RAT states that crime emerges from the opportunity in crime due to the daily routines that people have (Leukfeldt & Yar, 2016). An example is the increase of home invasions during the holidays when people are not at home.

An & Kim (2018) state there are three elements in play in the emergence of crime. A likely offender, suitable target, and the absence of capable guardians. Taking cybercrime as context: the likely offender is the seller and buyer of crimeware. The suitable targets are vulnerable individuals or organizations. The absence of a capable guardian is the lack of security surrounding systems (An & Kim, 2018). Other models that explain (cyber)crime are similar, like the crime triangle (Lallie, Sheperd, Nurse, Erola, Epiphaniou, Maple & Bellekens, 2020). Where crime is an interplay of a target or victim, the motive of the offender and opportunity for the crime. Leukfeldt & Yar (2016) conducted a literature review about the application of RAT surrounding cybercrime. They concluded that authors do not agree on whether RAT is appropriate as a framework for cybercrime. The reason for this is the broad definition of cybercrime. Leukfeldt & Yar (2016) argue that some cybercrime can be explained by RAT, but not all. High tech cybercrime, such as malware and hacking could partially be explained. Other types of cybercrime, such as stalking could not be explained. They did however find that one aspect of RAT, simply being visible or being online, had a significant effect on all types of cybercrime. This has increased due to the Corona-pandemic as more people work from home.

The Corona-pandemic has a large impact on the current social and economic situation (Interpol, 2020). Likely offenders might have increased due to less opportunity in traditional crime and switched to cybercrime. The suitable targets have increased dramatically due to the increased employees working from home instead of the office. The absence of a capable guardian is security when working from home is less rigorous. Because of this, traditional crime might have shifted towards cybercrime. It can be read later in the literature review that the intensity of cybercrime has increased, and RAT might be able explain that. As criminals want to generate revenue now that social circumstances have changed and want to benefit from these changes. Are previous cybercriminals more active during the Corona-pandemic or has the number of cybercriminals increased? I do however hypothesize that criminals have transitioned towards cybercrime due to the pandemic as predicted by Interpol (2020).

Chapter 2.3: Cybercrime as a business model

Cybercriminals have become sellers of cybercrime for less-technical criminals and have turned it into a business model. Ritter & Pedersen (2020) define a business model as: "The set of activities which a firm performs, how it performs them, and when it performs them so as to offer its customers benefits that they want and to earn a profit". Cybercriminals who run cybercrime as if it were a business are likely professional hackers and they are often driven by financial gain. They target their cybercrime mostly towards vulnerable individuals and organizations. From this, entire underground economies have developed surrounding this type of cybercrime and is a multibillion-dollar industry (EUCPN, 2015). It has thus attracted more cybercriminals also wanting to gain revenue. If cybercrime is viewed as a business model, the motives for cybercrime can vary. Where cybercrime is offered as a service, the criminals

ordering a cyber-attack can do this out of a political motive. Whereas the providers of the cyber-attack can have a different motive like financial gain.

Cybercrime-as-a-service (CaaS)

Surrounding cybercrime, highly sophisticated business models have emerged such as Crimeware-as-a-Service or CaaS. CaaS is one of the main business models that is used in cybercrime (An & Kim, 2018). However, there are relatively few academic studies focusing on CaaS. The term crimeware-as-a-service is self-explanatory: it is a business model that offers different types of crimeware or cybercrime as a service. This service model allows for cybercrime to be more accessible. CaaS is designed by highly skilled developers/programmers to be used by cybercriminals without great technical skill. CaaS has become an important trend for cybercriminals (An & Kim, 2018; Huang, 2018). As previously mentioned, little research has been done regarding the business model of cybercrime. Huang et al. (2018) wrote an extensive theoretical framework for the business model of cybercrime. The theoretical paper of Huang et al. (2018) is therefore an important one, but empirical studies lack.

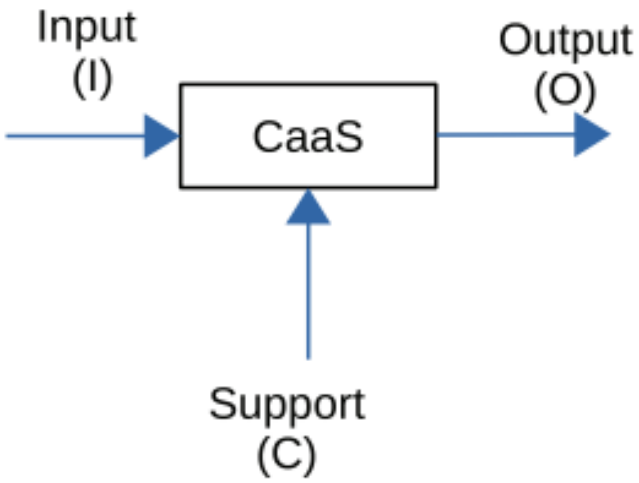
Interpol (2020) expected to see an increase in the use of CaaS by cybercriminals. As mentioned in the introduction, cybercrime has seen a shift from product oriented towards service-oriented models (An & Kim, 2018). Examples of such a service is Ransomware-as-a-service (RaaS). Where previously one had to have the technical know-how how to code ransomware, know how to deploy ransomware, launder the proceeds of the ransomware attack, criminals can now order complete attacks on illegal markets (Huang et al., 2018; An & Kim, 2018; Wegbreg, Klievink & Eeten, 2017). The highly skilled (criminal) developers in this case do not carry out the cyber-attack themselves, which allows these cyber-criminals to be more anonymous.

An & Kim (2018) discussed there are certain roles associated with CaaS. Developing a hacking-tool used in CaaS, setting up a cyber-attack, performing a cyber-attack, providing infrastructure for the cyber-attack, and laundering of the proceeds gained. This is in line with the value chain that Wegbreg et al. (2017) developed. Huang et al. (2020) models the business of cybercrime in a value chain in two types of activities, the primary and support activities. The difference with the models of An & Kim (2018) and Wegbreg et al. (2017) is that Huang et al. (2018) adds support activities explicitly. The primary activities mainly focus on the cyber-attack that is carried out. It starts with vulnerability discovery: where are the weak spots of a security system where the attack can take place? Then, the tool that carries out the attack is developed and delivered. Finally, the attack is carried out. To facilitate a cyber-attack, there are support activities. Attack life-cycle management, human resource (like a hacker community), an illegal marketplace where cybercrime products can be sold, and a money laundering scheme are all examples of support activities. They all exist to let the primary activities run as smooth and effective as possible. The primary and support activities are also offered as a service, as mentioned by both Huang et al. (2018) and Wegbreg et al. (2017). For example, the vulnerability discovery in the primary activities is offered as-a-service in Huang et al. (2018) as described below. The support activity “money laundering” is offered as a service as described in Wegbreg and colleagues (2017) as Money Mule-as-a-service. The marketplace is also offered as ‘platform as a service’, where criminals can create their own illegal marketplace. Here they can sell their own selection (or created themselves) of products, hacking tools or other products/services (Huang, 2018). This also is a central point in one of the challenges of understanding and mapping cybercrime. Studies often focus on large illegal marketplaces and not on smaller marketplaces such as created by ‘platform as a service’. These markets are potentially hard to gain access to as they require an invite to enter and thus hidden or at least harder to access for research purposes.

In figure 4 is a visual illustration from Huang et al. (2018) how CaaS tends to operate.

Huang et al. (2018) gave the example of "vulnerability discovery" is offered as-a-service under VDaaS (vulnerability discovery-as-a-service). Here the input is the target victim, output is the discovery of vulnerability of the target, support is the vulnerability discovery tool used.

Figure 4. Operation of Cybercrime-as-a-Service by Huang et al. (2018)



Chapter 2.4: DDoS, ransomware, and phishing.

The types of cybercrime, DDoS, ransomware and phishing, are chosen because these are the types of cyberattacks that were mostly reported by Interpol (2020) in August 2020 during the pandemic. They are also offered as-a-service. Huang et al. (2018) provides an intensive and broad theoretical framework to on how different types of CaaS tends to operate. This helps greatly to get an understanding of how these services might operate and what the economics of these types of attacks look like. After this section, the impact on the intensity of these cyberattacks, but not the business model, is addressed.

DDoS-attacks

In a Distributed Denial of Service attack or DDoS attack, the target (for instance online banking) becomes flooded with a large amount of request coming from many computers. The goal and the result of such flooding is that the servers are no longer able to complete requests from actual visitors and thus the servers become unreachable. An example from practice how this can cause damages the victim is a large online store that is unreachable for customers and thus the store loses revenues. The health services are more targeted due to the corona pandemic and thus the increased stress. DDoS attacks do not aim to steal data but can disrupt functioning of systems (An & Kim, 2018; Netscout, 2020). Cybercriminals can threaten with DDoS attacks if certain conditions are not met, like a sum of money. In that way ransomware attacks and DDoS attacks are similar. If the victims fail to pay a ransom, their systems become unusable.

The first DDoS attack was recorded in 1999 (MIT, 2019). Then, DDoS attacks could only be pulled off by someone having the tools and proper technical skills. Today, DDoS attacks can be bought of a marketplace as-a-service. The applicant of the DDoS attack does

not need any tools or technical skills as the attack is being done for them. Following Huang et al. (2018) a DDoS attack as-a-service can be understood as TAaaS (Traffic as a service). Looking at the figure by Huang et al. (2018): the input would be selecting targets, the support activity is the use of traffic generating tools, and the output is a DDoS attack with the intended purpose. According to Huang et al. (2018) this type of CaaS is currently existing, has a pricing model of the type subscription and has been observed to cost around between $300 and $999 a month in 2017 and 2018. Deloitte (2018) estimates that DDoS-services cost around $36 to $62 dollars per hour. However, Deloitte notes that price clearly depends on the sophistication of the DDoS-attack. To attack a government website costs significantly more than ordering an attack on a trivial website.

Ransomware

Ransomware is a type of malicious software that prevents individuals or companies from reaching their files. The files can be retrieved when the user(s) pay the amount of money that the distributer of the ransomware or attacker demands. Ransomware can prevent the user to access files in in two ways (Kiru & Aman, 2019). Either by preventing the user to access the operating system until the ransom is paid. Or the files are encrypted and thus not accessible until the user(s) pay the ransom. The payment today is usually in form of cryptocurrencies like Bitcoin.

Academic literature is conflicting when addressing ransomware-as-a-service. Huang et al. (2018) views RaaS as a combination of services offered. Kim & An (2018) view ransomware not as CaaS, but as crimeware products that can be bought on dark web marketplaces. However, Huang et al. (2018) wrote that the buyer of crimeware products can integrate these crimeware products into their expertise and become a service provider. In other words, the buyer of crimeware can commercialize their specialization/expertise as services for other cybercriminals to use.

Ransomware-as-a-service is being offered as-a-service on dark web markets. Example cases of RaaS are the Philadelphia tool and Fatboy in 2017. If illustrating in the figure proposed by Huang et al. (2018) the input would be the victim, the support activity is a RaaS tool, such as the Philadelphia tool, and the output would be the ransomware attack and ransom paid by the victim.

Deloitte (2018) found different strategies for the pricing model of RaaS. Some types of RaaS can be bought for a one-time payment, a subscription or a percentage of the revenue gained is paid to the developers or sellers of the RaaS. The most expensive RaaS was around $1500. Cheaper, more detectable types of RaaS were sold for $39. The average price, according to Deloitte (2018), was $1044. Subscriptions vary from $21 up to $125 per month.

Meland, Bayoumy & Sinde (2020) studied RaaS from 2018 to 2019 and concluded that the threat of RaaS is only a modest one, perhaps contrary to what other papers and media suggest. The amount of RaaS listings declined from 2018 to 2019 even though the total listings increased (Meland et al., 2020). They only found a small number of marketplaces that offered RaaS, even though these were the most popular dark web marketplaces. In the opinion of the researchers many of these listings were questionable in their authenticity (Meland et al., 2020). However, this thesis was written before the Corona-pandemic and RaaS might have seen an increase since.

Phishing

Chawla & Chouhan (2014) define phishing as the act of sending an email to a victim like an organization or individual and claiming to be someone else. For instance, the senders of the phishing email claim to be employees of a bank to steal credentials or to trick the victim into downloading malicious software. Phishing is not limited to emails, however. Fake websites and mobile messages claiming to be official organizations are also examples of phishing.

Since the Corona-pandemic phishing emails often impersonate health officials tricking victims into sending their credentials. The problem with phishing email is that these emails look identical to email from actual legitimate organizations and often sound urgent. Especially with the use of e-mail spoofing, where email addresses of official organization can be nearly copied, many victims do not get suspicious and trust the email. Phishing especially has seen a large increase during the Corona-pandemic (Interpol, 2020; Lallie et al., 2020; NOS, 2020).

Phishing is generally done via two techniques (Chawla & Chouhan, 2014), deceptive phishing and malware-based phishing. Deceptive phishing uses social engineering techniques tricking the victim into believing the email came from legitimate organizations as the examples given above. Malware-based phishing tricks the victim into clicking on an attached link or file, thereby installing malicious software. This malicious software can then steal the victim's credentials directly.

Cybercriminals do not have to carry out a phishing attack themselves but can buy phishing attacks as-a-service. Huang et al. (2018) models this service as Deception-as-a-service or DaaS. Where the input is information about a specific target, like how to evade the security of an organizations network. Then the phishing attack is realized by using the support of a deception development tool. The output would be receiving stolen credentials or receiving money as a result. NOS (2020) reported that in the Netherlands a suspect was arrested for developing software that can be used to create and impersonate a online banking website and deceive victims into filling in their credentials. At the time of writing, it is the first time that a suspect was arrested for building such software.

DaaS can be used for a phishing attack service and was observed to have a pricing mechanism in both subscription and commission. A subscription is raging from $85 to $115 a month. While the commission is around 40% of profit made (Huang et. al, 2018). NOS (2020) reported a case where cybercriminals can buy a complete fake online Dutch banking website for 262 euro. This is in accordance with the estimation of Deloitte (2018), where more sophisticated phishing, such as impersonating a bank, is available for $300. More simple types of phishing services are available starting at $10.

Chapter 2.5: The impact of Corona-pandemic on intensity on cyberattacks.

In this section the relationship between the intensity of cybercrime and the Corona-pandemic is discussed. Lallie et al. (2020) note that cybercriminals often try to benefit from a crisis. During the hurricane Katrina in 2005, many fraudulent domains and phishing attempts were launched in the name of officials. Similarly, this also happened in 2016 during the aftermath of the earthquakes in Japan. Is there a relationship between the intensity of cybercrime and the impact on the business model of cybercrime? If there is no observed change in the intensity of cybercrime, it is likely that there is no change in the business model of cybercrime. However vice versa could apply: if there are indications that the intensity has increased, perhaps the business model was also influenced. But it might also indicate that existing cybercriminals have become more active.

