Investigating Target Selection and Financial Impact of Service Fraud

An empirical research into criminal activities on underground markets and their implications for businesses

Master Thesis

April 2020

Elena Tsvetkova

TU Delft

Investigating Target Selection and Financial Impact of Service Fraud

An empirical research into criminal activities on underground markets and their implications for businesses

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Management of Technology**

Faculty of Technology, Policy and Management

by

Elena Tsvetkova

Student number: 4739396

To be defended in public on April 29th 2020

Graduation committee

Chairperson : Prof.dr.ir. P.H.A.J.M. van Gelder, Section Safety and Security Science
First Supervisor : Dr.ir. C. Hernandez Gañan, Section Organisation and Governance
Second Supervisor : Dr. H. Asghari, Section Organisation and Governance

Acknowledgement

I would like to express my gratitude towards the members of my thesis committee, without whom this project would not have been possible: Prof. Dr. Pieter van Gelder, Dr. Carlos Gañan and Dr. Hadi Asghari. I would like to thank Carlos, my first supervisor, for providing me with the opportunity to work on such research, and for the guidance and feedback he offered whenever I was in doubt during the research process. The commentary and suggestions from Pieter and Hadi have further helped to improve the final version of this thesis. I want to thank them all for their support and understanding throughout this period.

In addition, I would like to thank my friends and especially my family, for their unwavering support and encouragement during the process of writing this thesis and throughout the entirety of my studies.

Elena Tsvetkova
The Hague, April 2020

Executive Summary

Cyber security can be considered a major issue nowadays, with breaches in security costing millions of dollars annually. One of the many aspects of cybercrime is the creation and exchange of malicious software, fraudulent information, and other potentially harmful goods and services in the dark web. The impacts of such activities include not only financial losses, but also psychological trauma for individuals, and reputational and brand damage to organizations. A portion of the digital fraud happening in the dark web comprises the illegal exchange of vouchers, coupons, and stolen accounts. Despite its existence, this type of fraud, along with the factors that influence target selection have not been previously examined, which is the purpose of this research.

Acquiring an idea of the impact of service fraud is the first step of evaluating the gravity of the issue and taking measures to mitigate its effects. This subject has only briefly appeared in the literature, and mostly in reports of the security industry, no scientific papers which focus on it were found. However, by exploring data gathered from the underground markets in the dark web, it is evident such fraud exists and affects various industries and organizations. Therefore, this thesis tries to provide an initial understanding of the matter by mapping out the role of service fraud in the digital fraud landscape; exploring the different types of such fraud; quantifying the direct costs related to it and finally, providing an explanatory model explaining the company factors which influence cybercriminals in the target selection process. Hence, the goal of this research is to examine **what company characteristics influence the likelihood and the financial impact of service fraud conducted on the underground markets in the dark web**.

The first part of this project consists of a detailed overview of the relevant literature in order to provide clear understanding of service fraud. The dynamics of the illegal underground trade are explored, along with the credential theft lifecycle. A criminal business model for service fraud is suggested, describing its value chain and main attack vectors and targets. Furthermore, literature in the field of target selection is consulted to identify characteristics which could influence the number of fraud incidents experienced by companies. Studies examining target selection in the area of banking malware or data breach incidents were the most relevant sources discovered. After detailed research, various factors are proposed based on their relation to the value, visibility and accessibility aspects of target suitability as described in the Rational Activity Theory. The final selection of characteristics is based on their prominence and applicability for the analysis and consists of: *company size*, as measured by: *revenue, net income, total assets, total equity, and number of employees; company reputation; domain popularity,* and *area of service*.

Following the literature study, the examined dataset is processed and cleaned manually since it includes inconsistent data. The initial dataset contains information about 44,671 listings and 564,204 transactions made on digital goods from eight underground markets, for the years from 2011 to 2017, and includes data about product category, date of transaction, marketplace, vendor, order amount. After irrelevant data is removed, the remaining 2,318 listings are sorted into four categories depending on the type of item: account, loyalty program, pirated software, and voucher; and linked to the company which is affected by the fraud. Examination of the most intensely-targeted companies led to the selection of 47 entities to be studied further.

Additional data is collected or extracted on the basis of the outcome of the first part of the analysis, so that the target frequency and the losses of the examined companies can be assessed. After the costs incurred by companies were quantified, it became evident that the sustained losses are relatively low compared to the figures reported in various resources.

Data for the previously selected company characteristics is also gathered for the chosen businesses. All supplementary datasets required extensive data preparation in order to produce variables suitable for the analysis: multiple imputation, data transformation, and principle component analysis were employed to deal with the various issues arising from the data.

Explanatory analysis is performed to explore whether the collected external variables: *company size, reputation, domain popularity, location, relative price*, are related to the target frequency and total losses incurred by companies. The analysis consists of the generation of two types of regression models. Firstly, a negative binomial regression is used to model the relationship between the independent variables and the frequency of being targeted. The second model is a multiple linear regression aiming to establish whether the losses suffered by businesses from service fraud can be explained by the main company characteristics. Both models employ the variables category and mean lifespan of items as control variables.

Despite the various limitations presented by the data and the decisions which had to be made, some of the company characteristics seem to have an influence on the examined variables. The visibility factors: reputation and domain popularity, are able to explain to an extent the frequency of being targeted. Furthermore, companies operating locally, as well as smaller businesses seem to have higher financial losses. This could be due to the more limited security capabilities of smaller companies, or their lack of awareness of the issue. An explanation related to the examined theory is provided as to why visibility seems to explain target frequency better than company value.

Based on the results of the analysis and the present limitations, several recommendations are suggested. Expanding the current model with other variables or additional data could lead to more reliable results. The current analysis considers only the value and visibility aspects of the RAT model, while it has been reported that the accessibility aspect is of growing importance in cybercrime. Exploring characteristics related to accessibility, such as employed security controls and strategies, would undeniably enhance this study. Moreover, different research methods can be utilized in order to gain insights which would be hard to capture with solely quantitative analysis.

Lastly, in spite of the various limitations and restrictions of the current study, it can nevertheless serve as a basis for future research, or as means of raising awareness about a topic which is only now gradually gaining attention.

Contents

List of Figures

List of Tables

Abbreviations

EMCDDA	European Monitoring Centre for Drugs and Drug Addiction
ENISA	European Union Agency for Cybersecurity
FAIR	Factor Analysis of Information Risk
GDPR	General Data Protection Regulation
IRC	Internet Relay Chat
ITU-T	International Telecommunication Union Standardization Sector
NIST	National Institute of Standards and Technology
NPV	Net Present Value
PCA	Principal Component Analysis
PII	Personally Identifiable Information
RAT	Routine Activity Theory
RCT	Rational Choice Theory
ROA	Return on Assets
ROI	Return on Investment
ROSI	Return on Security Investment
SD	Standard Deviation
TARA	Threat Agent Risk Assessment
VIF	Variance Inflation Factor

1 Introduction

Over the past few years, reports and articles revealing the costs incurred by businesses through cybercrime are becoming more and more common, with values ranging from insignificant figures to millions and billions of dollars (Anderson, 2013). At the same time, the main sites where cybercriminal goods and services can be obtained: the underground markets in the dark web, have been steadily evolving. They are no longer the chaotic and poorly regulated chat forums of the past, but rather sophisticated platforms turning criminal activity into a commodity, easily accessible by consumers.

While the biggest amount of goods traded on these markets are drugs, the exchange of stolen financial information, identity theft, fake accounts, as well as various types of crimeware also have a stable share. A smaller portion of these goods can be assigned to a category comprising miscellaneous fraudulent items: stolen accounts, vouchers, coupons, gift cards, loyalty program accounts. This particular category, which will be further referred to as service fraud, as it concerns fraud with various services from media streaming to online retail, is at the focus of the conducted study.

Businesses are constantly trying to keep up with the trends and developments in customer preferences, introducing various digital offers and services to increase their appeal to users. Nonetheless, it is not uncommon that such innovations are implemented before security controls and practices have been brought up to date with these developments, creating more opportunities for cybercriminals. With the growing amount and value of personal data, it is of no surprise credential theft is becoming more and more common. Consequently, part of the targeted businesses are e-commerce sites, media service providers, travel and hospitality corporations, whose products end up on the dark web.

The purpose of this study is to examine and categorize this type of digital fraud, which has remained unexplored, and draw conclusions about the factors and characteristics which make a company an attractive target for fraudsters. Gaining even a little understanding of the issue could aid in the development of appropriate security management policies and mitigation strategies.

1.1 Background

To provide better understanding of the discussed topic, an overview of the main definitions and characteristics of the underground markets is made. Martin (2013) defines these structures as online forums where the exchange of goods and services takes place between parties who have concealed their identities through digital encryption. Underground markets, also known as dark net markets or crypto markets, operate through a network of websites hidden in the dark web (EMCDDA, 2017). To ensure the secure trade, they usually make use of: anonymization services, encrypted communication, authentication mechanisms, use of cryptocurrencies. Martin (2013) further specifies some characteristics typical for the cryptomarkets: they are usually accessed through the Tor network, providing the anonymity of both the users and the owners of the markets. In addition, to further anonymize users, cryptonyms are used. Payments are processed with cryptocurrencies such as Bitcoin, Ethereum, Litecoin and Monero, combined with the use of tumbling or mixing services, so as to ensure the untraceability of the transaction through the blockchain (EMCDDA, 2017). The transactions are further secured through the use of cryptography and authentication mechanisms.

Despite the relatively recent creation of the first crypto market: the "Silk Road", appearing in 2011, and the short life span most markets have had of less than a year on average (EMCDDA, 2017), the underground economy has been flourishing. It is acknowledged that there has been an increase in the volumes of goods and services flowing through the crypto markets, as well as in the range of products offered (Broséus et al., 2016). This could be credited to the developments in technology, making participation in the illegal trade more secure. The addition of various services like escrow payments, feedback ratings, and discussion forums has increased the users' feeling of anonymity and safety (Buxton and Bingham, 2015). It can be further pointed out that law enforcement operations seen as successful in closing down certain markets, have actually led to the emergence of even more dark net markets, utilizing more sophisticated technologies and spurring innovations in the process (Soska and Christin, 2015).

Despite the wealth of research on the topic, the majority focuses on drugs, which are the primary goods traded illicitly on the dark web. The main challenge in researching fraud facilitated by the underground markets is obtaining the data needed for analyzing the markets, as naturally there are no official records or numbers published from their owners.

One such dataset, containing detailed data from eight markets spanning several years, is explored in this thesis. As previously mentioned, a category of items present in the dataset indicates the existence of service fraud. This type of criminal activity affects numerous companies from various industry sectors: from retail stores, fast-food chains, supermarkets, to airline companies, and some of the biggest e-commerce sites worldwide. Considering some of the items offered on the cryptomarkets amount to thousands of dollars, such fraud is unlikely to be going unnoticed. However, no recent research has tried to identify what role service fraud has in the digital fraud landscape and what motivates cybercriminals to pursue particular targets.

When dealing with cybercrime, companies face both explicit and implicit costs (Anderson et al, 2009). While the more explicit and tangible financial costs are easier to estimate, the implicit are hard to quantify. They could be reflected in brand or reputational damage, reduced customer loyalty, or slower adoption of services, and often could have long-term effects, resulting in higher overall costs for the company (Lagazio, et al., 2014). Therefore, assets such as value, reputation, brand image, are often seen as central part of the functioning of a business. These may as well be the attributes which make a criminal prefer some companies over others, as the expected gains could seem greater. Hence, the technical aspects of a business should be considered along with the socio-economic and governance features in determining the attractiveness of a company.