Corona pandemic and DDoS-attacks

Netscout (2020) reported that the Corona-pandemic has functioned as fuel for DDoS attacks. Since the lockdowns started in the United-States and Europe in March, there has been a 25% increase in DDoS attacks. Khan, Brohi & Zaman (2020) also state there was an increase in DDoS attacks. In May, Netscout (2020) reported there were 929.000 detected DDoS attacks, which is the largest number of attacks in a single month. Additionally, Khan et al. (2020) and Netscout (2020) mentioned that DDoS are concentrating on the organizations that play critical roles in the pandemic, such as the health services. These organizations are the most vulnerable

when being extorted, and thus criminals focus their attention to those organizations who are most likely to pay ransom. However as mentioned before, these types of attacks were less prevalent in the results of Lallie et al. (2020).

Corona pandemic and ransomware

There has been a large increase in the amount of ransomware attacks during the pandemic (Carbon Black, 2020; CFC, 2020; Khan et al., 2020; Lallie et al., 2020). There especially was a large increase in the first two week of April. However, the CFC (2020) expects more victims as many systems could have been infected, but the ransomware might not have been employed. Carbon Black (2020) reported a 149% increase of ransomware-attacks in March compared to February. The researchers of Carbon Black (2020) also report that spikes in ransomware attack correlated significantly with key news reports about the Corona-pandemic. Reports such as the first victim in the United-States and the following lockdowns in the United-States and Europe. Lallie and colleagues (2020) found an increase in individuals and organizations who became a victim of ransomware. Lallie and colleagues (2020) gave the example of Corona specific ransomware, where ransomware is disguised as a Corona heat map. Khan et al. (2020) also found Corona-specific ransomware that lures victims into downloading ransomware that is disguised as a Corona info application. The locked data can later be used to threaten to release this data if further ransoms are not paid. Similar to what the Netscout (2020) reported on DDoS-attacks, on an organizational level health services are most targeted for ransomware-attacks (Lallie et al., 2020). This is most likely due to the critical role that the health services play during the pandemic and thus the vulnerability due to the risk of the casualties of patients (Khan et al., 2020; Lallie et al., 2020; Netscout, 2020). An interesting point can be read in Lallie et al. (2020) where they state that leading cybercriminals have promised to stop attacking the health services, at least until the stress due to the pandemic has reduced.

Corona pandemic and phishing

There was a significant increase in Corona-related phishing and fraud (Interpol, 2020; Khan et al., 2020; Lallie et al., 2020). Fraud related crimes include making use of supply shortages and fake medications. Phishing email increased during the Corona-pandemic trying to steal credentials and passwords, in other words deceptive phishing. The senders of these phishing emails often impersonate health officials, government, and employees of a company like a CEO. This agrees with the paper by Lallie and colleagues (2020) who states there was a particular large increase in phishing during the pandemic. According to Khan et al. (2020) and Lallie et al. (2020) this was expected as cybercriminals react to a crisis with phishing emails, as mentioned before. NOS (2020) also reported there was an increase of phishing attempts impersonating Dutch tax authorities. There was an increase from 2000 notifications per week before the pandemic to 10.000 to 12.000 notifications during the Corona-pandemic. Since June 2020, this amount has reduced ranging from 4000 to 6000 notifications per week. Interpol (2020) noticed an increase in malware-based phishing as well. Many emails that were sent in the name of health organizations contained malicious software. Malware like ransomware, but also software designed to steal sensitive information. This agrees with the paper from Lallie and colleagues (2020) who conclude that many of the cyberattacks during the pandemic start with a phishing-campaign. From this phishing URLs and/or attachment with malware (in cases ransomware) is spread. To increase the chance of success, these phishing campaigns are timed to announcements or events. For instance, official announcement from health officials and/or government (Lallie et al., 2020). The increase in phishing attempts impersonating Dutch tax authorities combined the pandemic and the timed the moment to file tax returns.

Chapter 2.6: The impact of the Corona-pandemic on activity on dark web markets.

In the next section the impact of the corona-pandemic on dark web marketplaces is being discussed. Deloitte (2018) wrote that the underground economy is an interrelated ecosystem where a mixed assortment of tools and services are available. CaaS can be bought through listings on dark web marketplaces and other studies focused on other types of listings and activity on dark web marketplaces during the Corona-pandemic. This might give clues on what happened to the CaaS listings during this time.

Dark web markets

Dark web marketplaces (DMW) play a key role in the economy of cybercrime (An & Kim, 2018; Bracci, Nadini, Aliapoulios, McCoy, Gray, Teytelboym, Gallo & Baroncehlli, 2020; Vu, Hughes, Pete, Collier, Chua, Shumailov). Among other places like hacker-forums, dark web markets (DMW's) are one of the places where cybercriminals meet and where illegal products can be bought. Drugs, weapons, stolen credit cards and stolen passports, but also cybercrime products and services like CaaS. DMW's are often only accessible using encrypting web-browsers like TOR. Additionally, the product sold can be paid for by using the previously discussed cryptocurrencies such as Bitcoin. These measures provide anonymity of both the buyer and the seller. An & Kim (2018) mention that DMW's have at least the following elements. First, actors: people involved in CaaS. Mainly they consist of developers/programmers of the crime-tools, operators of the crime tools and the buyers of CaaS. Second, value chains: chains of operation that are used to add value. Such as value chains used in Huang et al. (2018). Third, modes of Operation. For instance, CaaS or crime tools. CaaS and Crime tools differ, however. Where Crime Tools are do-it-yourself crimeware, with CaaS the criminals outsource the activities (An & Kim, 2018).

The impact of the Corona-pandemic on black markets

Research has been done to understand how the Corona-pandemic influenced the economics of DMW's (Bracci, et al., 2020; Vu et al., 2020). Cybercrime-as-a-service might have gained popularity during the corona-pandemic, as predicted by Interpol (2020) due to increased opportunity for criminals. Previous studies focus on the effect of the Corona-pandemic on Dark Web Markets, but not on CaaS listings.

The studies done on the effect of the Corona-pandemic on DMW's concluded there was a large peak during the beginning of the Corona-pandemic, but this quickly went down. The study done by Bracci and colleagues (2020) analyzed the trends in categorizations of Corona-related items. A paper by Vu and colleagues (2020) investigated transactions and members on these markets and this data was analyzed before and after the pandemic.

Bracci et al. (2020) assessed the effects of the Corona-pandemic on DMW's listings that are Corona-related such as personal protection equipment, medicines, and vaccines. From January 2020 until July 2020 listings of Corona specific health supplies on 23 different DMW's were analyzed. These Corona-related listings were also compared to public attention derived from Twitter and Wikipedia visits concerning corona health supplies. An influx of public attention on the Corona-pandemic was due to the Wuhan quarantine and corresponded with the emergence of Corona-related listings on dark web markets. This is also found by Interpol (2020), Lallie et al. (2020) and Carbon Black (2020) concerning the increase of cyberattacks during the pandemic. A second influx of public attention and emergence of Corona-related listings was in March due to the quarantine of countries in Europe. It was found that this influx was short lived: after the quarantine in Europe was step by step released, there were less and less Corona-specific listings. Bracci et al. (2020) also found that listing prices correlated with degree of public attention. The median prices experienced a sharp increase. The explanation given is perhaps not popularity but could be due to speculation in expected demand.

Vu et al. (2020) took a more economic view of Dark Web Markets, looking at transactions and number of users. In this paper similar results were reported. DMW's were observed during the years 2018, 2019 and 2020. There was special attention to the period before and during the Corona-pandemic had emerged. Vu et al. (2020) found that at the beginning of the Corona-pandemic in March 2020 there was a large increase of transactions observed in DWM's. Especially a large increase in users who make a one-time sale was seen. However, this sharp increase did not last long. In April 2020, a month after the Corona-pandemic started, there was a drop in users and transactions. Therefore, Vu and colleagues (2020) argued that the Corona-pandemic was a stimulus for DWM's, rather than a transformation. This result is similar that was found by Bracci et al. (2020) where after an initial peak, the Corona-specific items reduced.

Chapter 2.7: Impact of the Corona-pandemic on business models.

Some businesses can benefit from a crisis such as the Corona-pandemic, like online retailers, food delivery and perhaps CaaS as predicted by Interpol (2020). However not all business model prospers during a crisis, as is evident in the many bankruptcies during the Corona-pandemic. Ritter & Pedersen (2020) distinguish five different types of business models on how they function during a crisis that can be placed in two categories. Category 1 resilient: antifragile, robust, adaptive, and suspended business models. Category 2 vulnerable: aided and retired business models Antifragile business models perform better in times of a crisis than in normal circumstances. This might be due to increased opportunity and/or flexible nature of the business model. In other words, the current business model itself is not affected by a crisis but performs better than it normally would. The Robust business models can experience a decrease in revenue, but the business model itself is not affected. This is the case with an adaptive business model, where the business model is affected by a crisis and needs to change. Examples that Ritter & Pedersen (2020) gave were the transfer from in-class education towards online education, companies offer remote support instead of physical support. The impact of a crisis can be costly on this type of business model, but the business quickly reemerges. Suspended business models must close their business due to the crisis but are reopening after the crisis. An example in the Corona-pandemic is the entertainment sector. At the time of writing this, in many countries this sector is on hold. However, these businesses can support themselves during closure. And many businesses in that sector are expected to be profitable again after re-opening. Aided business models are not able to support themselves during a crisis. The businesses are expected to be profitable again after the crisis but need external support to survive a crisis. External support can come from government or investors who speculate that the business will generate profit again. Finally, the retired business model is a business that will not survive a crisis. Even with external support from government, but investors are not interested in investing this company as they suspect the business will not recover after business. Ritter & Pedersen (2020) argue that these business models need to shut down and start fresh after a crisis.

Resilient business models can survive on their own and are making profit during a crisis or are expected to make a profit again after a crisis. Vulnerable business models can not survive a crisis without external help. With external help from government and/or investors, some businesses are expected to make profit after the crisis. However vulnerable business models can also cease to exist during or after a crisis.

Ritter & Pedersen (2020) assessed the impact of the Corona-pandemic on business models via interviews with eight companies with each more than 500 employees. These interviews provide depth, but for understand cybercrime this method of research is not viable. Criminals tend towards anonymity. Ritter & Pedersen (2020) used different questions to

assess the business models. Whether value proposition has changed or terminated since the pandemic, how did the Corona-pandemic impacted channels of conducting business and are there changes in the capabilities of your business? Demand-curves were made to show how demand has grown or declined since the pandemic and a project for after the pandemic.

Pointers to assess how a business model was affected by the Corona-pandemic can be inferred from trends observed in CaaS from DMW's. An & Kim (2018) reason that, according to the previously discussed RAT theory, DMW's follow economic principles like supply and demand. Trends in supply can infer a trend in demand. Mainly by analyzing median prices: if demand exceeds supply, then the median prices will increase. But also, by the amount of advertisements surrounding CaaS. Changes in value propositions/demonstration and capabilities from before and after the pandemic can perhaps be observed in DMW's by following changes in scope of promised features.

Conclusion

Studies indicate that the intensity of cyberattacks has increased since the Corona-pandemic, signaling that cybercriminals are taking and have taken advantage of the lockdowns. Carbon Black (2020), CFC (2020), Interpol (2020), Khan et al., (2020) and Lallie et al. (2020) found that there was a particular large increase in phishing, spread of malware through phishing and ransomware attacks. These activities correlated with Corona-related news reports. Lallie et al. (2020) argues there were less cyber-criminal activities in the form of DDoS-attacks and hacking. However, Khan et al. (2020) and Netscout (2020) did report a clear increase in DDoS attacks. The effect of the pandemic on dark web markets, where cybercrime services are available, were studied during the pandemic. Similar to the increased amount of cyberattacks since the pandemic, there was also a sharp increase in the Corona-specific listings and transactions. However, this increased activity on DMW's was short-lived as after an initial peek, the activity went down. Therefore Vu et al. (2020) argued the pandemic acted as a stimulus rather than a transformation. Various researchers (Khan et al., 2020; Kashif et al., 2020; Lallie et al., 2020;) argue that the increase of cyberattacks is due to the impacted economic and social circumstances caused by the Corona-pandemic. This thesis zooms in on phishing, a subset of cybercrime. We try to answer what the impact of the Corona-pandemic has been on phishing. From the literature review it becomes clear that the intensity of phishing has increased significantly during the pandemic. But are cybercriminals using Corona-specific topics (other than the ones mentioned in the literature study) in their phishing campaigns, similarly to Corona-specific listings on DMW's? This will be studied further and answered in this thesis.

Chapter 3 Methodology

Research approach

From the literature review it is clear that the intensity of cybercrime has increased since the start of the Corona-pandemic. If we understand better how cybercrime is affected by the pandemic, it can be addressed more effectively. We hypothesize that criminals want to make use of topics related to the Corona-pandemic to generate revenue. So, we expect a significant change in the topics after the start of the Corona-pandemic. A phishing dataset was used to understand what the effect might have been on the topics used by cybercriminals. This dataset contains many phishing emails collected over a period of a year. Previous research found that the intensity of cyberattacks has increased since the Corona-pandemic. The method explained in this chapter aims to test the hypothesis by inductive analysis. First it is explained where the dataset about phishing was retrieved from. Secondly, the steps undertaken to clean the data and transformation of the data are described. Finally, the approach to the data-analysis is presented and which statistical tests will be used.

Chapter 3.1 Data collection

The dataset about phishing is retrieved from the Anti-Phishing Working Group (APWG). This organization collects data about cybercrime and in particular phishing data and makes this accessible for researchers to use. The provided dataset contains historical reported phishing emails from September 2nd, 2019 to October 9th, 2020. The archival data of phishing is used to compare differences in phishing methods before and after the Corona-pandemic. The goal is to understand what effect the start of the Corona-pandemic had on the methods of phishing. According to the World Health Organization (2020), the Corona-pandemic started at the 11th of March 2020. The time-period that is used in this thesis is taken between September 2nd, 2019 and July 2nd, 2020. So, the data is taken roughly six months before and four months after the WHO declared the pandemic for a total of ten months of data. Table 1 summarizes the content of the phishing dataset.

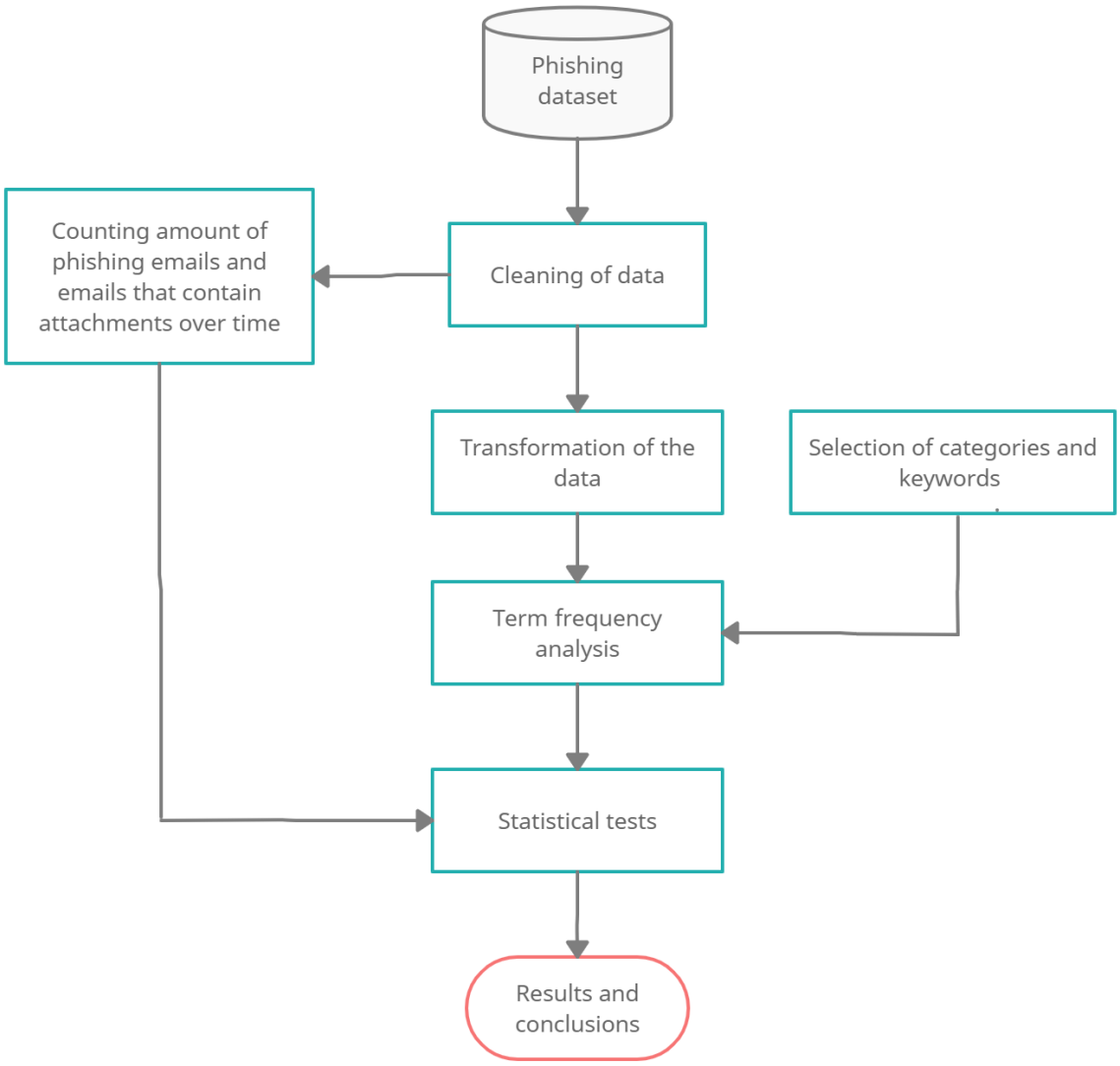
Table 1. Content of the phishing dataset from APWG

Contains	Phishing emails
Retrieved from	Anti-Phishing Working Group
Timescale dataset	September 2nd, 2019 to October 9th, 2020
Structure	Set of phishing emails per day
Total number phishing emails	2.345.505
File type	JSON and HTML

Chapter 3.2 Approach

The approach to the analysis of the phishing dataset is visually represented below in the flowchart. The method section will follow the steps as showed. These steps are derived from Knowledge discovery in databases or KDD. It plays an important role in understanding data as it provides clear steps on how to approach large amounts of data and ultimately find meaningful patterns to base your results on (Fayyad, Piatetsky-Shapiro & Smyth, 1996). This can include data from any field: either it being large amounts of financial data, health services or data on cybercrime.

Figure 5. Flowchart of the methodology



Amount of phishing emails that contain attachments.

The phishing dataset has various keys stored into the emails. Information about whether a phishing email contained an attachment is particularly interesting due to the distinction between deceptive and malware-based phishing. This can be easily done by first counting the number of phishing emails that contained attachment and divide this by the number of phishing emails reported that day. The attachments in the phishing emails could contain malicious software. Analyzing whether there has been a significant increase in attachment in phishing emails, might give an indication about changes in the method of phishing. Lallie et al. (2020) wrote that phishing campaigns that include attachments often contain malware. This can act as a vehicle for financial fraud as malware can steal credentials or this malware can be in the form of ransomware.

Cleaning and transformation of the data

The cleaning of data is an important step when analyzing data. Clean data allows for analysis to be more efficient. Data cleaning gets rid of information that is irrelevant, but also gets rid of data that are duplicates or contain entries that can cause errors like large blank spaces. Also, cleaning the data helps to reduce the size of the dataset which can reduce the processing load and time when analyzing. Transformation of data is also needed to properly assess the data. Like smoothing the data to get rid of noise, correcting misspelled words, but also transforming the emails in a manner that no HTML-code is present as this can lead to errors. These steps can help in improving the quality and reliability of the data which in turn helps in the analysis and conclusions.

The dataset was reduced to 1.373.324 phishing e-mails by taking the data between September 2nd, 2019 and July 2nd, 2020. First a vocabulary was built for each phishing email on each day in a bag of words. This was done by using the NLTK library in python. NLTK stands for Natural Language Tool Kit and is used for many applications in natural language processing. This was done by extracting visible text from the emails HTML. Then numbers and punctuation were removed from the text, replacing whitespaces with single spaces, removing special characters which were not encodable by utf-8 encoding, lowercase all text to prevent duplicates, removal of words shorter than three characters, spellchecking and removal the words that did not occur in the dictionary. The spellchecking was done using the python package Pyenchant. Which is solely used check the spelling of words and suggest corrections. A term frequency–inverse document frequency (tf-idf) calculation was done on the vocabulary using the Sklearn python library. This library is used for various applications in machine learning. In this thesis is it used for doing the tf-idf calculations. Tf-idf calculates the relative importance of a word by counting the occurrence of a word and total number of emails and dividing this by the number of emails containing the word. Words with very high and very low tf-idf scores were removed. Words with low tf-idf scores are often stopwords that can be found frequently in emails and are thus of no or little importance. Words that have high tf-idf scores come across so little that they also have little importance. After this was done, the size of the dataset was reduced dramatically by transforming the emails into a Sparce matrix via a Countvectorizer found in the scikit-learn package using Python. This is a matrix with mostly zeroes but counts the indices where a word was found in an email. Because this reduces the size of the dataset, the analysis can be done in less time and less computational power. The factors regarding the order of words in an email is lost, but this does not matter since in keyword analysis the occurrence of a word is important. Then the occurrence per keywords can be plotted over time and analyzed. The plotting was done by making use of the Plotly python library. This library is used to present data visually. The graphs and the cut-off points were made using this library. From this data a trend can be inferred on how these keywords and thus categories were influenced by the Corona-pandemic.

Term frequency analysis

The method of analysis is inspired by previous studies on other types of cybercrime, like underground markets (Bracci et al., 2020; Broadhurst, Lord, Maxim & Woodford-Smith, 2018; An & Kim, 2018). The analysis is both done on qualitative data like the content of the emails and quantitative data like the date of receiving the emails, whether the email had attachments and/or links. Bracci et al. (2020), Broadhurst et al. (2018) and An & Kim (2018) used a method of keyword analysis to mine data from the datasets concerning underground marketplaces. This is also done by Bracci et al. (2020) on Corona-specific listings on

underground markets. A trend can be inferred from the mentioned keywords and the effect of the Corona-pandemic can be measured. Similarly, this can be done in analyzing the phishing dataset. Employing keywords analysis or term plotting over phishing emails can reveal categories evolving over time. These clues can give insight in changes in the methods of how cybercriminals use phishing to gain revenue during the Corona-pandemic. Bracci et al. (2020), Broadhurst et al. (2018) and Kim & An (2018) made categories per listing with keywords to find these listings. For the phishing dataset: certain categories are associated with various keywords. In table 2 is an overview of the categories with the associated keywords in the wordlist. This is however not a complete list in the sense that there might be keywords associated with the given category which is not captured. This is also true for the categories themselves. This is addressed further in the discussion in chapter 5. The categories "Corona related mentions" and "Medical and protection equipment" are taken from the study of Bracci et al. (2020). Where they studied the prevalence of these categories on dark web markets to understand how dark web markets reacted to the pandemic. The category "financials" was inspired by the study of An & Kim (2018). This category and the category "order and delivery scams" were also inspired by the report of Fireeye (2012). They made a list of the most common words used in phishing emails in 2012. Terms that are on top of the listed are often related to financials, order, and delivery. Since the corona-pandemic, there have been reports of increased use of online shopping due to the closure of stores. Perhaps cybercriminals want to use this increase for their own benefit. Finally, there are also types of phishing emails that I expect to reduce in prevalence. Emails concerning "dating, beauty and care" are expected to have become less relevant during the pandemic. Due to the lockdown people are less interested in dating since this requires meeting in person. Also, due to people staying at home, perhaps people are less interested in personal beauty like losing weight by trying diets.

Table 2. In the table below are the keywords associated with a category. These keywords will be used in the keyword-analysis to discover trends in the categories over time

Category	**Wordlist**
Corona related mentions	*"corona", "disease", "pandemic", "virus"*
Medical and protection equipment	*"chloroquine", "mask", "masks", "vaccines", "ventilators"*
Financials	*"bank", "donation", "payment", "donate", "debt", "invoice", "subscription",*
Order and delivery scams	*"amazon", "account ", "delivery", "order", "shipment",*
Dating, beauty and care	*"date", "dating", "diet", "fat", "meet", "single", "skin" "women", "weight"*

The keywords that are analyzed were counted as unique mentions in the phishing emails. As can be read later in the results, the number of phishing emails after the start of the Corona-pandemic has increased significantly. Reporting only the number of times a keyword was found can give a wrong view as it is expected that certain keywords are found more often as the amount of phishing emails increase. This is also true for the number of emails that have

attachments. Therefore, the amount of unique mentioned keywords is normalized by giving the average mentions per day. This means dividing the number of mentions by the number of phishing emails on a given day. This gives a normalized count for keywords. Day X simply means a given day. For instance, the number of unique mentions of keyword "corona" on 15th of April 2020 divided by the number of reported phishing emails on the 15th of April 2020.

$$Normalized\ average\ count\ = \frac{\text{Number of unique mentions of a keyword on day X}}{\text{Number of emails on day X}}$$

This is also done for examining the number of phishing emails containing attachments. Dividing by the number of emails on a given day gives the proportion of the emails per day that contain attachments.

Timeline of the Corona-pandemic

Below is a table that provides context on how the Corona-pandemic evolved over time. This information is retrieved from the World Health Organization (2020). For us, the most important date is the 11th of March 2020, since the WHO then declared the Corona-situation a pandemic.

Table 3. Overview of the timeline of the pandemic based on the information retrieved from the WHO (2020).

Date	
31 December 2019	Wuhan Municipal Health Commission in China reported that a cluster of pneumonia cases were identified in Wuhan. The novel coronavirus was eventually identified
5 January 2020	The World Health Organization provided an official publication to the scientific and health community and the media. This contained risk assessments and advice and what had been happening in China regarding the then called pneumonia cases in Wuhan.
13 January 2020	The first Corona case was recorded outside of China, in Thailand.
30 January 2020	The WHO reported that there were 7818 total cases confirmed worldwide. Most of the cases were in China, 82 were outside of China in 18 different countries. The WHO gave a high global risk assessment.
11 March 2020	The WHO declared the global Corona-pandemic. From here lockdowns in Europe and the United-States were issued in the following days.

Statistical analysis

Statistical tests on the data are processed in SPSS. The statistical analysis has the purpose to test for significant increases or decreases in the data from before the pandemic and after/during the pandemic. The data collected is assumed to be normally distributed.

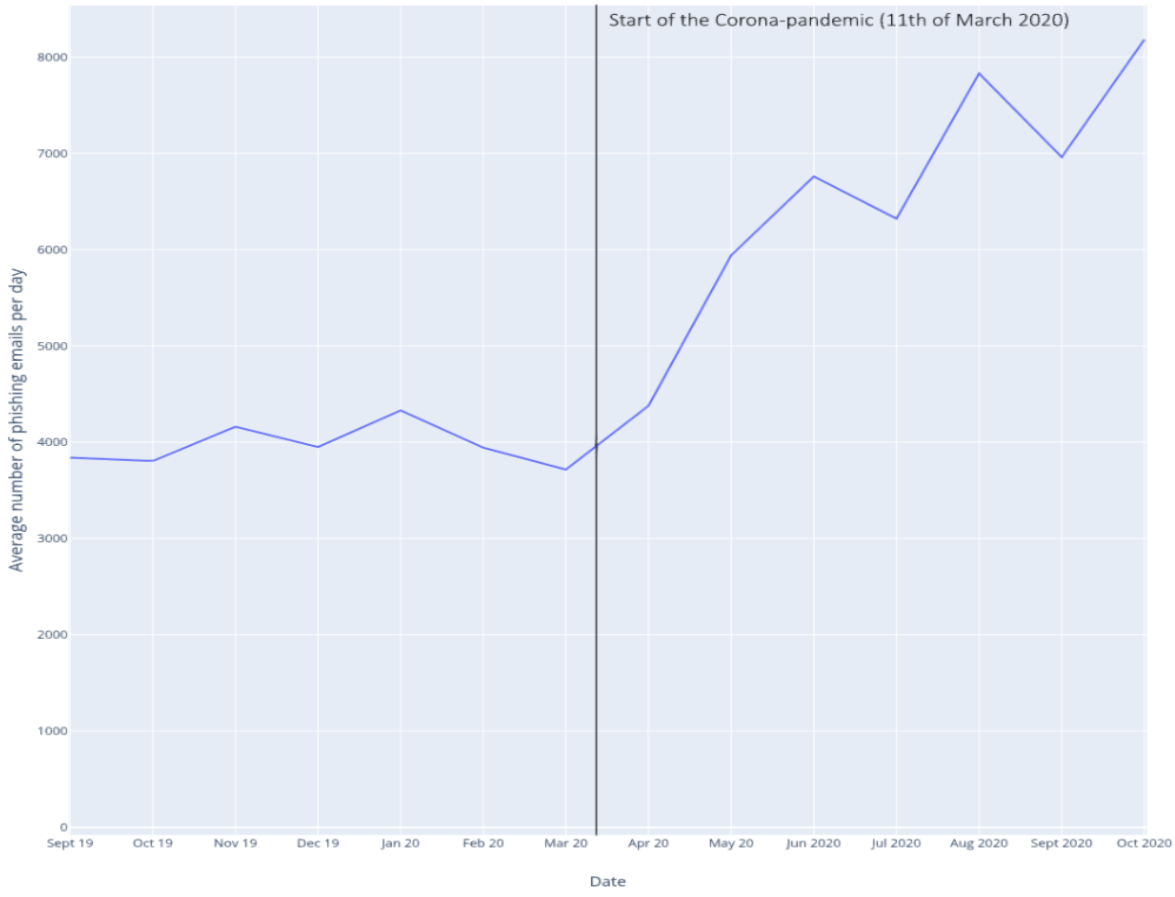
The number of phishing emails over time will be tested by a two-sample T-test to be tested for significant changes since the Corona-pandemic. A two-sample T-test is a method that is used to test whether the means of two groups are equal. In our case the two groups are

before and after the start of the Corona-pandemic. The proportions of emails that contain attachments are calculated for every day. The cut-off point is set to the 11th of March 2020 (the start of the Corona-pandemic). To test whether there has been a significant change in the proportion of phishing emails that contain attachments, a Mann-Whitney U Test for proportions is used. T-tests cannot be performed on the proportion emails that contain attachments, since these tests test averages between two groups. Finally, the results of the term frequency analysis will also be subject a two-sample T-test. The normalized count that will be tested for significant changes between the groups before the pandemic and after the pandemic. Both the Mann-Whitney U test and the two-sample T-tests will produce a p-value. This p-value will be compared to alpha (α=0,05). If the p-value, calculated by the Mann-Whitney U test and the two-sample T-tests respectively, is lower than this value then the groups differ significantly from each other. In other words, if the p-value is lower or equal to α=0,05, then a significant change has been observed in the data before the pandemic and the data after the pandemic.