Target selection is a relatively novel research area, mostly present in research investigating financial malware. Among the first are Tajalizadehkhoob et. al. (2013), who analyzed services targeted by the Zeus malware. Cheung (2017) has assessed country-level factors playing a role in distributed denial of service attacks in the financial sector. The studies of (Van Moorsel, 2016; Natalius, 2018; Hoppenreijs, 2019) have explored various characteristics, such as bank size or authentication type, which influence the target selection in different families of banking malware. The work of Sen & Borle (2015) has examined various factors influencing the risks of a data breach in an organizational context.

Overall, previous research on the topic of target selection has predominantly focused on the financial services industry. Therefore, given the lack of literature researching the particular topic of service fraud, the current study will use as reference some of these works in creating a conceptual framework.

1.2 Knowledge Gap and Research Problem

The purpose of the underground marketplaces to facilitate illicit trade and thus, stay hidden in the dark web, predetermines one of their main characteristics: untraceability. This inherent quality has made obtaining access or gathering data for research a rather difficult task. Most empirical studies in the area seems to focus on the functioning of the markets and the activities of the criminals, not so much on the effects these activities have on businesses. Moreover, a great amount of research is performed by private companies and organizations, which inevitably introduces some bias in the presented findings. The issue of service fraud and its impact has not previously been examined, especially the quantification of the costs. There is little knowledge of the factors playing a role in the selection process criminals go through before attacking a company.

From the outlined knowledge gaps, a problem which can be identified is that the issue of service fraud is yet relatively unknown and thus there is no study which has empirically analyzed the impact it has had on the affected parties. It is evident that cybercriminals engage in such fraud, though it is unclear what the consequences of these activities are.

1.3 Research Objective and Research Question

The research objective of the project is to explore the role of service fraud in the overall digital fraud landscape; categorize the different types of such fraud; provide a quantitative representation of its costs to affected companies; and give insights into what influences cybercriminals in the target selection process.

Based on the outlined knowledge gaps and research objective, this study will try to answer the following research question:

Which company characteristics influence the likelihood and the financial impact of service fraud conducted on the underground markets in the dark web?

To answer the main research question, the following sub-questions are to be investigated:

1. *What is the current underground digital fraud landscape and how does service fraud fit into it?*

   The answer to this question entails exploring and analyzing the current digital fraud landscape observed in the dark web, processes, and relevant theories for the following parts of the analysis.

2. *What company characteristics can be identified as potentially able to explain the differences in security incidents?*

   Possible factors used to characterize companies in the literature are examined and those deemed suitable are selected for further analysis.

3. *How can service fraud be classified and what are the main targets affected by it?*

   After exploring the available dataset, categories of service fraud are defined, along with a detailed analysis of main trends, targets, and costs incurred from such fraud.

4. *Which of the collected company characteristics are able to explain the differences in target frequency and incurred losses by service fraud?*

   Based on the previously identified company characteristics in Question 2 and the outcomes of Question 3, an explanatory analysis is performed in order to answer the main research question.

1.4 Research Methodology

Answers to the described research questions will be obtained through an empirical research approach. The purpose of such research is to draw conclusions and conceptualize problems based on empirical evidence (Brains et al., 2011). The following methods are going to be utilized in the three main parts of the research:

1. **Literature review and desk research**
   To provide answers to the first and second sub-questions, a literature review and a desk research are conducted. This would give a better idea of the current landscape of digital fraud, what are the estimated financial losses in the industries of interest for this research, and how service fraud fits into this landscape. The relevant data is obtained from searches in academic databases, online libraries and journals, as well as government and consultancy reports.
2. **Quantitative data exploration**
   The answer to the third sub-question is obtained by analyzing data on the selected types of digital fraud. The available database, consisting of various items sold on the underground markets, is manually explored, as the format of the data does not allow for an automated search. The refined results are analyzed and classified according to fraud type, while repeated targets are identified. Furthermore, the results from this analysis would serve to estimate the costs for the identified companies.
3. **Data Analysis**
   The final phase of the research consists of explanatory analysis aiming to provide answers to the fourth sub-question and the main research question. This is done through statistical techniques and more specifically: regression analysis, which is based on the output of the previous phase and additionally collected data for the purpose. Explanatory modeling was preferred to predictive, as the purpose of this analysis is to discover whether there are any underlying relationships between the collected variables and those extracted from the available dataset, while building up on the theory in the area of target selection. As such, the focus is on finding variables which may explain the differences in the number of incidents, thus indicating what possible remedying actions could be taken. Predicting future outcomes, given the hectic nature of the underground markets, was seen as a task lying outside the scope of the current research.

1.4.1 Research flow and structure

The research design is outlined in Figure 1.1. It is based on the different phases of the research process, while each phase is performed after the completion of the previous. The phase corresponds to the outcome of the research question.

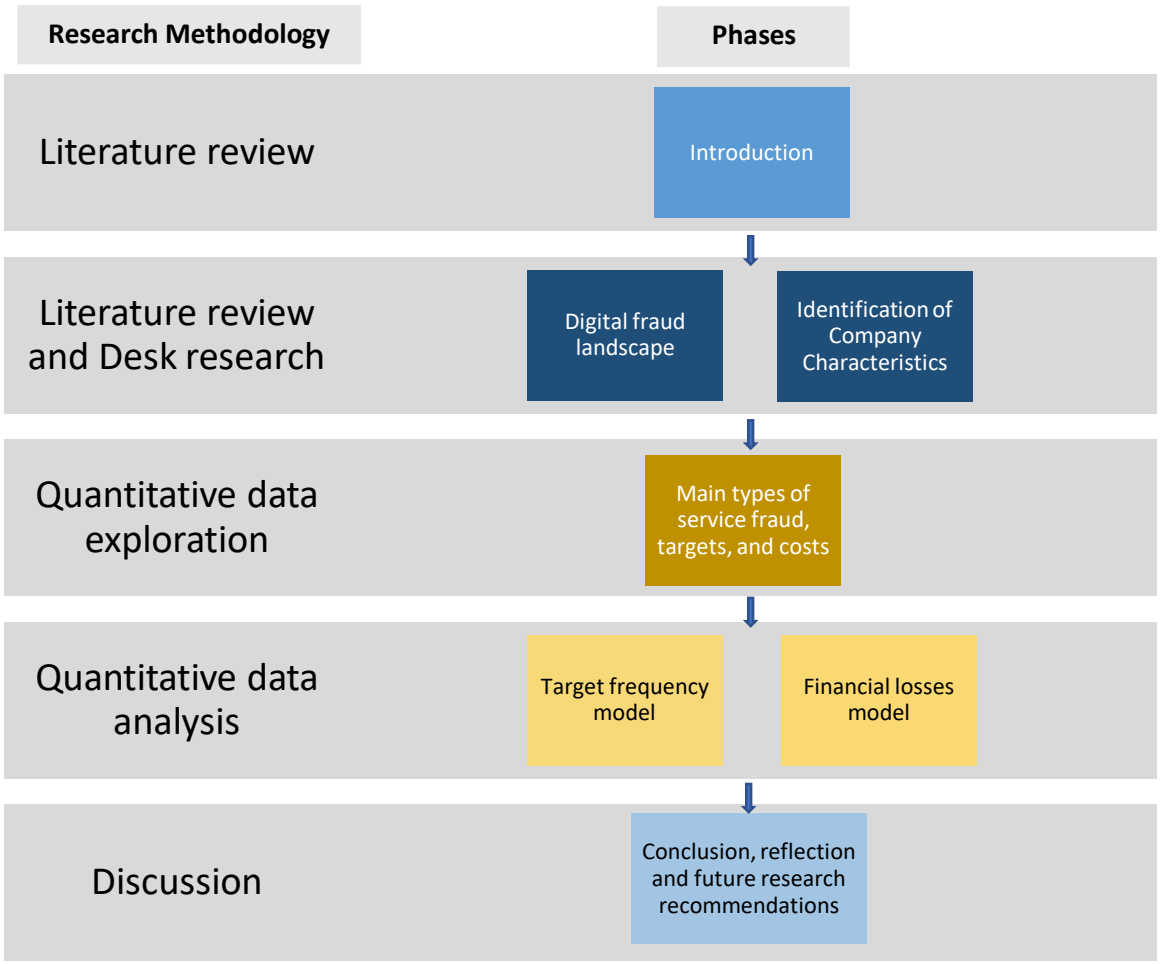


Figure 1.1 Research Flow

The outline of this thesis is the following:

- Chapter 2 provides an extensive review of the current literature starting with the basic concepts in cyberspace and cybersecurity, cyber security management, and economics of cybersecurity; the theoretical foundation of this study; the digital fraud landscape and the role of the underground markets in it; an overview of service fraud; and finishes with identification of various company characteristics influencing cyber fraud.

- Chapter 3 consists of an overview and preparation of the initial dataset.

- Chapter 4 presents the descriptive analysis of the provided data, the categorization of service fraud, and the estimated costs to companies.

- Chapter 5 details the collection and preparation of the additional datasets required for the regression analysis.

- Chapter 6 describes the results of the explanatory analysis: the performance of the regression models and analysis of the outcomes.

- Chapter 7 presents a conclusion providing answers to the research questions, details the limitations of the research, and gives suggestions for future research.

- Chapter 8 explains the scientific and social relevance of this thesis, and offers practical recommendations to affected parties.

2 Literature Review and Identification of Characteristics

This chapter introduces all concepts and theories related to this research. To understand the threat companies face, definitions of cyberspace and cybercrime are presented, along with an overview of the prominent cyber risk management frameworks, and the existing methods for estimating the cost of cybercrime. The basics of the Routine activity theory and Prospect theory are described as able to explain the rationale behind cybercriminals' target selection process. Then, the service fraud landscape is mapped out by defining the main characteristics and dynamics in the underground markets, the way criminals obtain credentials, and the types of service fraud found in the literature. Finally, the company characteristics seen as potentially influencing target selection are listed and a selection is made for the following analysis.

2.1 Cyberspace and Cybersecurity

2.1.1 Cyberspace and Cybercrime

The more information and communication technologies and their applications become prevalent in modern societies and intertwine with the everyday activities of people and organizations, the more the term 'cyberspace' is being used. Multiple definitions exist for what this notion constitutes, some putting more emphasis on the hardware and network infrastructure, while others include the information and data generated and stored in this environment (Rajnovic, 2012). Cyberspace can be seen as comprising all these, along with the users, services, and systems connected directly or indirectly to networks (ITU-T, 2008).

A conceptualized idea of cyberspace, encompassing the different aspects of the term, has been introduced by van den Berg et al., (2014). The model integrating the three components of cybersecurity, along with the various sub-domains which constitute the different sectors and industries, is illustrated in Figure 2.1. In the centre of the model is the technology layer, considered the first to have developed, as historically greater attention was paid to the technology and the infrastructure enabling cyberspace. The proliferation of various online applications and services lead to the formation of the second layer: the socio-technical, incorporating the relationships between users and their devices. On top of these two layers is the governance layer, where the numerous decision-making and policy enforcing actors are situated (van den Berg et al., 2014).