Chapter 4 Results

The results show that there has been a significant increase in the number of phishing emails reported after the start of the Corona-pandemic on 11 March 2020. The number of reported phishing emails doubled from October 2019 to October 2020. The average amount of phishing emails before the start of the pandemic was 3998 per day. During the pandemic this increased to an average of 6483 emails per day. This signals that cybercriminals took advantage of the Corona-pandemic and the following lockdowns. In figure 6 is a graph showing the monthly average reported phishing emails.

Figure 6. Average number of phishing emails per day



The analysis shows that the increase was significant by running an independent T-test for the difference in the mean between the group before the pandemic and the group after 11th of March. The null hypothesis is that there is no change in the of the mean of phishing emails per day between before and after the pandemic. The T-test shows there is a significant increase $t(285,69) = 16.153$, $p < .01$. This is in line with previous studies that found an increase in cybercrime during the pandemic.

There was also a significant increase in the proportion of phishing emails that contained attachments. Before the pandemic, an average of 343 or 8,43% of the reported emails contained attachments. This started to increase after the start of the pandemic and

especially between May 2020 and September 2020 where an average of 1715 or 24,07% of the emails contained attachments. Using a Mann-Whitney U Test to test for difference between before and after the pandemic shows this increase is significant *U(before = 191, after = 244,) = 7252, p < .01*.

Corona related mentions

The keywords "corona", "disease", "pandemic", "virus" were used to test the category "Corona related mentions". Independent T-tests showed that all the keywords in this category increased significantly during the pandemic. The data shows that the terms "disease" and "virus" were already trending before the pandemic. This might be due to the terms not being only Corona-specific. However, the mentions of these keywords increased significantly during the pandemic. The term "corona" had a positive strong significant correlation with the terms "pandemic" *r(303) = .725, p < .01* and "virus" *r(303) = .658, p < .01*. These terms correlated with a lesser extend with the term "disease" *r(303) = .241, p < .01*. Contrary to "disease" and "virus", the terms "corona" and "pandemic" were mentioned only a few times before the start of pandemic. This started to increase around February 2020 and gained momentum during March and April 2020. The t-test showed that the mentions of "corona" and "pandemic" increased the most during the pandemic. The maximum of all the keywords, "corona", "disease", "pandemic" and "virus" were between 11 March 2020 (start of the pandemic) and the end of March 2020. After that time, the prevalence in the keywords went down. However, the mentions did not disappear and remained higher than before the pandemic. This can be seen in the graph regarding these terms.

Figure 7. Average "Corona-related" mentions per day

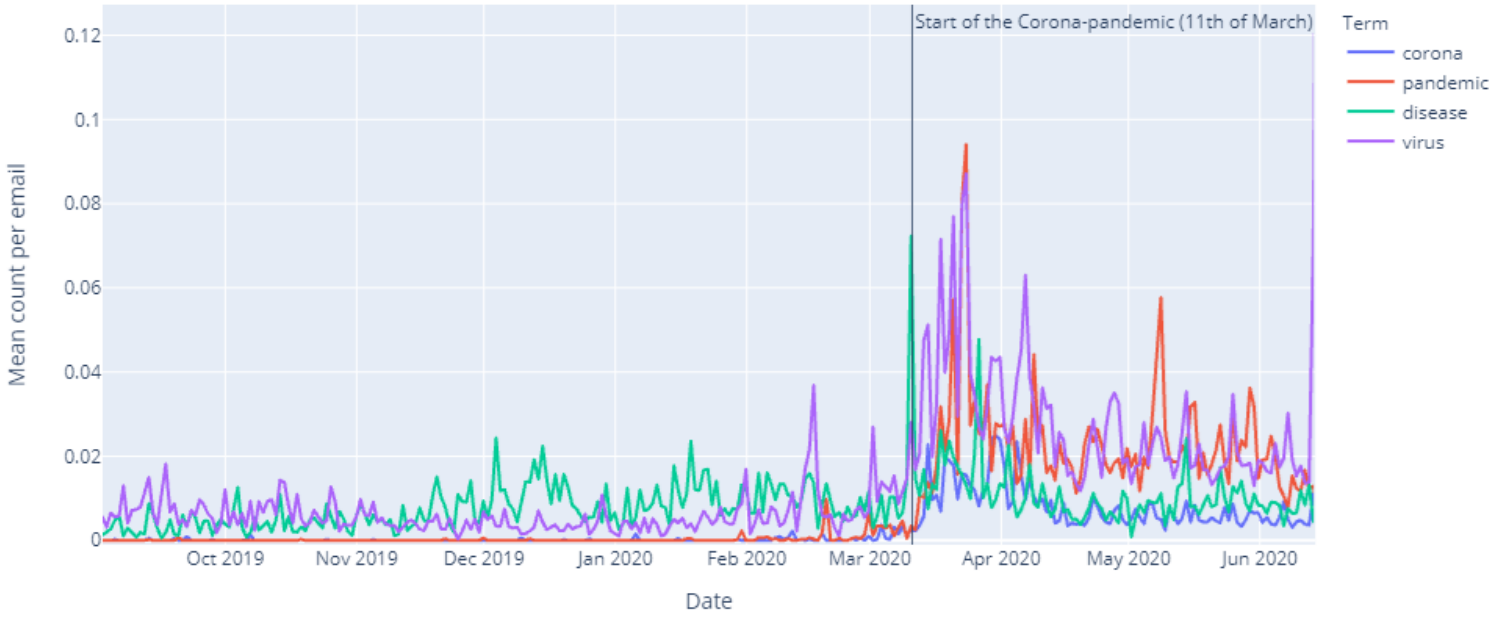


Table 4. Mean number of "Corona related mentions"

Note. in the column "keyword" are the keywords that are associated with the category "Corona-related mentions". The cut-off point in the analysis is chosen to be on the 11th of March 2020. On this day the WHO (2020) declared the start of the Corona-pandemic. The data is n=191 days before the pandemic and n=114 days after the pandemic for a total of 305 days. The mean value and the standard deviation of this data are presented. In the column "t-value", is the t-value that is calculated by comparing the mean before the pandemic and after/during the pandemic. In the column "prob" is the p value. The probability that is tested against an alpha of 0.05 for significance.

Keyword	Mean Standard Deviation	Days before pandemic (n=191)	Days after pandemic (n=114)	*t*-value	prob
"corona"	M SD	.000163 (.000520)	. 007377 (.005249)	-14.663	<.001
"disease"	M SD	.007435 (.004726)	.010608 (.008245)	--4.273	<.001
"pandemic"	M SD	.000295 (.001061)	.021545 (.012612)	-17.952	<.001
"virus"	M SD	.006093 (.004534)	.027203 (.22071)	-10.086	<.001

Medical and protection equipment

In the "Medical and protection equipment" category, medical and protection equipment are tested via the keywords: "chloroquine", "mask", "masks", "vaccines", and "ventilators". As can be read in table 5: the t-tests showed there is a significant increase of mentions of all the keywords. There is however a large difference in the amount that the average terms increase during the pandemic. The largest increase was in the keyword's "mask" and "masks". For comparison: "vaccines" was mentioned on average from 1.57 before the pandemic to 5.14 times per day during the pandemic. The keyword "masks" was mentioned on average from 2.4 times to 59 times per day. "chloroquine" and "ventilators" had even fewer mentions than "vaccines". It did show a significant increase in the average mentions, but this might be due to not being mentioned before the pandemic. The terms ventilator and chloroquine are left out in the graph due to the very few overall mentions. Here the terms "mask", "masks" and "vaccines" are plotted as normalized values.

Figure 8. Average "Medical and protection equipment" mentions per day

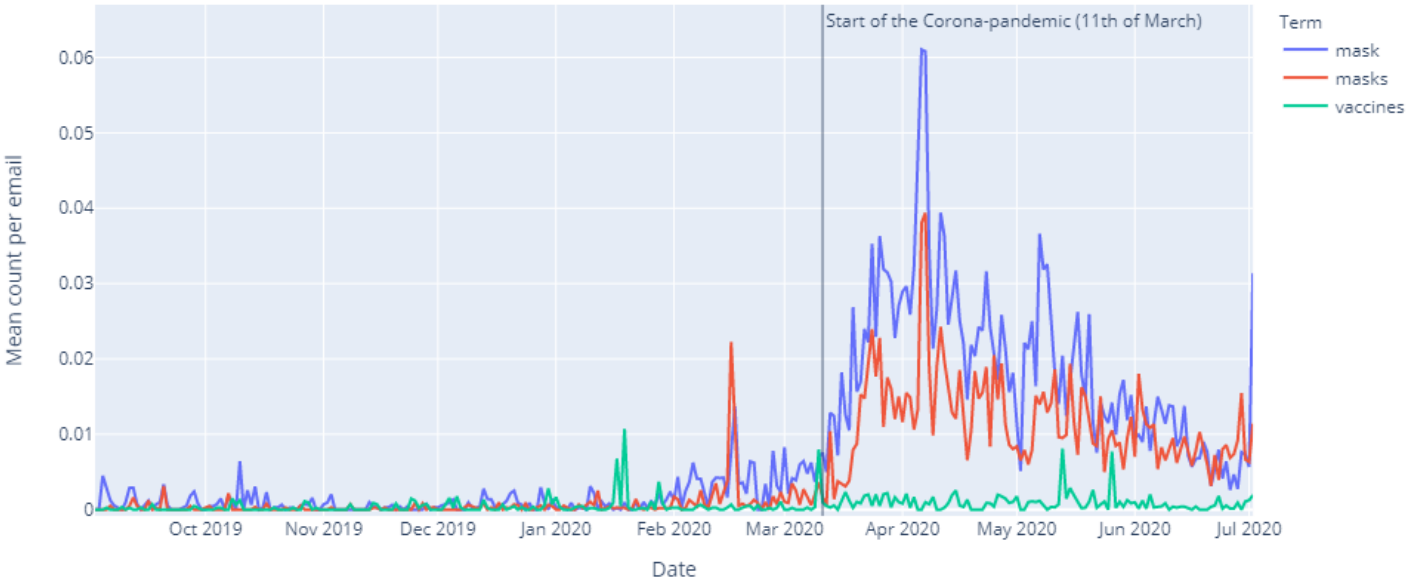


Table 5. Mean number of "Medical and protection equipment mentions"

Note. in the column "keyword" are the keywords that are associated with the category "Medical and protection equipment mentions". The cut-off point in the analysis is chosen to be on the 11th of March 2020. On this day the WHO (2020) declared the start of the Corona-pandemic. The data is n=191 days before the pandemic and n=114 days after the pandemic for a total of 305 days. The mean value and the standard deviation of this data are presented. In the column "t-value", is the t-value that is calculated by comparing the mean before the pandemic and after the pandemic. In the column "prob" is the p value. The probability that is tested against an alpha of 0.05 for significance.

Keyword	Mean Standard Deviation	Before pandemic (n=191)	After pandemic (n=114)	*t*-value	prob
"chloroquine"	M SD	0 (0)	.000209 (.000547)	-4.078	<.001
"mask"	M SD	.001424 (.001424)	.018698 (.011075)	-16.485	<.001
"masks"	M SD	.000633 (.001916)	.011625 (.0062145)	-18.370	<.001
"vaccines"	M SD	.000382 (.001151)	.000933 (.001172)	-4.002	<.001
"ventilators"	M SD	.000001 (6.0)	.000018 (3.9)	-5.922	<.001

Financials

In the "Financials" category the following keywords were tested: "bank", "donation", "payment", "donate", "debt", "invoice", "subscription". The T-tests show that all the terms increased significantly, except for the term "debt" as can be seen in table 6. The term "debt" decreased over time, but this was not a significant decrease. The biggest significant increase of average mentions was the term "donate", which is related to the term "donation". The second largest increase is the term "payment" which is related to the term's "bank", "donate" and "subscription" which all showed a significant increase and a significant moderate positive correlation with payment. The term "invoice" also showed a significant increase. Looking at the graph, it can be seen that "payment" and "bank" are especially prevalent in phishing emails looking at the normalized count, this then increased during the pandemic. In figure 9 the three most prevalent keyword are graphed: "bank", "donate" and "payment".

Figure 9.1 Average "Financial" mentions per day

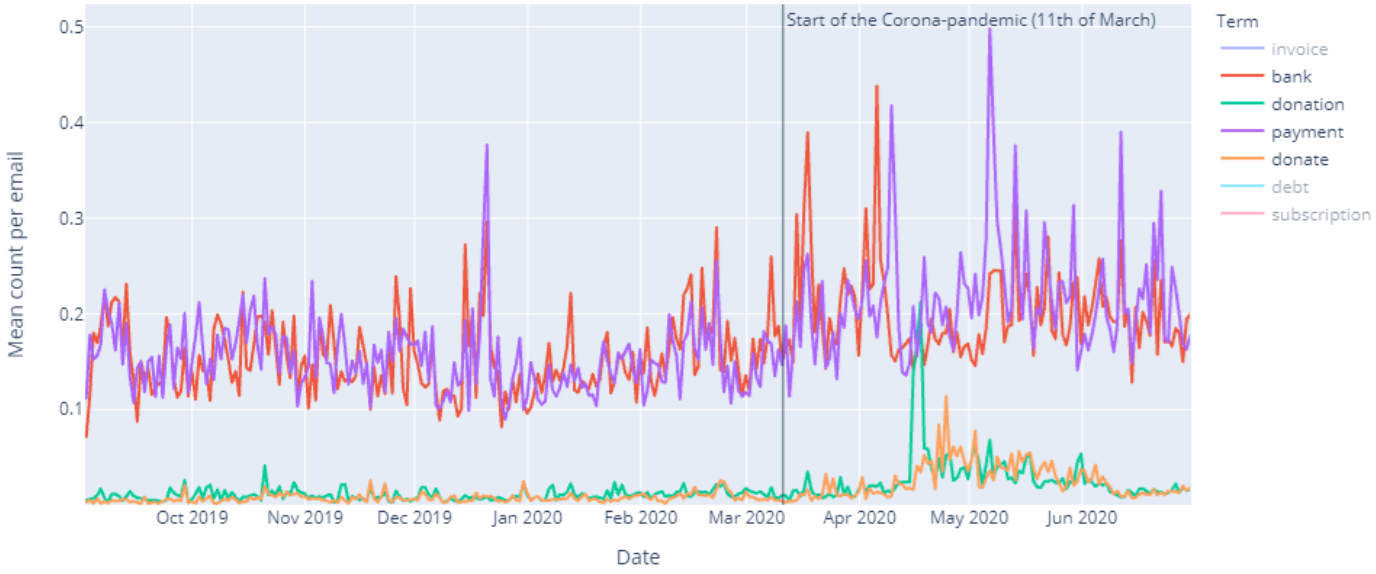


Figure 9.2 Average "Financial" mentions per day.

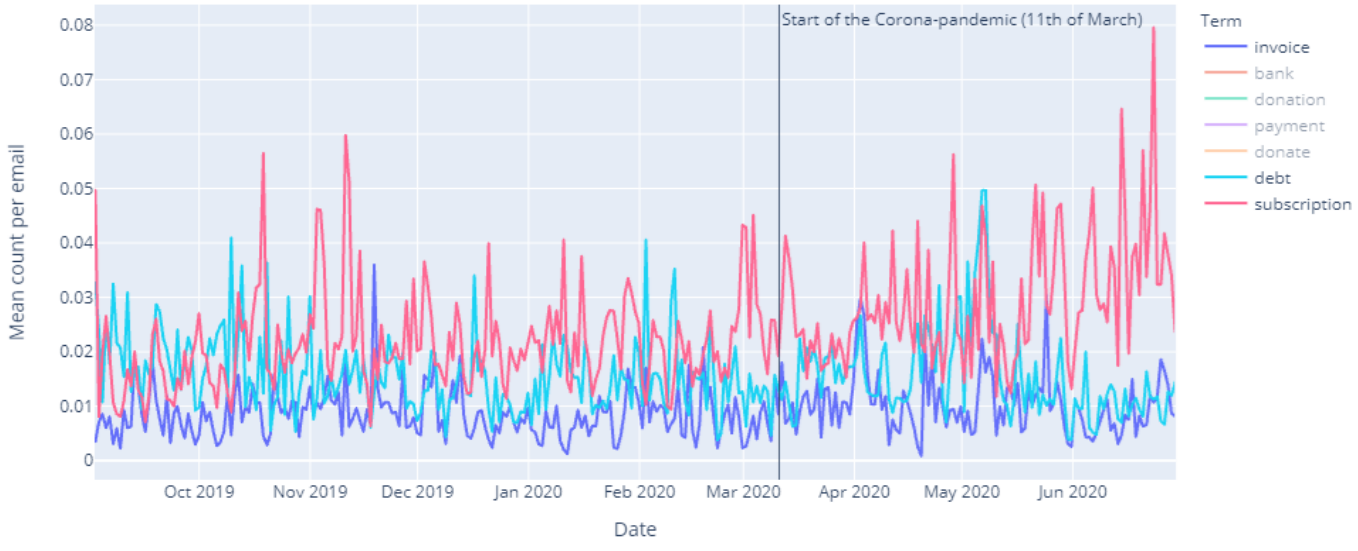


Table 6. Mean number of "Financial mentions"

Note. in the column "keyword" are the keywords that are associated with the category "Financial mentions". The cut-off point in the analysis is chosen to be on the 11th of March 2020. On this day the WHO (2020) declared the start of the Corona-pandemic. The data is n=191 days before the pandemic and n=114 days after the pandemic for a total of 305 days. The mean value and the standard deviation of this data are presented. In the column "t-value", is the t-value that is calculated by comparing the mean before the pandemic and after the pandemic. In the column "prob" is the p value. The probability that is tested against an alpha of 0.05 for significance.

Keyword	Mean Standard Deviation	Before pandemic (n=191)	After pandemic (n=114)	*t*-value	prob
"bank"	M	.152856	.201274	-9.015	<.001
	SD	(.039851)	(.048379)		
"donation"	M	.010639	.027874	-6.499	<.001
	SD	(.005586	(.027983)		
"donate"	M	.007341	. 025505	-10.172	<.001
	SD	(.004564)	(.018737)		
"debt"	M	.015789	.015668	.135	.892
	SD	(.006976)	(008390)		
"invoice"	M	.008503	.010314 (.005457)	-2.987	.003
	SD	(.004504)			
"subscription"	M	.021505	.029399	-4.002	<.001
	SD	(.008988)	(.011636)		
"payment"	M	.153333	.215297	-9.871	<.001
	SD	(.036922)	(.060645)		

Order and delivery scams

The data shows that all the terms concerning the category "Order and delivery scams" have increased significantly since the pandemic. This can be read in Table 7. The highest increase was found in the keyword "order", this went from averaging around 289 mentions per day before the pandemic to 647 mentions per day on average during the pandemic. The highest overall average mention was het keyword "account". It is interesting to note that "order" correlated with all the keywords. It has the highest correlation with "shipment" $r(303) = .38$, $p < .01$ and "delivery" $r(303) = .351$, $p < .01$. It also correlated, but to a lesser extent, to "amazon" $r(303) = .238$, $p < .01$ and "account $r(303) = .201$, $p < .01$.

Figure 10.1 Average "Order and delivery scam" mentions per day

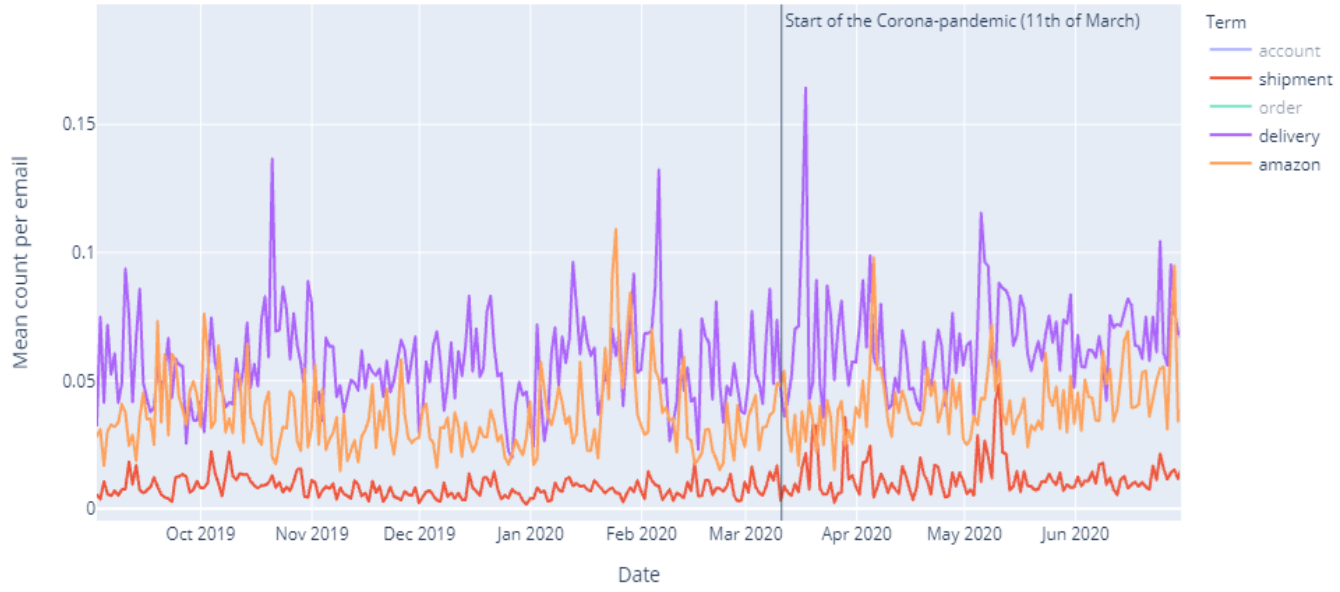


Figure 10.2 Average "Order and delivery scam" mentions per day

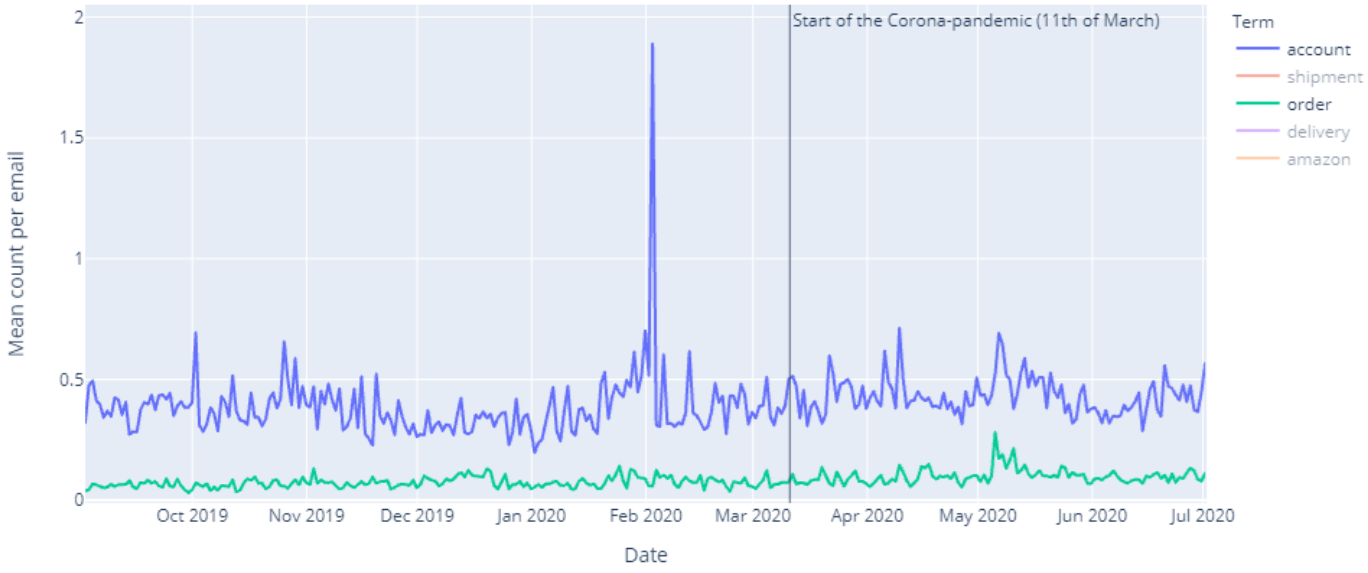


Table 7. Mean number of Order and delivery scam mentions

Note. in the column "keyword" are the keywords that are associated with the category "Order and delivery scam mentions". The cut-off point in the analysis is chosen to be on the 11th of March 2020. On this day the WHO (2020) declared the start of the Corona-pandemic. The data is n=191 days before the pandemic and n=114 days after the pandemic for a total of 305 days. The mean value and the standard deviation of this data are presented. In the column "t-value", is the t-value that is calculated by comparing the mean before the pandemic and after the pandemic. In the column "prob" is the p value. The probability that is tested against an alpha of 0.05 for significance.

Keyword	Mean Standard Deviation	Before pandemic (n=191)	After pandemic (n=114)	*t*-value	prob
"amazon"	M	.035570	.041401	-3.485	.001
	SD	(.014642)	(.013247)		
"account"	M	.382429	.433217	-3.552	<.001
	SD	(.140045)	(.078412)		
"delivery"	M	. 055758	.065810	-4,762	<.001
	SD	(.017317)	(.018674)		
"order"	M	. 0724	. 099729	-8.076	<.001
	SD	(.020961)	(.032297)		
"shipment"	M	.007967	.012276	-5.654	<.001
	SD	(.003754)	(.007605)		

Dating, beauty, and care

In the final category "Dating, beauty and care", keywords associated with other types of phishing were tested. The graph showing the results are split into two graphs. Figure 11.1 displays the trend of the terms concerning dating, figure 11.2 shows the trend of keywords related to dieting. The results in table 8 show that the keywords "dating", "diet", "fat", "meet", "single", "women" and "weight" have decreased significantly since the pandemic. Only the term "skin" did not differ significantly. The largest decrease of average mentions was in the term "dating". This went from around 725 average mentions per day before the pandemic to 34 average mentions per day after the pandemic. Terms such as "meet", "single" and "women" have a significantly positive correlation with "dating". This is expected as these terms can be related to each other. The term related to appearance/dieting that decreased the most were the terms "fat" and "weight" as can be read in table 8. These terms also correlated positively with each other $p(303) = .561, p < .01$ and "weight" with "diet" $p(303) = .482, p < .01$. This is also expected as these terms are also related to each other. In figure 11.1 are the terms associated with dating. In figure 11.2 are the terms associated with weight loss.