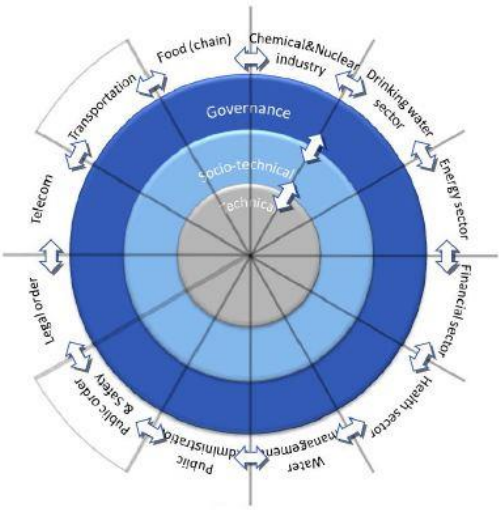


Figure 2.1 Conceptualization of cyberspace in layers and sub-domains. (van den Berg et al., 2014)

Along with the opportunities allowed by the advances in technology, also grew the threats enabled by them. Hence, cybercrime and cybersecurity are often viewed as integral parts of cyberspace. There is no uniformly accepted definition of cybercrime, as outlined by the European Union Agency for Cybersecurity (ENISA), which suggests the one from (Casey, 2004): "*Any offense where the modus operandi involves the use of a computer network in any way*", where the technology can be either the object or the tool of the crime (ENISA, 2017b). Furthermore, the European Commission has classified cybercrime in three general categories (European Commission, 2019):

- Internet-specific crimes, such as phishing and hacking;
- Online fraud and forgery, for example: identity theft, spam, malware;
- Distribution of illegal online content, especially such concerning terrorism, discrimination, child sexual abuse.

Arriving at a definitive term for cybercrime is additionally made harder by the constant developments within the cybercriminal world and its increasing interrelatedness with the physical realm. Cybercrime exists in the cyber environment, differentiating from traditional crime as being enabled by the electronic communication and information systems; however, its effects often transfer to the physical domain.

2.1.2 Brief Overview of Cyber Security Management Practices

Cybersecurity can be seen as an extension of the previously known disciplines of information and network security (van den Berg et al., 2014), as it deals with all aspects of the cyber environment and the risks within it. While information has long been viewed as a key asset in cyberspace, with the main focus put on preserving its Confidentiality, Integrity, and Availability (CIA), in recent years more attributes have been defined as part of cybersecurity, such as ensuring the Reliability, Safety, Resilience, Authenticity of its components (van den Berg et al., 2014). Overall, cybersecurity should encompass the prevention, forecasting, detection, mitigation, removal, analysis and investigation of cyber incidents (ENISA, 2017a).

Given the continuously evolving cyber threat landscape, security is increasingly becoming more expensive for organizations, and they have to carefully manage their security investments. Risk quantification theory (Zhao, 2007) is a framework commonly applied to the identification of threats, assets and vulnerabilities, and how to quantify investment (Bojanc & Jerman, 2008). Nevertheless, it is reported that organizations rarely perform a more sophisticated financial analyses to determine their security investments, but take a rather reactive approach to security issues (Rowe & Gallaher, 2006).

There are various frameworks devised to manage cyber risks, such as the FAIR (Factor Analysis of Information Risk) methodology (Jones, 2005), and the TARA (Threat Agent Risk Assessment) (Rosenquist, 2009). FAIR outlines what constitutes information risk and the main factors influencing it, and provides a detailed method to analyze the various threats and possible risk scenarios. The TARA methodology, developed by Intel, focuses on the attacker model in risk analysis and explores the potential threat agents, while considering which areas are most vulnerable to attacks. The National Institute of Standards and Technology (NIST) has also proposed a detailed Cybersecurity Framework guiding organizations as to how to incorporate cyber security practices into their business processes, while considering the cyber, physical and human aspects of cybersecurity (NIST, 2018). A prominent point in cyber risk management is grading the cyber dependent activities based on their criticality for the organization: the starting point is recognizing and evaluating the organization's assets and vulnerabilities; identifying the

potential cyber risks, their likelihoods and impacts; determining acceptable levels of these risks; and devising a list of controls and response measures to mitigate them.

In addition to quantifying cyber risk, another major topic within cyber security management is deciding how much and how to invest. Various security investment models have been developed for the purpose. Among the most prominent, despite being criticized for being overly abstract, is the one introduced by Gordon and Loeb (Gordon & Loeb, 2002). It suggests deciding on a level of security investment relative to the expected loss without security investment, by using a security breach probability function. It takes into account the particular value of the security investment, and the probability of suffering loss. What the model has clearly demonstrated is that security investments have decreasing marginal returns, thereby there is an upper bound on investments after which the costs of security outweigh the benefits, as illustrated in Figure 2.2 (Anderson, 2001).

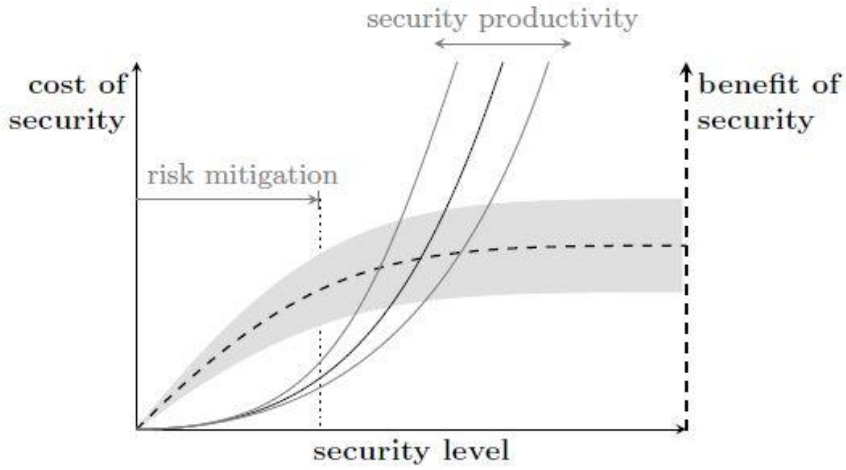


Figure 2.2 Security investment function as proposed by Anderson, 2001

An alternative method devised to aid businesses with managing cybersecurity is the Return On Security Investment (ROSI) model (Sonnenreich at al., 2006). It draws comparisons with the widely used in investment finance concepts of Return On Investment (ROI) and Net Present Value (NPV), thus facilitating its use in corporate settings. ROSI takes as parameters the solution cost, as well as the risk exposure and the risk mitigated. However, the difficulty in employing this method lies in the lack of accuracy in determining the correct values in order to calculate ROSI, as it relies on assumptions similarly to ROI (Bohme, 2010).

All these frameworks acknowledge that dealing with information security incidents poses many practical issues, and no methodology can be extensive enough to cover all possible risk scenarios. This is in part related to the fact that an information security incident concerns all aspects of cyberspace and the various parties responsible for each: while it may happen on the technical level, its impacts are reflected on the socio-technical, and consequently the governance layers.

Given the high levels of globalization and connectivity nowadays, no organization stands alone in the face of cyber risks, as one incident can easily be the cause of others and threaten the whole environment. Proper cyber security management could not only mitigate the adverse effects of potential cyber risks, but also assist in maintaining a positive image of the organization.

2.1.3 Cost of Cybercrime

With the rising importance of the social and human factors, there was a shift from the sole focus put on the technical aspects of cyber security to a more holistic view, incorporating an economical and more business-oriented approach (van den Berg et al., 2014). Cyber risks are now seen as a danger to not only networks and infrastructure, but involving the entire organization, along with its customers and users, and the ecosystem on the whole. Especially, considering that often the reason a system may be breached is not because of the underlying technology but the different incentives of the involved parties (Asghari et al., 2016).

Numerous reports by various actors, from governmental organizations to security vendors, have tried to estimate the costs of cybercrime. However, oftentimes the presented data is either over- or under-estimated, either intentionally or due to the measurement techniques employed, as explored by Anderson et al. (2012). The authors have introduced a framework for classifying the costs of cybercrime into four categories:

- Criminal revenue: the profits realized by the criminals by conducting the illegal activities.
- Direct costs: the actual financial losses and damage experienced by the victim. They may include lost money, time, effort, psychological suffering.
- Indirect costs: losses and opportunity costs caused by the sole act of committing the cybercrime, regardless of whether it was successful or not. These costs can be substantial as they often are related with the loss of reputation or customer trust, which could have long-term negative consequences for the organization.
- Defense costs: the financial costs invested in preventive measures. They can be direct: for the development, deployment, and maintenance of security products and services; or indirect: the opportunity costs and business inconveniences inflicted by the employed procedures.

The sum of the latter three categories: the direct, indirect and defense costs, constitutes the Cost to society, which has been shown to considerably exceed the revenue made by criminals. A reason for this is that cybercrimes are usually undertaken globally, thus spanning different jurisdictions, and affecting a wide range of actors.

Moreover, cybersecurity is described as exhibiting strong externalities: these are costs or benefits incurred by parties who did not take part in the decision-making process (Anderson, 2001; Asghari et al., 2016). As a result, while some security measures may seem reasonable to implement by companies, the indirect costs they may bring about could render them meaningless, or even detrimental to business interests. Thereby, organizations may have an incentive to forgo implementing certain security measures, provided they are willing to accept the risks and losses this would entail.

2.2 Theoretical Foundation

Similarly to other types of fraud, service fraud can be considered a crime committed in the cyber domain, making the application of criminology theories suitable for this case. Specifically, the Rational choice theory (RCT) and the Routine activity theory (RAT) are going to be examined. Furthermore, the Prospect theory used in behavioral economics could provide additional insights.

Among the central theories in economics is the rational choice theory (RCT), often used to explain economic and social phenomena, as well as individual behaviour. In addition to microeconomics, it has been widely applied in other fields such as political science, philosophy, and criminology.

RCT is based on the classical utilitarian philosophical paradigm, in which the utility of performing certain actions governs human behaviour (Clarke, 1997). It is assumed that men are reasonable beings, consciously weighing the costs and the benefits of their actions and making rational decisions, aiming to maximize their goals. RCT takes the point of view of the individual as central, regarding individuals as accountable for their choices and behaviour, and then translates it to the level of the larger social group or situation.

Some of the main assumptions made in RCT include (Abell, 1992):

- It is individuals who effectively take action, which eventually manifests in social acts;
- People always act as rational beings taking into consideration the changing conditions;
- Individuals make decisions of their own accord, disregarding external influences;
- People take an optimal course of action given the circumstances, preferring the best possible option.

Despite being a prominent theory in various fields, RCT is still widely criticized, mostly for the inability to empirically prove its assumptions, and its heavy reliance on rationality (Green & Shapiro, 1996; Sato, 2013). It is highly doubtful whether individuals engage in thoughtful calculations each time before they take an action, or that they are always guided by pure rationality.

One theory which implies that other psychological processes have an effect on actors' actions is Prospect theory, developed by Kahneman and Tversky (1979). It explores the decision-making process people go through when dealing with risk and uncertainty, concluding that in situations entailing certain losses people tend to exhibit risk-seeking attitudes, while when a certain gain is at stake, they are rather risk-averse. Furthermore, it stipulates that framing, or the way information and different alternatives are presented to individuals, can greatly influence the outcomes of people's decisions.

More recently, the concepts of heuristics and biases were added to the theory, highlighting the role of the mental shortcuts people take when making decisions, and the influence of prior knowledge and experience, as well as of situational and environmental factors (Kahneman, 2011). These have all been shown to affect perceived rationality, especially when making decisions concerning risk and security. When individuals are faced with situations where there is uncomplete information, they might resort to relying on known norms or habits, or make assumptions based on the provided limited facts.