Figure 11.1 Average "Dating, beauty and care" mentions per day

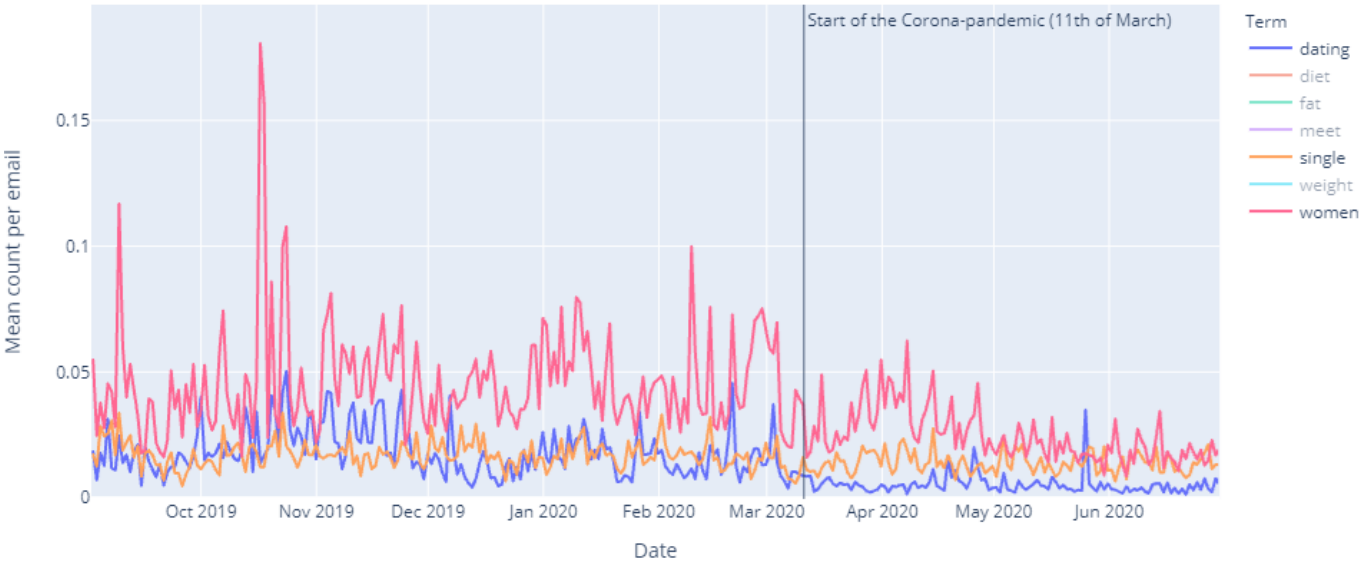


Figure 11.2 Average "Dating, beauty and care" mentions per day

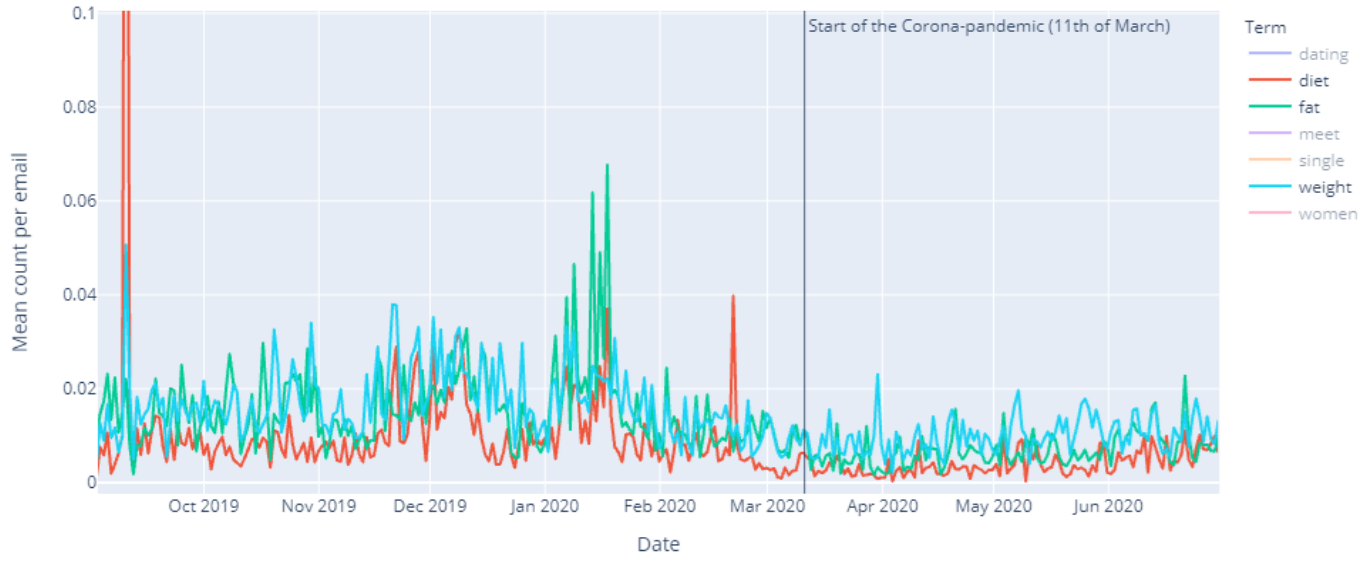


Table 8. Dating beauty and care mentions

Note. in the column "keyword" are the keywords that are associated with the category "Dating, beauty and care mentions". The cut-off point in the analysis is chosen to be on the 11th of March 2020. On this day the WHO (2020) declared the start of the Corona-pandemic. The data is n=191 days before the pandemic and n=114 days after the pandemic for a total of 305 days. The mean value and the standard deviation of this data are presented. In the column "t-value", is the t-value that is calculated by comparing the mean before the pandemic and after the pandemic. In the column "prob" is the p value. The probability that is tested against an alpha of 0.05 for significance.

Keyword	Mean Standard Deviation	Before pandemic (n=191)	After pandemic (n=114)	*t*-value	prob
"dating"	M	.018123	.005233	16.377	<.001
	SD	(.009577)	(.003984)		
"diet"	M	.011017	.004071	4.359	.001
	SD	(.021758)	(.002607)		
"fat"	M	.015662	.006911	12.063	<.001
	SD	(.008911)	(.003549)		
"meet"	M	.063196	.034315	11.262	<.001
	SD	(.031770)	(.012136)		
"single"	M	.017055	.013939	5.260	<.001
	SD	(.005458)	(.004137)		
"skin"	M	. 010408	. 008492	1.955	.052
	SD	(.009467)	(.005779)		
"women"	M	.046074	.024654	11.560	<.001
	SD	(.021715)	(.010486)		
"weight"	M	.016693	.009995	10.302	<.001
	SD	(.007525)	(.003792)		

Chapter 5 Discussion and conclusion

In this chapter the discussion, final conclusions, limitations, and suggestions for future work are presented.

Chapter 5.1 Discussion

This thesis peered into the implications of the Corona-pandemic on the way cybercriminals are using their phishing campaigns. As stated before, the WHO declared the start of the Corona-pandemic on 11th of March 2020. The conclusions are based before this initial start and after the start of the pandemic. Below are the conclusions derived from the results and followed by discussion points.

Number of phishing emails and attachments

The results indicate that cybercriminals are trying to take advantage of the Corona-pandemic by increasing and adapting their phishing campaigns. This is evident in the rapid increase in the number of phishing emails that were sent during the pandemic. From 3998 daily emails before the pandemic to 6483 daily emails during the pandemic. During the pandemic, many people were urged to work from home. Cybercriminals most likely wanted to make use of the increased number of people staying at home and being online. From the present results we cannot say whether traditional criminals resorted towards phishing or whether existing cybercriminals took it up a notch and increased the amount of phishing emails.

An interesting find is the great increase in the number of phishing emails that contained attachments. Since the pandemic, the number of phishing emails that contained attachments has seen an increase from 8.43% before the pandemic to 24.07% after the pandemic on average per day. Phishing emails that contain attachments can carry malware which can act as a vehicle for financial fraud Lallie et al. (2020). If the attachments contain malware, this can be in line with the finding that the amount of ransomware increased significantly since the pandemic. However, from the dataset we cannot conclude that these attachments are malicious. It is possible these attachments are harmless, like images. Besides the increase of the intensity of phishing emails during the pandemic, cybercriminals appeared to have responded and adapted their phishing campaigns in reaction to the Corona-pandemic.

Corona related mentions

It was found that phishing emails contained significantly more keywords from the Corona-related mentions than before the pandemic. Around the date the WHO declared the start of the pandemic (around 11th of March 2020), the keywords “pandemic” and “corona” peaked significantly and were hardly mentioned before this date. This signals that cybercriminals not only increased the number of phishing-mails since the pandemic, but actively adapted to the pandemic by using it as a theme in their phishing campaigns. It was not discovered how cybercriminals used this theme in their phishing emails, other than making use of the scarcity of medical masks as can be read in the section “Medical and protection equipment”. Previous studies suggested that phishing campaigns can be focused on impersonating health officials into tricking people to send credentials or open attachments. There was no evidence found for this claim. After the start of the pandemic on the 11th of March 2020 the keywords “virus” and “disease” started to increase, but earlier peeks can be observed. For instance, on the 1st and 17th of February 2020. This could be due to the ambiguity of the term “virus”, which can also refer to a computer virus. “Disease” had grown significantly since the pandemic, but to a lesser extent than the other terms. It also correlated the least with the other terms. It is possible

that “disease” was used in other types of phishing emails that used this term before the start of the pandemic.

Medical and protection equipment

There was an especially large increase in the number of phishing emails that contained the keywords: “mask” and “masks”. These keywords appeared in advertisements wanting or giving the impression to sell medical masks. The term “mask” and “masks” are most likely targeted for Corona-use since the phishing emails came mainly after the pandemic and made use of the scarcity of protection equipment like masks. Phishing emails containing the word “masks” often had a subject that read: “now 30% off for medical masks!”. The term “chloroquine” was mentioned very few times, but what is notable is that before the pandemic “chloroquine” never appeared in phishing emails. However little, after the pandemic it started to appear in phishing emails. The terms “vaccines” and “ventilators” also increase in the number of mentions in phishing emails, but there were still only very few mentions of these keywords. Especially compared to “mask” and “masks”. Perhaps the demand was higher for masks than for vaccines or ventilators. A peak in “mask” and “masks” before the start of the pandemic on the 16th and 17th of February 2020 can be seen. On the 17th of February 2020 there was also a sharp increase in the term “virus”. Perhaps there is a connection there, but there is no evidence for this. A peak in “vaccines” mid-January 2020 was found. Perhaps these spikes were caused by the lockdowns in China, but there is no evidence for this.

Financials

The results signal that cybercriminals already tried to gain revenue by faking financial emails before the pandemic. However, this increased significantly during the pandemic, signaling that cybercriminals adapted to the pandemic to gain even more revenue using phishing emails. Mainly by using terms around “payment”, “donation” and “bank”. To a lesser extend using terms like “invoice” and “subscription”. This signals that criminals increased their phishing-mails concerning payment and active subscriptions. These terms were already in use before the pandemic but increased during the pandemic. One can argue that the financial related phishing emails are the more traditional phishing emails. The exception is around the terms “donations” and “donate”. These terms did occur very little before the pandemic but started to increase during the pandemic. It is suggested by previous studies that cybercriminals tend to make use of crisis-situation in the form of faking charity and collecting donations from people. This seems obvious looking at the increased prevalence of “donate” and “donation”, but further inspection showed that the terms surrounding donations seemed to increase mainly around late April and May. This was also around the time that American elections started. The analysis shows that the terms surrounding donation correlated more with the term “Trump” than with “pandemic” or “corona”. The term “Trump” showed a strong positive correlation with the term “donation” and “donate”. The term “pandemic” a moderate positive correlation. This suggests that perhaps the increase in the mention of “donate” or “donation” was not due to the pandemic but had more to do with the election in the United States. The term “debt” was shown not to differ significantly, one could perhaps expect otherwise. Since many companies were on the brink of bankruptcy and many people lost their jobs during the pandemic.

Order and delivery scams

Due to the lockdowns, caused by the pandemic, people were forced to buy products online instead of going to the store. The results signal that cybercriminals tried to make use of this by launching phishing campaigns that impersonate online stores, such as Amazon. This is backed up by the keyword's "order", "delivery" and "shipment" that have increased significantly in the same period. Some phishing emails contained a warning about an order that has been cancelled, luring the victim into opening the phishing email and perhaps into downloading an attachment. These phishing emails often had subjects that read: "FW: order cancelled" or "Delivery of your boxes". The results signal that cybercriminals tried to benefit from the increased online orders by impersonating the company Amazon, faking orders, deliveries, account, and shipment details. There was a particular peak in the term "order" during May cannot be explained by the pandemic and the reason for this peak remains unknown. The term "account" had the highest overall mentions. This is perhaps due to the ambiguity in the meaning of the word, as the term had the lowest correlation with the other terms. Even though, there was a significant increase in this term, it might not relate to our category. This could also be true for "amazon". It had the second to last correlation between the terms. "Amazon" saw a significant increase since the pandemic, but "amazon" does not automatically have to mean the online webstore Amazon.

Dating, beauty, and care

There was an increase in the above categories. It was however expected that some categories became less prevalent since the pandemic. Even though the total amount of phishing emails that were reported have increased significantly since the pandemic, some categories became less prevalent even when correcting for the increase in phishing emails. The keywords in this category are less attractive for cybercriminals to misuse, perhaps due to the decreased chance of success. It is likely that phishing emails surrounding dating and meeting women has decreased due to the lockdowns and social distancing. Where many peaks in the term's "women", "single" and "dating" before the pandemic are seen, after the pandemic these peaks become rarer and less high. However, there is a peak in dating at the end of May 2020, which even peaks "women" for the first time. From the data gathered, it cannot be explained why this happened.

Significant decrease of mentions surrounding weight loss and dieting was also seen. This is perhaps due to the shifted interest of cybercriminals and the chance of success in the phishing campaign. Signaling that cybercriminals adapted to the pandemic by focusing less on appearances, weight loss and dating and perhaps tried to make use of the pandemic by using corona-related terms. It was hypothesized that cybercriminals would try to benefit from the pandemic by using keywords associated with the pandemic. Other types of phishing, concerning appearance and dating seem to have decreased since the pandemic since meeting in a pandemic is less appropriate and/or a priority.

Chapter 5.2 Main Conclusions

The motivation to do this research was to better understand how the Corona-pandemic affected cybercrime. Technology is still growing in importance in our daily life and in business. However, the threat of cybercrime is also growing in the last years and more people become victims. When we want to combat cybercrime more effectively it is essential to have the latest understanding of how cybercrime operates and how cybercrime evolves and reacts to worldly changes. A great change in the world is, at the time of writing, the Corona-pandemic. It has affected the daily life of many people and has forced us even more towards technology. Due to quarantines, the standard was working from home and meeting friends from home all via the internet. This has increased the potential victims to cybercrime tremendously. In understanding cybercrime better, we can address these issues more effectively.