Among the most prominent theories in criminology is the Routine activity theory, introduced by Cohen & Felson(1979). The central premise of RAT is that there are three necessary conditions for crime: a motivated offender, an attractive target, and the absence of guardians, all converging in time and space. The suitability of a target can be measured by four main properties: *value, inertia, visibility,* and *accessibility*. Value is linked to the potential gain the criminal can obtain from attacking the target. Inertia is related to physical features of the target which could cause a hinderance to the offender when committing the crime. Visibility reflects how aware the criminals are of the target, and accessibility – the ease with which the target can be reached.

The applicability of the mentioned theories to the explored topic is going to be discussed in the following section.

2.2.1 The Theory in Relation to Service Fraud

The main principles of RCT can be employed to explain the behaviour of both the fraudsters and the targets on the underground marketplaces. From the criminal's point of view, engaging in cyber fraud is likely to be preceded by a careful weighing of the benefits, including factors such as financial gains, entertainment, retribution; and the consequences: the possibility of getting caught and the potential incurred penalty. As such, for offenders the benefits of committing fraud would often outweigh the costs, especially in cyber space, where the likelihood of being apprehended is minimal, and the costs of participation marginal.

Rational choice theory as presented by Becker (1974) could also be applied to the actions of those considered targets of service fraud. Becker admits that it would be impossible to completely eradicate crime, as the benefits from this would not justify the great costs which would be required to do so. Considering companies are constantly seeking ways to maximize their revenues while viewing their public image and reputation as one of their main assets (Lagazio, Sherif, & Cushman, 2014), it could explain why less attention is paid to service fraud, or why it is not publicly acknowledged. It is unclear how much of a threat non-financial account theft is considered by companies, and whether they prefer to invest to protect other assets which could be seen as more vulnerable.

A different aspect to the victim's behaviour can be provided by prospect theory. Despite that companies may have complex decision-making procedures in place with regard to security, the outcomes are still likely influenced by some of the psychological phenomena the theory presents. It is possible that the preoccupation of the media or security sector with certain types of cyber threats has pushed service fraud away from consideration as less significant. Moreover, prospect theory can be applied to the decision-making process of criminals, as they may be subject to the drawbacks of perceived rationality.

The rationale behind target selection can possibly be best explained through RAT. How relative the theory is to cybercrime has been examined by Yar (2015), who argues that it is still applicable, albeit with some modifications. The properties which define a target's suitability have been revised in relation to the online world: the usability of inertia in evaluating a target can be doubted, as the virtual environment does not pose the same constraints as the physical world. Moreover, Leukfeldt & Yar (2016) have examined the RAT elements in relation to different cybercrimes. The value aspect of a target has been shown to have little influence in selecting targets in consumer fraud, unlike visibility and accessibility. However, the study considers individuals as the victims, and not organizations, which is the focus of this thesis. Therefore, in order to identify characteristics influencing the target selection in service fraud, the three suitability elements suggested by RAT are going to be applied: value, visibility, and accessibility.

2.3 Digital Fraud Landscape and Role of the Underground Anonymous Markets

Considering data is increasingly being seen as a major asset from companies and individuals, it is of no surprise that information theft, loss, or attack was reported as the fraud most often experienced in recent years (Kroll, 2017).

Sensitive data is gathered and stored for various purposes by companies, governments, and other entities. While usually protected by a range of defensive measures, databases can often get attacked and compromised by hackers and cyber criminals, on occasions leading to the massive

exposure of information (Franklin et al., 2007; Holt, 2012; Holt et al., 2016). In previous years, mass data breaches have occurred, compromising the personal information of millions of customers (Holt et al., 2016).

Naturally, after acquiring such vast amounts of information, hackers would need a place to advertise and distribute it. This has led to the emergence of various online markets specialising in the trade of illegally obtained data and numerous other types of criminal goods (Franklin et al., 2007; Holt et al., 2016; Motoyama et al., 2011). These markets have different structures, some being operated as Internet Relay Chats (IRC), while others bear more resemblance to forums and websites (Malin, Gudaitis, Holt, & Kilger, 2017). The main language used may also differ based on the region they are operated from. Some are available on the open web, and as such are accessible to everyone (Thomas et al., 2013), while others are hidden in the dark web, protected by various encryption techniques (Martin, 2013), thus isolating their users from the general public.

Despite the fact that malicious software and compromised sensitive data can be found and obtained on the open web, it is believed the publicly accessible sources cannot offer the same level of organization and commodization as the underground online marketplaces (K. Thomas et al., 2015; van Wegberg et al., 2018). Moreover, since the greater part of the wares offered on the dark web are drugs, as well as other illegally attained products, it is imperative for criminals to try and conceal their identity to the best of their abilities. Hence, while some of the trade may now occur on the surface web, the majority is likely to remain on the encrypted networks in the dark web (Europol, 2018).

2.3.1 Evolution of the Underground Digital Trade

The digital underground economy is constantly evolving and changing, thus making it difficult to get a comprehensive view of the way it is organized. What started as forums used primarily for the purpose of sharing experience and techniques by hackers and individuals interested in honing their skills, has become a complex network mostly driven by profit (Thomas et al., 2015).

Online illegal trade has existed in one form or another for several decades. Hackers and other interested parties are reported to have been communicating and sharing files through message and bulletin boards as early as the 1980s (Malin et al., 2017). Such endeavors were made easier by the emergence and later the global spread of the Internet, which in turn led to the creation of better organized networks and criminal groups. Thomas et al. (2006) describe the initial actual digital underground markets as platforms utilizing IRC: a standard protocol for real-time text messaging over the Internet (Oikarinen & Reed, 1993). Users on the illicit markets would utilize IRC to share availability and pricing information about various products and services, such as credit card information, compromised accounts, botnets, malware (Thomas & Martin, 2006).

Gradually, this type of structure was replaced by web forums, where more varied information was shared under unique threads, and access was often more restricted than in the open IRC chats (Motoyama et al., 2011). The existence of such forums in various geographical regions has been explored previously: Holt et al. (2016) present an analysis of markets operating in Russian and English, Broseus at al. 2016 explore dark net markets in Canada, while Zhuge et al. (2008) focus on forums in China.

Eventually, the web forums evolved into a market structure, similar to the e-commerce platforms popular on the open web. The predecessor of the numerous underground markets which have existed in the last decade and the first to boast an organization of this kind was the Silk Road market, launched in 2011 (Christin, 2013). It is reported that by the time it was shut down in law

enforcement operations in the United States and Australia in 2013, a revenue of around $1.2 billion had been made through the site (Barratt, 2012). Its closure, however, did not lead to a decline in the underground trade online, as several other markets of a similar structure emerged, drawing in the users of the Silk Road (Malin et al., 2017).

2.3.2 Current Underground Landscape

Despite the relatively recent creation and following closure of the Silk Road, the underground economy has been flourishing. Buxton and Bingham (2015), Soska and Christin (2015), Broseus et al. (2016) argue that there has been an increase in the volumes of goods and services flowing through the crypto markets as well as in the range of products offered. This could be credited to the developments in technology, making participation in the illegal trade more secure.

Soska and Christin (2015) have shown that since the existence and disappearance of the first online anonymous market, the number of sellers has significantly increased along with the high competitiveness among suppliers. However, most of the examined markets seem to reach vendor saturation, or never expand sufficiently, due to law enforcement operations shutting them down; self-destructing mechanisms such as exit scams performed by the markets' owners; or voluntary closures (Europol, 2018).

Despite the limited growth of the markets and their short life span, the activity on them remains high as shown by Broseus et al. (2016), who examine the structure and the functioning of eight dark net markets over two months in Canada. Their findings are consistent with the previously reported (Soska and Christin, 2015) diversification and replication of vendors on different marketplaces, aiming to increase profits and reputation or mitigate risks of potential shut-downs. The study serves to confirm what other authors have observed, despite its regional character owing to the fact that the analysis is based only on data pertaining to markets operating from Canada.

Although law enforcement operations are seen as successful in closing down certain markets, they have actually led to the emergence of even more dark net markets, utilizing more sophisticated technologies and spurring innovations in the process. As witnessed in 2017, when three of the most significant global dark web markets: Alphabay, Hansa, and RAMP were shut down in law enforcement operations, these actions only led to the trade being shifted to other existing markets or to newly-found smaller, privately run vendor shops, along with more regional secondary markets operating in particular countries or languages (Europol, 2018). It is expected that the smaller scale and more targeted approach of these marketplaces is possibly not going to attract the same level of attention from the authorities or the media, as the larger platforms.

2.3.3 A Market for Lemons

The inherent characteristics of the underground markets as being outside the law, and the measures employed to hide the identities of their participants, while achieving the goal of anonymizing users to law enforcement agencies, have the negative effect of rendering the buyers at a disadvantage. Clients, in the case of being cheated or provided with unsatisfactory service, usually have no formal mechanisms to protect their rights as buyers (Holt, 2012). Since they are engaging in activities deemed illegal, there are no outside regulators, such as law enforcement or governmental organizations, which could aid them, hence they have to rely solely on the discretion of the vendors.

This seems to suggest that the sellers may have the advantage in the trade on the dark web, as they likely have much more information available to them about the quality of the offered product than the buyers (Herley & Florencio, 2010). It is suggested that the crypto markets suffer from strong information asymmetry effects, which along with the lack of formal regulatory mechanisms, makes them no different than a market for lemons as defined by Akerlof (1978). The implications of this would be that primarily those vendors who try to pass off low-quality or fake products would remain on the market, driving quality sellers away, and eventually leading to a market failure, until the information asymmetry is corrected.

However, the discussed paper by Florencio and Herley (2010) focuses on Internet Relay Chat markets, which have chiefly been replaced by other forms of exchange as mentioned in the previous chapter. It is unclear whether these evolved market structures, which usually include feedback and reputation systems, in order to improve both seller and buyer credibility, would display similar effects to those presented by Florencio and Herley (2010), as the underground economy has shown to be particularly resilient.

Such an exploration is made by Allodi et al. (2016) as the authors draw a comparison between the IRC markets and the forum-based marketplaces, by observing and examining two online anonymous markets: a failed market based in Germany, and another successful one – in Russia. The results of their study confirm in practice that a market lacking reputation or other regulatory systems may eventually fail due to the information asymmetries and high level of scammers on such a forum. Furthermore, they acknowledge that more evolved markets of the type mostly present nowadays, are not as easily affected by such effects, as they boast better structure and regulation. The study by Allodi et al. (2016) makes valid scientific contributions to the theoretical field of the underground economy. However, being based on the analysis of only two markets operating in two distinct countries may limit the applicability of the findings. Geographical specifics could influence the functioning of the two examined markets, such as less governmental prosecution in certain regions, or historically established traditions in cybercrime.

2.3.4 Factors Influencing the Dynamics in the Underground Markets

The factors described by Holt (2012) and Malin et al. (2017) as having most influence on the crypto markets and their functioning are: price, customer service, trust, and reputation.

Price

Understandably, the price of products has an influence on the flow of the trade (Holt, 2012).Considering that the pricing information is publicly visible for all participants, customers can compare and select offers which would appear to provide the best value of return, thus promoting competitive pricing (Franklin et al., 2007; Holt, 2012; Motoyama et al., 2011). Similarly, openly advertising the price of a product or service, may serve as a regulatory mechanism for vendors, since those demanding prices deemed to be either too high or low, would find it difficult to attract customers willing to pay unfavourable fees for their wares. Herley and Florencio (2010) argue that the advertised price could be an indication of the validity of the data or service provided, and the legitimacy of the sellers. Especially, when the sale of credentials or accounts is concerned as they become stale over time.