The main research question is: What is the impact of the Corona-pandemic on the business model of cybercrime? The scope of this thesis was reduced to a subset of cybercrime: phishing. To answer this question, the sub-questions stated below were formulated. The first two sub-research questions are answered by knowledge gained in the literature review. These are the questions: “What is cybercrime?” and “What is the business model of cybercrime?”. These questions were answered by the means of the systematic literature review in chapter 2. The third sub-question is: “How can we empirically test the impact of the Corona-pandemic on phishing?”. This question was helped answered by the methodology which can be found in chapter 3. In this chapter approach to the analysis of the phishing dataset and the approach to statistical testing is explained. The fourth and final sub-question is answered by deriving conclusions from the results and arguments found in the discussion section.

1. *What is cybercrime?*

Cybercrime has many definitions and thus no one correct answer exists. Cybercrime comes in many forms: from phishing, DDoS-attacks, malware, ransomware to drug trafficking and sextortion. Sabillon et al. (2016) define cybercrime as any crime where computers and networks played an integral in committing the crime. The European Commission (2015) defined it as crimes done via the internet. Poonia (2014) hsd a similar definition of cybercrime: A crime done with a computer and/or the target is a computer or a system of computers. Huang et al (2018) stated that activities associated with cybercrime are so called “double edged sword activities”. This means that both an offensive and defensive are engaging in the same types of activities, but for different motives. The cybercriminals, the offensive side, are for instance looking for vulnerabilities in a security-system that they can exploit. Cybersecurity specialists, the defensive side, are also looking for vulnerabilities in a security system. Not to exploit this vulnerability, but to repair it before it can be exploited. However, Huang. Et al. (2018) and EUCPN (2015) express their concern in that the offensive side seem to have an edge over the defensive side. Meaning that cybercriminals tend to be a step ahead over the side that tries to prevent cybercrime.

Cybercriminals engage in cybercrime for various reasons. Cybercrime can be fueled by emotion and/or out of hatred, politics and/or religion, just for fun and for financial gain. The main conclusion that can be derived from the literature is that the main motive has shifted. From engaging in cybercrime “just for fun” to “acquiring financial gain”. Although, cybercrime “just for fun” does not have the main intend to gain money, it can cause damage.

It is motivated by excitement and entertainment, but if a large organization cannot operate due to these attacks it can cost a lot of money. Those motivated by financial gain do want to make a profit out of their cybercriminal activities. An example is the case of a ransomware-attack. During such attack, the entire network of an organization is being held "hostage". Files that may contain sensitive information are being locked until a ransom is paid. In many cases the organization comes to a halt until the organization pays the criminals ransom to get access to their files again. More elaborate examples are given in the second research question.

2. *What is the business model of cybercrime?*

The previous research question peered into what cybercrime is and gave various motives for people to commit cybercrimes. The shift toward financial gain has transformed how cybercrime operates. Cybercrime used to be for the technically skilled criminals, since launching a cyberattack requires great technical skill. However highly sophisticated business models have emerged in cybercrime such as cybercrime that is offered as a service model. Cybercrime has shifted from product to service-oriented business models. CaaS is a business model that offers different types of cybercrime as a service. In that way the technically skilled cybercriminals have become sellers of cybercrime. On illegal markets, one can order a complete ransomware or DDoS-attack without having any technical ability. This transformation has allowed less-technical criminals to engage in cybercrime.

The point is that cybercrime appears to be operating like an actual business. Where there are sellers and buyers of cybercrime products and services. For those needing help, customer service may be available. Cybercrime business models can roughly be modeled in two types of activities. The primary and secondary activities. The primary activities focus on the cyber-attack itself that is carried out. These activities are focused on finding weak spots in the system and what type of tools are used for the attack. Secondary activities are all the activities that facilitate the primary activities. This can be the illegal marketplace to buy the tools needed for an attack or a money laundering scheme to help launder the proceeds gained. The goal of secondary activities is to let the primary activities run as smooth as possible. Many, if not all, the primary and secondary activities are also available as-a-service. So, criminals can orchestrate an attack and buy services where needed to facilitate the attack. An example given in the literature review is the platform-as-a-service. This service allows for criminals to create their own underground marketplace where they can buy and sell their own selection of cybercrime products. Here, other services can be offered, like the product or service to buy a custom ransomware-kit or Moneymule-as-a-Service.

3. *How can we empirically test the impact of the Corona-pandemic on phishing?*

To test what the impact of the Corona-pandemic has been on phishing we used an appropriate dataset retrieved from the Anti-Phishing Working Group. It contained historical reported phishing emails from September 2nd 2019 to October 9th 2020.

The data was cleaned and transformed to make it suitable for analysis. A term frequency analysis was used to monitor trends in the data and how the Corona-pandemic might have had an impact on these trends. The number of reported phishing emails per day were taken into consideration. We monitored for any trend changes since the Corona-pandemic in five different categories. For each category associated keywords were formulated which were based on literature. The categories were: "Corona related mentions", "Medical and protection equipment", "Financials", "Order and delivery scams" and "Dating, beauty, and care". An overview of the keywords used per category can be found in table 2 in chapter 3. In order to identify changes in the topics/themes, we defined a definite timestamp before

the pandemic and after the pandemic. This timestamp was chosen on the 11th of March 2020. This was the date that the World Health Organization declared the start of the Corona-pandemic: following worldwide lockdowns. From there we observe how the topics behave before and after this timestamp. Then the normalized counts per day are retrieved before and after the timestamp. Finally, the data is analyzed in SPSS for any changes in the topics between the groups using a two-sample T-test. This is compared against the null-hypothesis which states there is no change in the topics before and after the start of the Corona-pandemic. Having this as a basis we can test what the impact of the Corona-pandemic has been on phishing.

4. What are the implications of the Corona-pandemic on phishing?

Based on the results, there several implications of the Corona-pandemic on phishing. The number of phishing emails increased significantly after the Corona-pandemic started. Other studies found that other types of cybercrime also increased drastically since the pandemic, and our findings suggest that phishing is no exception. Before the pandemic there is an average of 3998 reported phishing emails per day. This increased to an average of 6483 phishing emails per day during the pandemic. This even increased towards 7500 emails per day in August 2020.

Not only the number of phishing emails per day increased significantly, the proportion of emails that contained attachments had also increased significantly. From 8,43% on average before the pandemic to 24,07% on average during the pandemic. Or from 343 emails before the pandemic to 1715 emails on average per day during the pandemic that contained attachments. Attachments in phishing emails can act as a vehicle for malware and in turn for financial fraud, for instance via ransomware. However, it has not become clear whether these attachments were malicious.

It appears that the intensity of phishing emails is not the only way cybercriminals reacted to the pandemic. The results indicate that cybercriminals have adjusted their phishing-campaigns toward Corona-related topics. Especially terms associated with Corona have seen a large increase. This signals that cybercriminals focused their phishing-topics accordingly reacting to what was happening at that time. This can also be seen as various other topics like dating has seen a decrease as might be expected during a viral pandemic due to quarantine.

The largest increase was found in “Corona related” and “Medical and protection equipment” mentions. Before the pandemic, these topics seldom came up, but during the pandemic these topics increased. Phishing emails containing the Corona-related terms “pandemic”, “corona”, “mask” and “masks” increased the most. The topics “financials” and “order and delivery scams” also increased, but the difference is that these topics also occurred before the pandemic. Terms concerning the topic “Dating, beauty and care” decreased since the pandemic. Especially the terms related to dating. This is expected since the possibility for dating during a pandemic is more limited.

One can conclude that cybercriminals responded to the lockdowns by stepping up their game. This is evident with the significantly increased intensity of phishing emails since the pandemic. The theme that is present in the phishing emails appeared to have changed with what is happening in the world. Since the pandemic, cybercriminals adapted their phishing themes in such a way that it fits current events. Corona-related terms, medical and protection equipment saw a big uprise in presence in the emails. Cybercriminals also adapted their

phishing campaigns to the lockdowns and buying products online by increased emails concerning financials, order, and delivery. But decreasing the phishing emails concerning dating and dieting.

Chapter 5.3 Limitations

The data that is used to study the keywords ran from September 2019 and July 2020. The dataset ran from September 2019 to October 2020, but the analysis ran into error when it reached the end of July 2020. The error that was received was: "KeyError: None of [Index['Email_Body'], dtype='object')] are in the [columns]". The analysis was done on building a vocabulary of the emails and this was extracted from the key "Email_Body" which contained the email itself. However, from the 3rd of July 2020 onward get the error that there is nothing to extract from the data-key "Email_Body". We were not able to solve this error, therefore it was chosen to remove these from the analysis.

Results and trends found by the analysis gives information on how the Corona-pandemic affected the way how criminals focus their phishing mail campaigns. This type of analysis maps the occurrence of keywords and from there a trend can be inferred. However, it cannot dive deeper into changes in the business model of cybercrime. It is also important to note that a limited scope of phishing is captured by this dataset. Other methods of phishing occur by redirecting victims to fake websites, which is not done only by phishing emails. Phishing is also on the rise via other techniques such the mentioned WhatsApp fraud.

The analysis of the dataset Certain increases and decreases were found where conclusions are based upon. It is however possible that certain keywords or categories show a certain cyclical effect where increases and decreases come and go every year. It is therefore possible that effects attributed to the Corona-pandemic are the result of cyclical events. A dataset with a larger timescale can be used to correct for this. However, keywords that were not found before the pandemic are most likely the result of the pandemic. These keywords include "mask", "masks", "pandemic" and of course "corona".

Another consideration is the risk of suggestion bias. The trends that are found in the analysis were based on keywords. This only reveals trends in keywords and perhaps categories that were suggested and found in existing literature. It is possible that the keywords suggested do not represent the topic in question properly. Perhaps if different keywords were chosen for "Financials", this might have resulted in different outcomes. A more advanced method in machine learning called topic modeling can yield better and more reliable results. Instead of suggesting keywords, topic modeling allows for machine-learning to suggest topics or categories. This approach takes the whole dataset and suggests the most relevant topics and is thus free of suggestion bias. Topic modeling also captures topics, which were perhaps missed by only searching for specific keywords.

Chapter 5.4 Future work

Our suggestion for future work relates to the limitations discussed in the previous part. To get a full picture of how phishing evolves over time, one can make use of a dataset that extends over a longer period. This can result in firmer conclusions about what the impact of the Corona-pandemic has been by correcting for cyclical effects.

Future research can peer into how certain sectors are affected by phishing. Instead of focusing on how theme's/topics in phishing campaigns shift over time, future research can peer into how various sectors are targeted by phishing. For instance, have cybercriminals

focused their phishing campaigns more towards the technology sector or the health services during the Corona-pandemic? Have certain sectors since see a relieve in targeted phishing campaigns? And how does this evolve over time?

Because the results are timeseries, future research can make use of a more sophisticated model when testing the effect of a large event on the methods of phishing done by cybercriminals. This can be done by using a dataset which also contains non-phishing emails. This could be a model based on logistic regression that is able to predict whether an email is a phishing email or not by using certain combination of words. It is possible that before a large crisis, certain keywords or combination of keywords called X are good predictors in giving a probability that a given email is a phishing email. Then, a structural break can appear when combinations of X are no longer significant predictors and suddenly combination of keyword Y become a significant predictor. Combinations of X and Y should ideally be given by topic analysis. This gives indication about a change in the method of phishing. An example of such structural break can be found in the appendix in figures 16.1 and 16.2.

To get rid of suggestion bias as much as possible, we can make use of methods like topic modeling found in Natural Language Processing (NLP). Natural Language Processing can be a big help when it comes to large datasets that involve language. Latent Dirichlet Allocation or LDA is a common type of topic modeling that reverses the approach to analyzing topics than what has been done in this thesis. It is important to work with cleaned data. So, a vocabulary/ bag of words of the dataset is created to further work with. From here we run the LDA model where we can set the number of topics we expect in the text. LDA extracts major themes found in large quantities of text. LDA does this by finding common trends in text, like how often do certain words appear together and the probability that words belong to a certain topic. LDA does not suggest the topic, only the words that are associated with the assumed topic. Below are the topics and some graphs when LDA was performed on the phishing dataset used in the thesis. Most topics are too vague to be understood what they mean. The LDA shows 10 topics with 15 associated words per topic. The graphs that are shown are the topics that vaguely can be understood where they are about. The graphs are found in the appendix.The graphs can be interpreted in the following manner: the LDA analysis is done in batches of 3300 emails. The X axis shows the number of the batch and can therefore be loosely interpreted as time. The amount of batches times the emails per batch then amounts to the total number of emails in the dataset.

Literature

An, J., & Kim, H.-W. (2018). *A Data Analytics Approach to the Cybercrime Underground Economy*. *6*, 17.

Bracci, A., Nadini, M., Aliapoulios, M., McCoy, D., Gray, I., Teytelboym, A., Gallo, A., & Baronchelli, A. (2020). The COVID-19 online shadow economy. *ArXiv:2008.01585 [Physics]*. http://arxiv.org/abs/2008.01585

Chawla, M. (2014). A Survey of Phishing Attack Techniques. *International Journal of Computer Applications*, *93*(3), 4.

Deloitte. (2018, december). Black-market ecosystem Estimating the cost of Pwnership. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-risk-black-market-ecosystem.pdf

European Crime Prevention Network. (2015). Cybercrime: A theoretical overview of the growing digital threat. EUCPN. https://eucpn.org/sites/default/files/document/files/theoretical_paper_cybercrime_.pdf

Fireeye. (2012, September). *Top Spear Phishing Keywords Used in Spearphishing Attacks: Successfully Compromise Enterprise Networks and Steal Data*. https://www.fireeye.com/content/dam/fireeye-www/global/en/current-threats/pdfs/rpt-top-spear-phishing-words.pdf

Huang, K., Siegel, M., & Madnick, S. (2018). Systematically Understanding the Cyber Attack Business: A Survey. *ACM Computing Surveys*, *51*(4), 1–36. https://doi.org/10.1145/3199674

Hummel, R., Hildebrandd, C., Modi, H., Dobbins, R., Bjarnason, S., Belanger, J., & Arzamendi, P. (2020). *Netscout Threat Intelligence Report: Cybercrime: Exploiting a Pandemic.* Netscout. https://www.netscout.com/threatreport

Interpol. (2020a, May). Global Landscape On COVID-19 Cyberthreat. https://www.interpol.int/content/download/15217/file/Global%20landscape%20on%20COVID-19%20cyberthreat.pdf

Interpol. (2020b, August). Cybercrime: Covid-19 Impact. Interpol. https://www.interpol.int/content/download/15526/file/COVID-19%20Cybercrime%20Analysis%20Report-%20August%202020.pdf

Khan, N. A., Brohi, S. N., & Zaman, N. (2020). *Ten Deadly Cyber Security Threats Amid COVID-19 Pandemic* [Preprint]. https://doi.org/10.36227/techrxiv.12278792.v1

Kiru, M. U., & Jantan, A. B. (2019). The Age of Ransomware. *Artificial Intelligence and Security Challenges in Emerging Networks*, 1–37. https://doi.org/10.4018/978-1-5225-7353-1.ch001

Lallie, H. S., Shepherd, L. A., Nurse, J. R. C., Erola, A., Epiphaniou, G., Maple, C., & Bellekens, X. (2020). Cyber Security in the Age of COVID-19: A Timeline and Analysis of Cyber-Crime and Cyber-Attacks during the Pandemic. *ArXiv:2006.11929 [Cs]*. http://arxiv.org/abs/2006.11929

Leukfeldt, E. R., & Yar, M. (2016). *Applying Routine Activity Theory to Cybercrime: A Theoretical and Empirical Analysis*. 19.