Customer service

Customer service was identified as the second factor determining the dynamics in the underground online markets (Holt, 2012). Satisfactory communication between buyers and sellers, as well as a positive user experience with other participants, are observed to have an impact on customer engagement. Moreover, the quality and usability of the delivered goods and services play a key role in customer satisfaction and retention (Malin et al., 2017). An additional factor is the promptness of sellers in contacting clients, and in delivering the products ordered (Holt, 2012).

Sellers employ diverse customer service techniques to attract and retain clients: posting samples of the offered goods and real time customer support messaging are common examples (Holt et al., 2016). Providing bulk discounts on products sold in large quantities, such as compromised data and accounts, is another extensively used mechanism, similarly to offering lifetime warranty and account exchanges in the occasion the provided credentials are no longer valid (Wilson, 2019).

Trust

The third major factor distinguished in the literature is trust, which has been found to strongly relate to both price and customer service (Malin et al., 2017). As buyers may not be able to assess the quality of the offered goods, nor legally protect themselves in the case of being cheated, they have to be cautious when selecting a vendor, in order to limit any potential losses. Holt et al. (2012) have defined some of the existing informal mechanisms introduced in the crypto markets to guarantee a certain security level for the participants.

- Providing formal validation of the items offered on the market by the forum administrators, usually done by checking a sample of the data, or testing the service.
- Use of escrow services. This introduces a trusted third party acting as a guarantor for the transaction. The payment made by the client is not transferred to the seller directly but remains in the escrow agent, until the product or service is delivered and the buyer has confirmed the successful transaction (Buxton & Bingham, 2015).
- Customer feedback (Christin, 2013; Holt et al., 2016; Malin et al., 2017). On the majority of the dark web markets, customers are expected to provide user feedback and post reviews after a transaction has been completed. This process aids in establishing reputations and distinguishing quality sellers from frauds.

Reputation

Although preserving anonymity is one of the main priorities on the dark web markets, reputation is still considered an influencing factor, mainly achieved by managing usernames and received feedback (Aldridge & Decary-Hetu, 2014; Esparza, 2019). Vendors have been shown to maintain certain identities over different markets, in order to establish a name for themselves and increase their presence and consequently, their market share.

A table summarizing the factors and the methods through which they are influenced is presented below.

Table 2.1 Main factors influencing the trade dynamics on the underground markets.

Factor	Means
Price	Openly advertised prices (Holt, 2012)
Customer service	Buyer-seller communication, Quality and usability of data (Malin et al., 2017) Promptness in delivery, Customer support line (Holt, 2012) Lifetime warranty for accounts (Wilson, 2019)
Trust	Admin validation of the offered data (Malin et al., 2017) Escrow services, Customer feedbacks (Buxton & Bingham, 2015; Christin, 2013; Holt, 2012)
Reputation	Username management (Aldridge & Decary-Hetu, 2014; Wilson, 2019)

In addition to providing a suitable environment for hackers, cyber criminals and fraudsters to exchange information, tools and experience, the illicit markets have further facilitated criminal behavior by lowering the threshold for those willing to participate in illegal activities (Mirian et al., 2019; van Wegberg et al., 2018; Wilson, 2019). Various tools and services, the development of which would previously have required certain expertise and sufficient skills, are readily available for sell or rent to even the most inexperienced. This, in turn, has contributed to the commercialization of fraud and commodization of crime.

2.4 Service Fraud

2.4.1 Impacts of Credential Theft

Since their establishment, the dark web markets have always been associated with the online distribution of drugs, and it still constitutes the bigger part of the trade on most (Buxton & Bingham, 2015; Europol, 2018). Nonetheless, a significant increase has been noted in recent years in the retail of tools and services aiding crime, and even more so for compromised data (Figure 2.3.). Among the commoditized products sold on the dark web, data is regularly highlighted as the second or third largest commodity (Europol, 2018), with data breaches in 2018 claimed to be costing globally around $3.86 million on average (IBM Security, 2018).

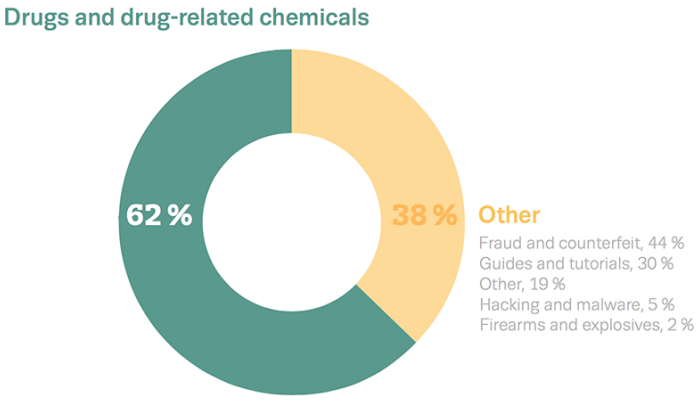


Figure 2.3 Dark web markets content. (Adapted from EMCDDA, 2017)

A data breach can result in substantial financial losses from: mitigation and recovery procedures, legal costs, regulatory enforcement (Wilson, 2019). However, the indirect effects of failing to protect user data can be extensive: interruption of business continuity, loss of reputation, employee trust and customer loyalty can prove costly in the long run (IBM Security, 2018). This is made evident in Figure 2.4., which illustrates the direct and indirect costs of data breaches in various countries for 2018.

The monetary losses of credential theft for companies and organizations are substantial, but the negative consequences of such an attack go beyond the merely economic. It has been reported by the majority of companies who had suffered a fraud or a cyber incident in recent years, that the strongest negative impact was in fact felt by the employees, as their privacy and morale were negatively affected, and their feeling of safety lessened (Kroll, 2017). More than half of respondents also stated the adverse impact fraud has on their customers, reputation, revenues and relationships with regulators (Kroll, 2017). Given the relatively recent introduction of the GDPR (General Data Protection Regulation) in 2018 in Europe, repercussions of data mishandling are likely to have even stronger adverse effects on companies and provide criminals with additional motivations to target data held by companies, as they could realize higher gains from extortion.

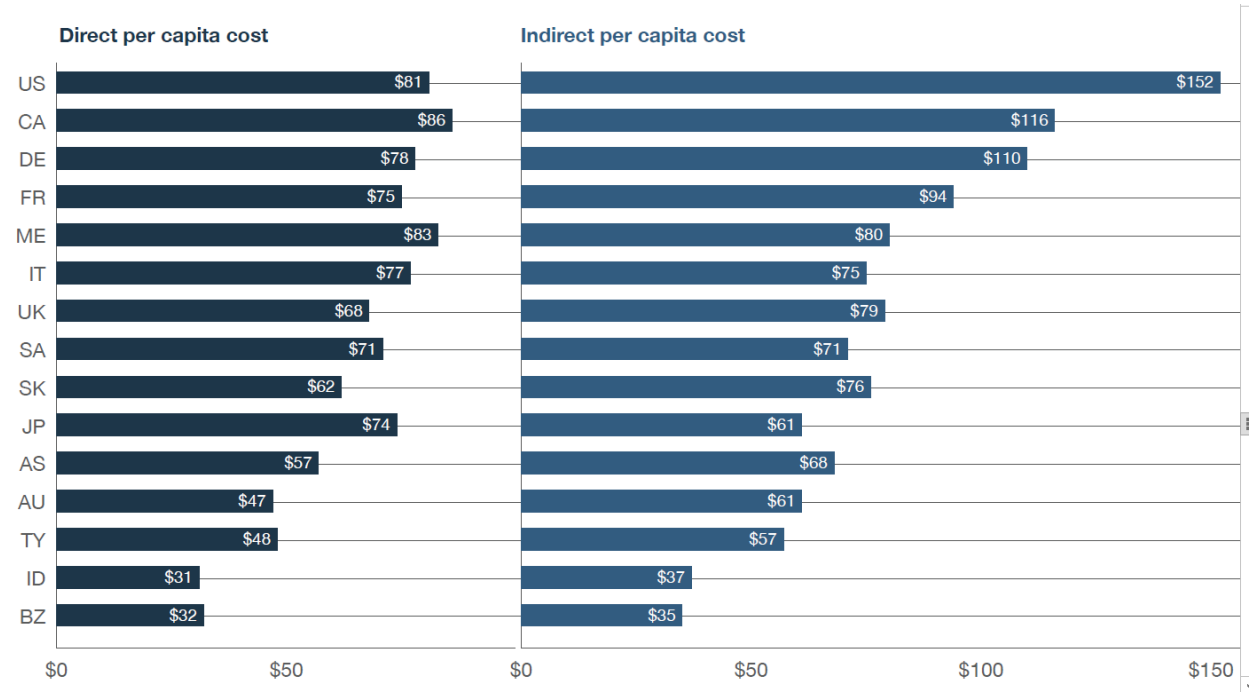


Figure 2.4 Direct and indirect costs of data breaches (IBM Security, 2018).

2.4.2 Attack Vectors

Thomas et al. (2017) outline the three prevalent techniques criminals employ to gain unrightfully access to credentials: malware, phishing, and through credential breaches. The most significant reported amount of credentials is exposed in credential breaches: 1.9 billion; 12.4 million usernames and passwords are obtained through phishing; followed by 788,000 users affected by malware. Yet, it is reported that phishing is most likely to lead to successfully hijacking an account, followed by malware, and lastly, data breaches. Moreover, phishing and malware credential attacks are more easily customized and tailored to particular targets and security mechanisms, potentially causing more harm (Mirian et al., 2019).

The focus of the study is on the email address as the basis of the online identity, since once attained, it can be used to acquire access to other online services, compromise user data, or serve as an initial point for further attacks. A limitation of their work is that it is based on a case study on Google accounts and as such may not apply widely to other services, utilizing different security mechanisms.

Other widespread techniques as reported by the security industry are: man-in-the-middle attacks, DNS hijacking; exploiting known vulnerabilities; brute force attacks; social engineering (Blueliv, 2018).

Credential Spill

Credential leaks are becoming commonplace nowadays, with data breaches affecting millions of users being reported on a regular basis (Accenture, 2017; Symantec, 2018). Attackers are going after major online services such as Yahoo, Gmail, Adobe, LinkedIn, Dropbox, government agencies and organisations, as well as non-commercial blogs and forums. Furthermore, the damage is usually not contained only to the services initially targeted, as users are consistently found to be re-using credentials over platforms or utilizing weak passwords, thus exposing third party sites

to subsequent attacks (K. Thomas et al., 2017). Cybercriminals are often going through email accounts searching for financial data, links to other services, or abusing them for spam attacks: all methods to monetize the credentials.

Malware

Malware infections are an efficient method widely applied by criminals in credential theft attacks, as a significant amount of data is permanently stored on computers in password vaults, configuration files, third-party applications (Blueliv, 2018). A popular type of malware used in credential theft are off-the-shelf keyloggers. Originally keyloggers' main functionality was to record and transmit the sequence of keystrokes on a machine, however, later editions, such as HawkEye, Predator Pain and Cyborg Logger, can also monitor user activity, steal passwords stored on the device or clipboard content (K. Thomas et al., 2017).

Alternative methods for data gathering include the use of botnets, such as the Torpig botnet, effectively employed for harvesting email accounts and passwords; as well as banking trojans, which have evolved and are increasingly being used for stealing non-financial credentials.

Phishing

Among the oldest but still effective methods of gaining unauthorized access to data is by targeting the users, most often through phishing campaigns (Kroll, 2017; Symantec, 2018).Cybercriminals usually employ phishing kits: complete packages used to generate and configure phishing attacks, including support for reporting stolen credentials. The data obtained from victims by phishing kits can vary, though most often includes usernames, passwords, geolocations, phone numbers. Typically, this method provides criminals with more up-to-date credentials than exploiting a database (Blueliv, 2018; Shape Security, 2018).

Thomas et al. (2017) observed that on average there is little development in the functionality and capabilities of keyloggers and phishing kits in recent years, suggesting users are being exposed to similar threats, while service providers have not found a satisfactory solution.

Other Techniques

- Man-in-the-middle attacks. This type of attack relies on monitoring the connection between the user and the service they are trying to access. Attackers can sniff the traffic and extract the credentials information from the network, or redirect users to a phishing site, emulating the legitimate service.
- DNS hijacking. This attack is made possible by the common use of third-party DNS management services used to add, alter, or erase DNS records. Cybercriminals attack the DNS service provider, and through their website target their clients, thus obtaining more credentials.
- Exploiting well-known vulnerabilities or misconfigured systems. Oftentimes, criminals exploit SQL vulnerabilities by undertaking SQL-injection attacks: these allow attackers to send unauthorized commands directly to a database, due to inadequately managed user-submitted data. Furthermore, it is a widespread issue among numerous organizations to misconfigure their databases or servers, or forget to change the default settings, thus rendering them vulnerable.
- Brute-force attacks. Attackers manage to obtain valid credentials through methodically trying possible character combinations on websites until they guess the correct one. The success of these attacks is aided by the common lack of caution people practice by selecting weak or dictionary word passwords.

2.4.3 The Criminal Business Model of Account Theft

Actors in the cybercriminal world are long past being driven by mere curiosity of hacking systems, or simply causing a disruption, as the desire for profit has come firmly along these motives, or in some cases, completely replaced them. This could be attributed in part to the shift of commercial and financial operations towards the Internet, as well as the advances in technology making large-scale attacks more of a trivial matter (Thomas et al., 2015). The underground economy nowadays boasts well-established value chains aimed at streamlining the profit-making abuse. Credential theft is highly dependent on the underground markets, and the goods and information traded there. As various tools and services enabling crime are widely available for sale, people willing to engage in criminal activities no longer have to be able to master the entire process themselves, they can simply become part of the chain. Therefore, fraudsters possessing only certain skills may put these to use, and at the same time take advantage of the expertise and experience of others.

The various criminal business models identified in the dark web have previously been classified by Thomas et al. (2015), and later updated in van Wegberg et al. (2018)'s work. The model suggested by them for the resale of accounts is used in this work and further expanded.

A model of the value chain underpinning service fraud is presented in Figure 2.5. One of the most often employed techniques for account and voucher theft is through credential stuffing attacks, consisting of re-using compromised credentials to access other targeted websites. The following process is mainly based on this type of attack.

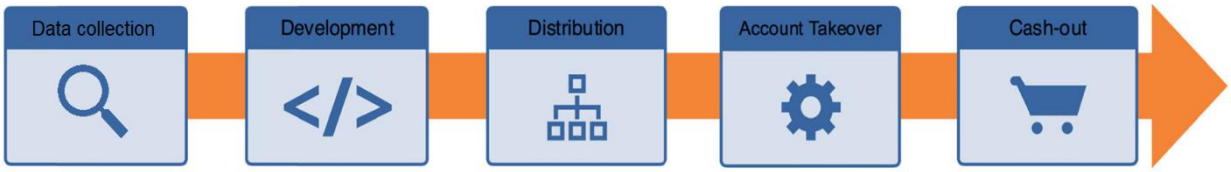


Figure 2.5 Service Fraud Value Chain. (Adapted from van Wegberg et al., 2018)

There are essentially five stages in the credential theft schemes: Data collection; Development of an attack tool; Distribution; Account Takeover; and Cash-out. These are going to be further elaborated on below.

Stages of Service Fraud

Data collection is the first step in executing credential and subsequently, account theft. There are numerous possibilities open to cyber criminals based on their abilities and resources, from performing an attack singlehandedly, to purchasing data from either a contact from their criminal network, or making use of the tools available on the dark web markets. Rates of credentials on the market can vary significantly, with the main price determinants being origin, freshness, attacker reputation, and credentials validity (Shape Security, 2018). In addition to purchasing credentials, some of the more popular data gathering techniques are: phishing, social engineering, malware distribution, leaked databases, exploiting existing vulnerabilities (Blueliv, 2018).

This stage in the process may also involve filtering, extracting and authenticating of the stolen credentials (Blueliv, 2018). Since usually the credentials are stored in databases or files, they need to be extracted and checked for validity, as the obtained accounts and passwords may not be useable any more. This task is completed by employing bots, credential checkers, or even manual checks for smaller amounts of credentials.

Development. The second stage in the chain can progress in a few different directions depending on the criminal. Some, especially the more proficient, may prefer to develop their own attack script or tool, which despite being a rather laborious task, would yield more reliable results in the long term. Other attackers who do not have the abilities or the desire to delve into writing their own software, can acquire a readily-made toolkit from the dark web (Shape Security, 2018).

Attackers who have developed their own software would usually have to go through an additional testing phase. Prior to distributing the attack, cybercriminals would first test the tool on the desired website, so as to ascertain whether the attack software is able to successfully circumvent the site's defenses. This is usually done by utilizing fake accounts, specifically created for the purpose. In the case that the attack script does not function as intended, based on their expertise, attackers may be able to assess and determine why the script is failing, and eventually fix it (Shape Security, 2018).

Distribution. Once the criminals have verified the efficiency of the attack tool, they can move on to distributing the attack usually by utilizing numerous IP addresses through proxy servers. Acquiring access to proxy servers is relatively cheap and simple to do, with various available options at different prices (K. Thomas et al., 2015). On the clear web there are open servers which are free to access, as well as paid service providers. However, they are usually overused and thereby, already known and blocked by organizations. The other option is to buy access to proxy servers owned by other cybercriminals, advertised on the underground markets (Shape Security, 2018). Attacks may often be aimed at specific targets or distributed in particular regions (K. Thomas et al., 2017).

Another alternative criminals have to developing or purchasing an attack tool, and distributing an attack themselves is to use an account checker service (Shape Security, 2018). This is usually leveraged by the least experienced of attackers, who still wish to take advantage of the numerous data spills, while lacking the required abilities. Account checkers are a type of software which validates the stolen credentials supplied by the attacker, by running the credential stuffing attack themselves and charging a small amount for each approved credential. They can often be found on the clear web, though are usually only available for a limited number of large targets, who have millions of potential accounts to compromise.

Taking over of the compromised accounts is the next to last step before attempting monetization.

Hackers may then either try selling the credentials or further exploit the data by undertaking additional attacks enabled by the obtained personal information, by for instance targeting specific companies or organizations (Esparza, 2019).

Cash-out. When opting to sell the credentials, attackers have the possibility to find buyers directly from their criminal connections, thus keeping the data confidential for longer and trading it for a higher price (Shape Security, 2018). Once the attackers and their private network have exhausted the credentials, they may go on to advertise them on the dark web (Esparza, 2019; Shape Security, 2018; Wilson, 2019). Cybercriminals who do not have an extensive set of contacts, or access to more private credential distribution channels, might offer the data for sale to other more professional cybercriminal groups, who could get more value out of the compromised credentials. Additionally, they could post the data for sale on forums and markets, pricing it lower than the average, or providing a sample from the data for others to validate, making the data leak public (K. Thomas et al., 2017). Some of these retail websites can be found on the clear net, while a big portion are based in the dark web.

2.4.4 Main Types of Service Fraud

Data acquired by criminals can be broadly categorized as financial, corporate, or personal data (Wilson, 2019). Financial data frequently consist of payment card details, bank account numbers, salary information, or payroll records. Corporate data, which is a highly sought commodity, comprises personal information about employees, management personnel and board members; data concerning external parties, such as suppliers, and clients; or other sensitive information like trade secrets, financial data, operational strategies.

Personal data may consist of PII (Personally Identifiable Information), medical data, or contact details (Wilson, 2019). A portion of this, which essentially constitutes a commodity of its own and is at the center of the current work, is the trade of account credentials. As these may range from accounts providing access to various services, to loyalty program credentials, the following section provides a brief overview.

E-commerce

Regularly cited as an industry targeted by cybercriminals, retail has been reported as having as much as 90 percent of its overall login traffic as credential stuffing attacks (Shape Security, 2018). The reasons for this have to do with the industry's focus on maximizing profit by attracting and retaining clients. As e-commerce websites strive to reduce the possibility of customers giving up midway through purchases and not going through with their orders, they make every effort to ensure users have a seamless experience. Thereby, the introduction of additional, and possibly more complicated security mechanisms as part of the process could hinder customers' convenience, making online retailers averse to tightening their security (Shape Security, 2018).

Furthermore, businesses are continuously developing and adding various services to entice more customers, which opens up new possibilities for cybercriminals, especially for monetizing the stolen accounts. In recent years, retailers with physical stores are increasingly adopting 'Buy online, pick up in store' practices, which considerably ease the process of cashing out a purchase made with a fraudulent account (Shape Security, 2018).

Moreover, since payments through mobile apps are gradually becoming more common, so are methods of abusing this service, which applies to online retail as well. One possible scenario is that after the criminal has obtained a valid account, they visit the physical store of the company and use the available balance or the payment method provided in the account, thus avoiding the delays and risks related to having items delivered to a specified address. Subsequently, the fraudsters can resell the purchased products, or try and carry out return fraud.

Besides monetizing through the physical front of a business, cybercriminals frequently cash out by purchasing digital goods or gift cards, which are easily transferrable and resold online.

Loyalty Programs

Other regularly suffering from credential theft industries are the airline and the hospitality, and more specifically, their loyalty award programs. It has been observed that there has been a surge in loyalty fraud, as the percentage of attacks has nearly tripled over the course of one year (PYMNTS, 2018). Accumulated air miles and loyalty points are exchangeable for air flights or hotel reservations, which presents a lucrative deal for criminals who can easily convert the points into real currency. What makes loyalty program accounts attractive for credential stuffing attacks is the low login rate of customers, as they access their accounts on a significantly less regular basis than other online services. This could considerably extend the time period between an account breach and its discovery (Armor, 2018). Furthermore, these businesses often do not have complex cyber protection in place: the majority require users to have basic credentials consisting of merely

a username-password combination and no other site-specific information, or other advanced authentication (Shape Security, 2018). While this may provide customers with a more smooth and convenient experience, it eases the task of the attackers, especially considering the high probability of these credentials being used on other unrelated platforms (K. Thomas et al., 2017).

The freshness of the accounts is of great importance for cybercriminals, as the more time passes from the account takeover, the higher the chance the breach may be detected and the accounts suspended (Blueliv, 2018). However, finding buyers for miles and other loyalty program points is not a straightforward process, since these are fairly specific goods. This has led to the rise of another criminal service: mileage brokerage (Armor, 2018). The role of a mileage broker is to mediate the sale of the award points between the cybercriminal and the prospective buyer. After the attacker has successfully performed the credential theft, they would sell the access to the compromised miles account to a broker, who verifies the account and further proceeds with monetizing it.