Li, X. (2017). A Review of Motivations of Illegal Cyber Activities. *Kriminologija & Socijalna Integracija*, *25*(1), 110–126. https://doi.org/10.31299/ksi.25.1.4

Meland, P. H., Bayoumy, Y. F. F., & Sindre, G. (2020). The Ransomware-as-a-Service economy within the darknet. *Computers & Security*, *92*, 101762. https://doi.org/10.1016/j.cose.2020.101762

Netscout. (2020, October 1). *Netscout Threat Intelligence Report Shows a Dramatic Increase in Multivector DDoS Attacks in First-Half 2020.* https://www.netscout.com/netscouts-threat-intelligence-report-1H2020

NOS. (2020a, June 11). *Veel meer meldingen over phishing uit naam van de Belastingdienst.* https://nos.nl/artikel/2336887-veel-meer-meldingen-over-phishing-uit-naam-van-de-belastingdienst.html

NOS. (2020b, November 5). *Voor het eerst bouwer van phishing-software gearresteerd.* https://nos.nl/artikel/2355325-voor-het-eerst-bouwer-van-phishing-software-gearresteerd.html

Ritter, T., & Pedersen, C. L. (2020). Analyzing the impact of the coronavirus crisis on business models. *Industrial Marketing Management*, *88*, 214–224. https://doi.org/10.1016/j.indmarman.2020.05.014

Sabillon, R., Cano, J., Caveller, V., & Serra, J. (2016). Cybercrime and Cybercriminals: A Comprehensive Study. *International Journal of Computer Networks and Communications Security*, *4*(1), 165–176.

Trautman, L. J., & Ormerod, P. (2018). Wannacry, Ransomware, and the Emerging Threat to Corporations. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3238293

UNC Health Sciences Library. (n.d.). LibGuides: *Systematic Reviews: Write the Review. Health Sciences Library*. Retrieved March 13, 2021, from https://guides.lib.unc.edu/systematic-reviews/PRISMA

VMware Carbon Black. (2020, November 11). *Global Incident Response Threat Report: The Cybersecurity Tipping Point*. https://www.carbonblack.com/resources/tipping-point-election-covid-19-create-perfect-storm-cyberattacks/

Vu, A. V., Hughes, J., Pete, I., Collier, B., Chua, Y. T., Shumailov, I., & Hutchings, A. (2020). Turning Up the Dial: The Evolution of a Cybercrime Market Through Set-up, Stable, and Covid-19 Eras. *Proceedings of the ACM Internet Measurement Conference*, 551–566. https://doi.org/10.1145/3419394.3423636

Wegberg, R. S., Klievink, A. J., & van Eeten, M. J. G. (2017). Discerning Novel Value Chains in Financial Malware: On the Economic Incentives and Criminal Business Models in Financial Malware Schemes. *European Journal on Criminal Policy and Research*, *23*(4), 575–594. https://doi.org/10.1007/s10610-017-9336-3

World Health Organization. (2020, 17 maart). *Events as they happen*. WHO. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen

Appendix

Table 9. Topics (10) found by LDA with associated keywords/terms (15)

Topic number	Keyword/terms
Topic 0	price shanghai fob mask min disposable ply washable valve reusable blue anna fashion kids factory
Topic 1	grime gallery socks racks trampolines sheds morris wreaths gazebos curated birdbaths annuity joey illnesses species
Topic 2	glass insulator antique survey app calcium responses mint olive spins casino beta wireless abba match
Topic 3	para con anger och med del module ser att din till som dig relaxation killer
Topic 4	dos spy spying exchanger browsing webcam everywhere wallet recorded gall infected rat capers benne huff
Topic 5	wed bank payment impact funds fund money able work card security transfer jun without thanks
Topic 6	pang tritium canvas prints antenna sienna custom boycott anti brightest refinery regards discount sprint barrels
Topic 7	pour par pas plus photo est energy pays hong aux son gaz sex ans dune
Topic 8	repetitive spamming scam scams links addresses randomly noticed seems skin dangers stop block sheer combinations
Topic 9	fat gadgets diet burn tech pill auction skinny hidden hottest habit regrowth belly sweeping autographed

Figure 12. LDA Topic 0

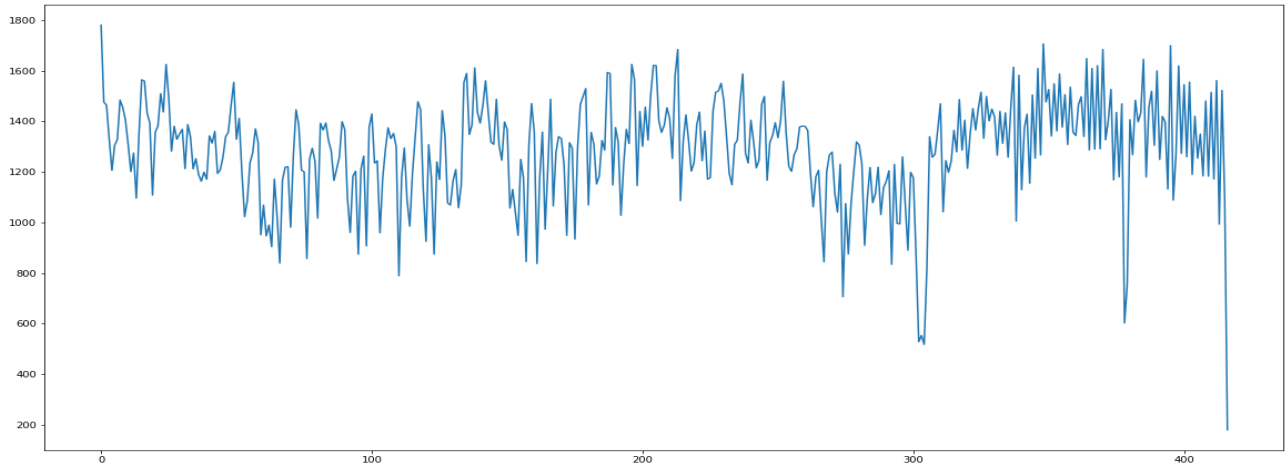


Figure 13. LDA Topic 4

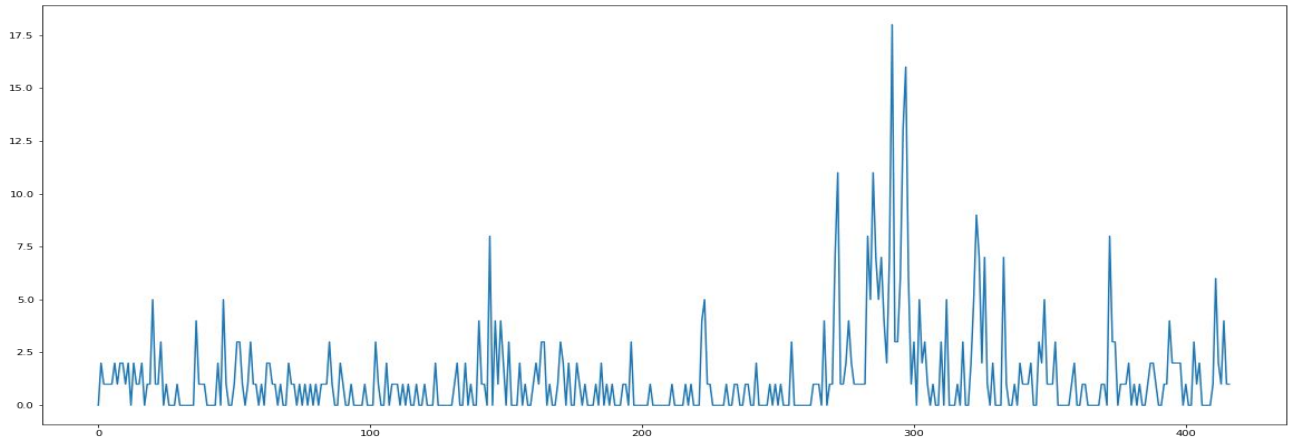


Figure 14. LDA Topic 5

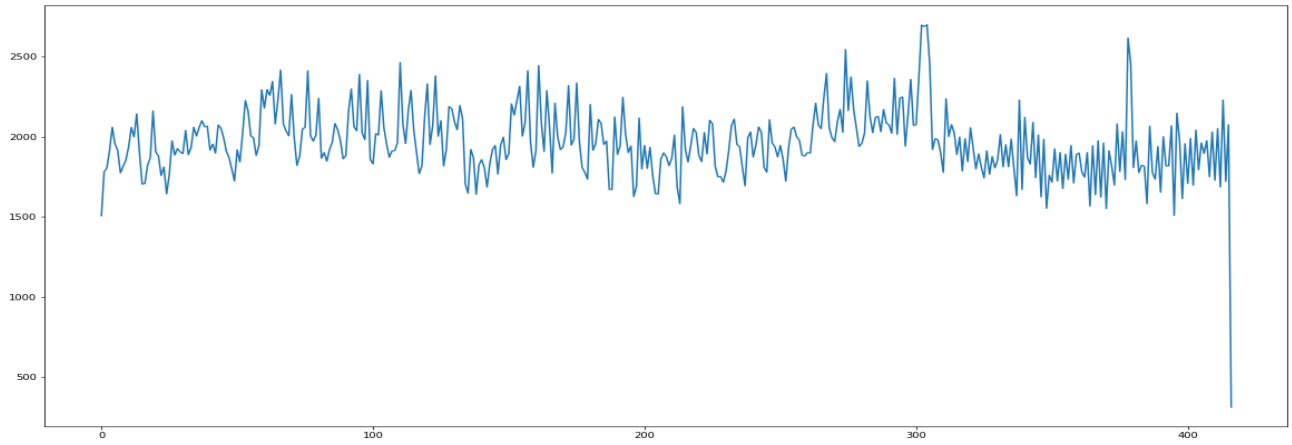


Figure 15. LDA Topic 8

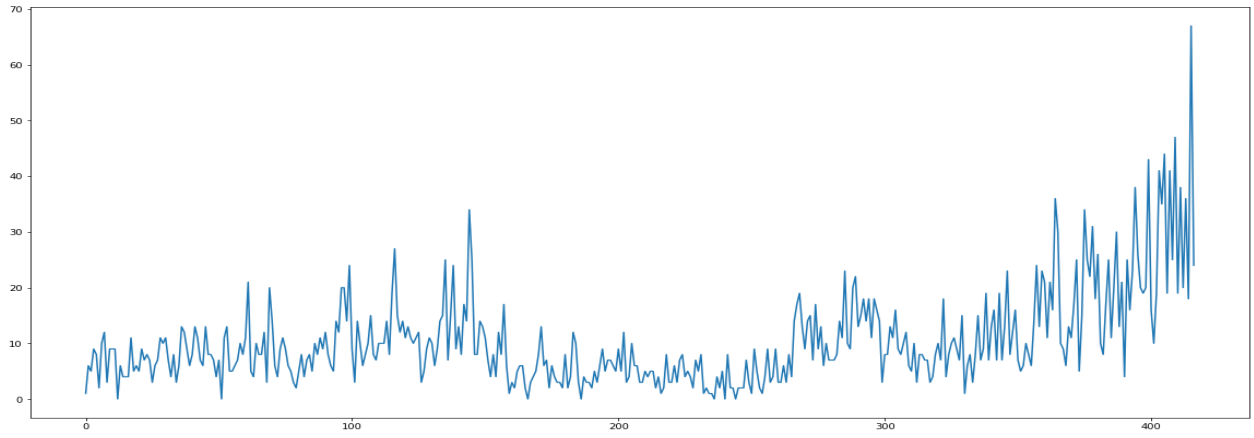


Figure 16.1 Structural break "bank"

Below is an example of a structural break. The red lines represent what the model can predict. This expands over time since the prediction becomes less certain over time. Every point (black line) that is between the red lines can be explained by the model proposed (for instance a logistic regression model). When the points go beyond the red lines, the results are no longer predicted by the model and thus a structural break appears. For the term "bank" the results are predicted by the model, however the term donate (figure 16.1) goes beyond the red lines and cannot be predicted by the model and is thus a structural break. One should note that this data is retrieved from the analysis done in this thesis. The graphs are for example purposes only. The structural break is only tested around the means of terms of a given day and thus does not yield usable information.

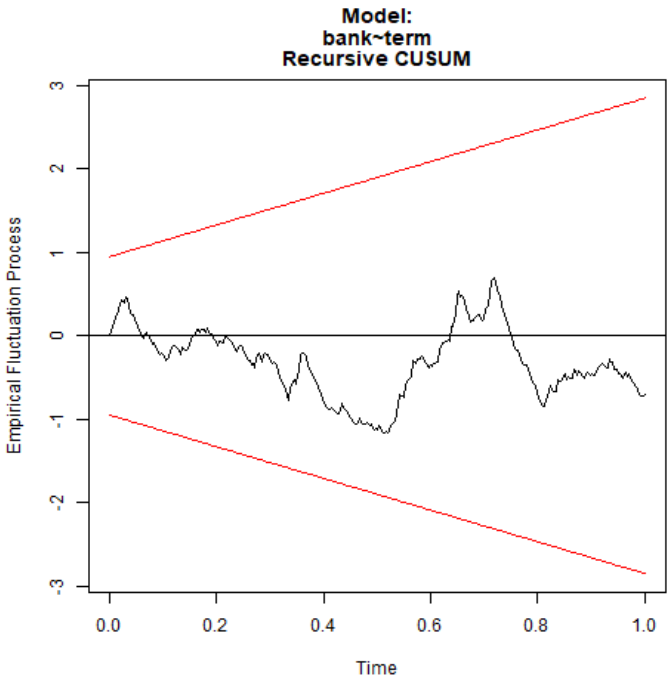


Figure 16.2 Structural break "donate"