Two methods for utilizing air miles have been reported (Shape Security, 2018): the first one hints that brokers find buyers themselves by advertising the miles on the internet, including on the dark web. The other method suggests that they cooperate with discount travel agencies, which offer heavily reduced fares for a fraction of the original price. The inclusion of an additional party in the distribution channel has the added effect of making it even harder to trace the miles back to the attackers.

Further options are available to cybercriminals in possession of hotel loyalty accounts: one common alternative is exchanging the points for items such as gift cards offered by the hotel's partner network or own gift shops (Armor, 2018). Moreover, the introduction of various mobile services by hotels aimed at achieving a higher competitive advantage in a vastly transforming industry, have expanded the monetization opportunities available to criminals. The increasing use of hotel apps for checking in, paying, and as digital room keys, has greatly reduced guests' interaction with hotel personnel and diminished the need for personal identification, leaving fraudsters at an advantage for impersonating unsuspecting clients (Shape Security, 2018).

2.5 Identification of Company Characteristics Driving Cyber Fraud

This section summarizes some of the characteristics mentioned in the literature as likely to have an influence on the target selection of companies and industries. As service fraud is still relatively weakly researched topic, such metrics were mainly found in studies examining various cyberattacks targeting firms and organisations, usually leading to loss of personal information. As resources were used scientific articles, as well as documents published by security-related companies and organisations.

A particularly prominent research area seems to be how cyber attack incidents influence firm value and public perceptions, and how corporate policies are adjusted as a result of an experienced security breach. Various firm characteristics have been suggested as indicative and described in this regard. However, few studies examining how company factors affect target selection were found. Metrics from these references were reviewed and extracted as relevant for this thesis. The results are presented in Table 2.2. The various characteristics are grouped according to the three aspects of the target suitability concept as outlined by RAT.

Table 2.2 Organizational Characterisitcs Influencing Cyber Fraud

Characteristics	Literature reference
Value	
Size, measured by: • Profits • Return on Assets (ROA) • Asset intangibility • Net income • Revenues • Total assets • Total equity • Market capitalization • Number of employees • Number of customers	The rise in targeted attacks, which are directed towards specifically chosen victims, in recent years in comparison to opportunistic attacks (Kshetri, 2005) has resulted in an increase in cybercriminal activity aimed at large corporations (Kaspersky, 2018). It is reported that nearly seventy percent of targeted attacks are against larger companies (Verizon, 2012). Companies turning in larger profits (higher ROA), having higher asset intangibility, and fewer financial limitations are reported as more attractive for criminals (Kamiya et al., 2018). On the other hand, smaller companies are more often the victims of opportunistic attacks (Verizon, 2012). Smaller firms seem as attractive targets, since they have more limited security budgets; may consider that are at a lower risk and consequently employ less stringent security controls; or be used as a way of gaining access to a larger third-party company (Hayes & Bodhani, 2013). Various metrics for measuring company size have been reported by Dang & Li (2015), who have performed an analysis of the most commonly used measures of firm size in corporate finance: net income, total assets,

	market capitalization, number of employees, net assets (Dang & Li, 2015). Commercial rankings such as Forbes Global 2000 use assets, sales, profits, and market capitalization, while Fortune 500 make use of only sales and profits in measuring company size.
Industry	Sen & Borle (2015) conclude that the industry an organisation is in affects the chances of it being targeted by criminals.
Digitization of value	Companies which have a higher dependence on digital technologies to conduct business are at greater risk of cyberattacks, as they have an increased exposure of their assets and sources of value (Kshetri, 2005).
Visibility	
Popularity	A report by Kaspersky (2018) specifies the popularity of a company as a reason for it to be more targeted by cybercriminals. Website domain popularity has previously been used as a measure of popularity in exploring target selection in financial institutions (Natalius, 2018; Hoppenreijs, 2019).
Reputation	Kamiya et al. (2018) highlight that firms which have a greater public image, as measured by their presence among the Fortune 500 companies, are more likely to be a target of a cyber attack.
Symbolic significance and criticalness	Kshetri, (2005) argues that the more an organisation is seen as significant or dealing with a critical infrastructure, the more likely it is to become a cybercrime target.
Accessibility	
Security management committee	Kamiya et al. (2018) indicate that firms which have a risk committee on the board are less targeted than those which do not.
IT security budget	Sen & Borle (2015) report that higher investment in IT security is counter-intuitively related to an increased risk of a data breach. They argue that the reason for this could be that investments are made inefficiently by focusing on one type of controls over others. For example, investing more on technical controls, rather that administrative or physical.

Industry competition and future growth opportunities	Companies experiencing less strong market competition and having more opportunities for growth are more targeted (Kamiya et al., 2018).
Location	It has been established by Romanosky et al., (2011) that the physical location of an organisation could influence cybersecurity incidents. This could be because different states and countries adopt various in severity laws and regulations, which could in turn have a deterrent effect on crime. Moreover, (Sen & Borle, 2015) argue that societies with stronger markets and economies could attract more cybercriminal activity, as it may lead to higher gains for the criminals. It has been investigated that the average total fraud loss differs by geographical regions (Verizon, 2012).
Weakness of defense mechanisms	As cybercriminals often look for unfixed software holes to conduct an attack, such mishaps could leave a company at a higher risk (Kshetri, 2005).

2.5.1 Selection of Characteristics

A selection has to be made from the characteristics described in the previous section for further analysis. Considering the novelty of the topic, no previous research evaluating the influence of the outlined characteristics in this type of cyber fraud was found. It would not be possible to explore the effects of all factors in this thesis due to time constraints, as well as availability of information. Among the main criteria for selection, thus, is the feasibility of collecting the required data, within the limited time period for this project.

In order to examine the influence of any of those factors on target selection, additional data has to be collected about all companies chosen for the analysis. Oftentimes, such information is difficult to obtain: it may not be possible to get access to it free of cost, or it may not be available at all. Privately-held companies are not required to publish official reports. Furthermore, there is no common repository containing financial data for all investigated entities, it is mainly scattered over various resources. Other factors, such as digitization of value or symbolic significance, are more subjective, and hard to estimate without an existing robust measurement system.

Exploring characteristics related to cyber security controls and measures taken by affected companies or their budgeting would have provided valuable insights. Unfortunately, it was not possible to collect such information, despite the efforts, as it is usually confidential and organisations are reluctant to provide it.

Based on the explored literature and the outlined constraints, the characteristics selected for further analysis are those most commonly mentioned as representative of the value of a company:

Size. Company size can be expressed by numerous indicators; however, it is not possible for this research to analyse all of them. Therefore, a further selection has been made from the previously listed indicators on the basis of their prominence in the literature:

- **Revenue**: revenue is the income a business makes from its normal business activities, also referred to as sales or turnover. It is considered a stable indicator which is based on actual figures, and not predictions (Hoppenreijs, 2019).
- **Net income**: also referred to as net profit, or net earnings, is the amount of revenue left after all expenses, taxes and costs have been extracted from the sales; used as a measure of a business' profitability and in calculating market capitalization.
- **Total assets**: regarded as one of the most commonly used measures of firm size (Dang & Li, 2015), it represents the total amount of assets owned by an entity.
- **Total equity**: represents the value a company has after its total liabilities have been subtracted from the total assets. It is considered a stable in time, more independent indicator of size, less influenced by the profits a business is making (Hoppenreijs, 2018).
- **Number of employees**: regularly encountered in studies and reports as indicative of size. A higher number of employees could be linked to a greater risk of cyber fraud, as people are often exploited by cybercriminals wishing to gain access to corporate assets. A disadvantage of this metric is that it accounts for full time employees only, and does not include part-time employees, which often form a main part of a business' regular activities.

These indicators were chosen as they represent different facets of an organization and could be considered indicative enough of its size. Additionally, the majority are accessible open source information, included in the balance sheets and annual reports of businesses. As manually exploring each company's financial documentation is a rigorous, time-consuming task, a limitation had to be put on the number of factors which could be examined.

Reputation. Reputation is often regarded among a company's highest valued and most easily damaged assets. As such, it would be interesting to examine what role it plays in the target selection process. The measure of reputation employed by (Kamiya et al., 2018): whether a company is present in the Global 500 list, will be applied to this research.

Other characteristics, such as **domain popularity** and physical **location** of a company are also included in the analysis. The factors identified in the literature, in correspondence with the RAT, as well as the selection made among them are illustrated in Figure 2.6.

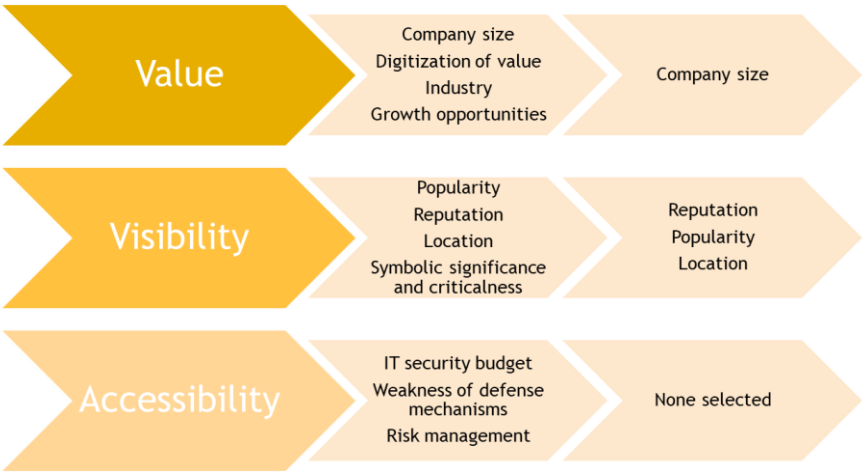


Figure 2.6 Identified and Selected Company Characteristics

Summary of Chapter

The purpose of this chapter was to set the foundation of this research and provide answers to the first two research sub-questions. The main concepts related to the research problem were outlined, and the role of the underground markets in facilitating digital fraud was described. Service fraud, which is at the center of this study, was introduced and placed within the cyber fraud landscape, along with its proposed business model. The five stages of committing service fraud were presented: starting with the collection of stolen credentials through various methods; the development of malicious software and its distribution; and finally, taking over and cashing out of the stolen accounts. As main types of service fraud mentioned in the literature were identified stolen online retail accounts and vouchers, and hospitality and airline loyalty reward programs.

The relation of the Routine activity theory, borrowed from criminology, to target selection in the cyber domain was presented. Furthermore, relevant company characteristics are examined from the literature and placed within the context of the theory. As answer to the second sub-question, a selection of factors potentially affecting target selection and financial losses was made: company size, reputation, domain popularity and company area of service. In order to obtain the necessary data on these characteristics, the initial dataset has to be explored and analysed, which is done in the following chapter.

3 Data Overview and Preparation

This section provides an overview of the dataset used in the thesis. As it included information deemed irrelevant for this work, an explanation of the preparation and selection procedures of the final data needed for the analysis is presented, including the collection of additional necessary data.

3.1. Overview of the Dataset

The used dataset includes records collected from eight prominent underground anonymous marketplaces from 2011 to May 2017, and consists of 44,671 listings and 564,204 transactions made on digital goods, grouped in 17 categories. The utilized data is presented in a SQL database contained in two tables: Items table comprising information about the *product category of the sold item, title, vendor, marketplace, time and location stamps*, among others. The second table includes *feedback values* for each purchase, *date of transaction, order amount, amount paid converted to USD* (at an exchange rate at the time of the transaction), *buyer, etc.* As there are no official records or statistics about the volume of sales transpiring on these marketplaces, it has previously been determined that the number of feedbacks is a relatively reliable measure for sales. Since most markets require customers to leave a feedback score after making a purchase, it can be assumed that these feedbacks correspond to the number of transactions which have taken place. Furthermore, data which was not required was removed, leaving the following variables: *product category, marketplace, date of transaction, order title, order amount, feedback value, vendor*, which were used in the following analysis.

All eight examined markets operate in the English language, and were active during various time periods. The dataset contains transaction information starting from the original Silk Road 1, and follows its successors, including one of the largest existing single marketplaces to date: Alphabay. While the trade on the markets revolves predominantly around drugs and similar substances, a portion constitutes digital goods, such as malware, bots, fake and pirated goods, as well as various types of information. The cybercriminal listings in the dataset have previously been classified by van Wegberg et al. (2018) into several categories, differentiated by whether they are aimed at other criminals, as in business-to-business, or to be used personally. This work focuses on the retail side of the trade, and as such that information was further examined. The data in question in the original dataset is grouped in the following categories: Accounts, Fake, Guide, Pirated, and Voucher, along with Custom and Others. An overview of the categories and the main data they consist of is provided in Table 3.1, along with the original number of listings in each category and the remaining number after the filtering process, explained in the following section.

Table 3.1 Categories in Provided Dataset, Type of Listings, Initial Number of Entries, and Entries after Filtering.

Category	Description	Initial entries	After filtering
Accounts	Accounts for media streaming services; pornography websites	3805	1071
Fake	Fake items: ID cards, passports, money	3429	23
Guide	Guides for various illegal endeavors; hacking; personal development	5097	63
Pirated	Pirated software and other digital content	1431	148
Voucher	Vouchers and gift cards for various stores and restaurants; lottery tickets	1305	762
Custom	One-time buyer-specific products or services	6378	28
Other	Miscellaneous items: clothing and accessories replicas; drugs; documents; cash	8491	223

In order to provide better understanding of the database, screenshots of the two utilized tables of the dataset are given below.

	hash_str	category	marketplace	title	vendor	vendor_hash	total_sales	first_observed	last_observed
	Filter	OTHER	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	00005f75...	other - guide	Silk Road 2	3-5g-devils-cocktail	sweettganjab...	a3c24c810...	584.09	2014-07-02	2014-10-29
2	0005ea2...	other	Evolution	Jason Ferruggia - The Renegade Diet (...	SteroidWareh...	305012dbc...	0.96	2015-01-14	2015-01-29
3	000720c3...	other - guide	Evolution	(eBook)	optiman	50e0ae980...	1.63	2014-11-16	2015-01-13
4	0008018...	other	Silk Road 2	rc-starter-kit-milligram-scale-and-tiny-s...	ReconnoiterC...	4842d0a30...	162.54	2014-02-05	2014-02-12
5	000cb18b...	other - guide	Agora	Ebay Auctions Theft, Make Money From...	passman	58af38ccad...	308.29	2014-01-17	2015-02-04
6	001146fa...	other - acco...	Silk Road 1	uploaded.to - 3 Months Premium Memb...	somono	ad22e1f2b8...	28.76	2012-05-22	2012-07-02
7	0015b9d...	other	Alphabay	SEVENFRIDAY - P series P2-1【KW-Fact...	sexyhomer	606106245...	2587.0	2015-12-03	2017-05-03
8	00164b0f...	other - guide	Silk Road 1	MySQL Database Usage & Administration	HipsterPikachu	5809deda7...	0.74	2013-04-07	2013-04-07
9	00177b6...	other	Alphabay	Rolex - MILGAUSS 40MM Special editio...	sexyhomer	606106245...	356.0	2017-01-08	2017-04-17
10	0017abec...	other - guide	Alphabay	Fiverr Bomber	etimbuk	398e7e66e...	8.0	2015-10-17	2015-12-19
11	001a67c6...	other - custom	Alphabay	Travel discount 30% flights- custom for...	gugustiuck	22218bd09...	198.0	2015-12-26	2015-12-26
12	001d9a0...	other	Evolution	aLantern2Lifver	RosesGarden	b7700e14a...	951.72	2014-10-14	2015-01-21
13	001fcbe8...	other	Alphabay	WonderWoman AAAAA+++++ (14Gram)	AbraKadabRa...	a568862e8...	297.8	2017-05-02	2017-05-02
14	0026c182...	other - pirat...	Agora	Windows 7 AIO (Win 7 + Office 2010 3...	wakeside917	e64c2c2b8...	69.25	2014-02-16	2015-01-13
15	0026f4c3...	other - custom	Alphabay	Custom for scotsman2015	HulkedBenzo...	eb68e9803...	394.14	2015-11-04	2016-03-07
16	0027957...	other - custom	Agora	LAMPLITE CUSTOM!	FattusCattera...	20f1f8fbb6...	198.54	2014-03-04	2014-03-04
17	0027b13...	other - acco...	Alphabay	UFC FIGHT PASS ACCOUNT (lifetime)	TheTopDigital	009d1600b...	15.0	2015-12-08	2015-12-10
18	002bb75...	other	Alphabay	Same Day Shipping! For Tuesday and T...	FarmersUnion	4264878bb...	10.0	2016-12-20	2016-12-20

1 - 19 of 29918 Go to: 1

Figure 3.1 Items table

	hash_str	marketplace	item_hash	date	giver	reciever	message	order_title	feedback_value	order_amount	order_amount_usd
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1ad9e323...	Silk Road 1	001146fa...	2012-05-22	Unknown ...	somono	I love you.	uploaded.to - 3 Months Premiu...	5.0	0.0	5.65
2	65465d3a...	Silk Road 1	001146fa...	2012-05-22	Unknown ...	somono	:)	uploaded.to - 3 Months Premiu...	5.0	0.0	5.65
3	a2a7d504...	Silk Road 1	001146fa...	2012-07-03	Unknown ...	somono	top.thanks	uploaded.to - 3 Months Premiu...	5.0	0.0	5.68
4	b040eca1...	Silk Road 1	001146fa...	2012-06-23	Unknown ...	somono	nice fast	uploaded.to - 3 Months Premiu...	5.0	0.0	5.59
5	cec62f6ec...	Silk Road 1	001146fa...	2012-06-15	Unknown ...	somono	top	uploaded.to - 3 Months Premiu...	5.0	0.0	6.17
6	5a4920b4...	Alphabay	0027b13a...	2015-12-10	t...7	TheTopDig...	Fast and exac...	UFC FIGHT PASS ACCOUNT (life...	0.0	5.0	5
7	5e0346bb...	Alphabay	0027b13a...	2015-12-10	b...r	TheTopDig...	works perfect ...	UFC FIGHT PASS ACCOUNT (life...	0.0	5.0	5
8	7ef6e1a4...	Alphabay	0027b13a...	2015-12-08	j...9	TheTopDig...	No comment	UFC FIGHT PASS ACCOUNT (life...	0.0	5.0	5
9	0092ce8a...	Silk Road 2	0035d0f7...	2014-05-15	Unknown ...	DrawkwarD	Fast delivery ...	/items/directv-premium-accoun...	0.0	0.0	10.09
10	01c36bb2...	Silk Road 2	0035d0f7...	2014-01-22	Unknown ...	DrawkwarD	Ordered 8 ite...	/items/directv-premium-accoun...	0.0	0.0	10.43
11	04401652...	Silk Road 2	0035d0f7...	2013-12-29	Unknown ...	DrawkwarD	customer fro...	/items/directv-premium-accoun...	0.0	0.0	10.37
12	05149652...	Silk Road 2	0035d0f7...	2014-09-13	Unknown ...	DrawkwarD	works like a c...	/items/directv-premium-accoun...	0.0	0.0	9.88
13	06ab58b2...	Silk Road 2	0035d0f7...	2014-03-13	Unknown ...	DrawkwarD	Perfect produ...	/items/directv-premium-accoun...	0.0	0.0	9.89
14	09f47085...	Silk Road 2	0035d0f7...	2014-09-16	Unknown ...	DrawkwarD	Great seller, g...	/items/directv-premium-accoun...	0.0	0.0	10.11
15	0b35454f...	Silk Road 2	0035d0f7...	2014-07-22	Unknown ...	DrawkwarD	have yet to tr...	/items/directv-premium-accoun...	0.0	0.0	10.21
16	0dc04409f...	Silk Road 2	0035d0f7...	2014-08-11	Unknown ...	DrawkwarD	Works! Thank...	/items/directv-premium-accoun...	0.0	0.0	9.88
17	1119a7fd...	Silk Road 2	0035d0f7...	2014-08-31	Unknown ...	DrawkwarD	Excellent	/items/directv-premium-accoun...	0.0	0.0	10.02
18	113dbe8a...	Silk Road 2	0035d0f7...	2014-05-03	Unknown ...	DrawkwarD	Perfect	/items/directv-premium-accoun...	0.0	0.0	10.18

1 - 19 of 304188 Go to: 1

Figure 3.2 Feedbacks table, including transaction information and price

3.1 Data Selection

3.2.1. Main Dataset

The focus of this research is on revenues made and lost due to the underground activities; therefore the relevant information had to be extracted from the full dataset. The data from the categories outlined above was inspected manually as the nature of the listings makes it impossible to automate the task. As illustrated in Figure 3.1, the titles of the listings may include various jargon terms, abbreviations, unrelated keywords, and other incomprehensible phrasing. Out of the original 29,918 records, 2,318 were selected for the next part of the analysis. The other listings were excluded as irrelevant to the research topic for one of the following reasons:

- listings concerning drug offerings which had slipped through in the digital goods categories during the initial categorization of the dataset;
- incomplete listings, where it is unclear what is the offered product, or which is the concerned company;
- pornography accounts, as they were considered out of the scope of this research as they cannot be traced back to corporate entities;
- unrelated listings: invitations for underground markets, lottery tickets.

As this study is mainly concerned with account, loyalty program and voucher fraud, the categories of Accounts and Voucher were considered as central to the subject, yielding 1,883 records. However, the remaining categories were also inspected for entries which could be relevant, which resulted in the addition of 485 more records.

The next step of the data preparation process entailed categorizing the selected listings depending on the type of the related product or service. This resulted in four categories: Accounts, Loyalty programs, Vouchers, and Pirated software. The first three have been briefly mentioned in industry reports, as displayed in the previous chapter, while the Pirated software category was added due to the recently adopted by the industry subscription business model, which has gradually replaced the previously common licensing model, thus obtaining characteristics similar to having a service account. Four subsequent datasets were formed based on this distinction.

Consequently, the entries from each category were sorted according to the company or business which was affected, in order to provide a notion of which organizations were targeted the most. A further selection was made, by including only those which had more than ten entries in the database per category, thus excluding the companies which had comparatively weak presence. Based on this classification, the sales information derived from the feedbacks table was selected and added to each dataset, hence providing data for the number of transactions for each listing, and the exchanged amount.

The final step of the data preparation process is the anonymization of the selected entities. As some of the information concerning the affected companies can be considered sensitive, the name of each company was replaced with an alphanumerical, based on the category the firm was placed in. When a company is represented by two products, or in two categories, that is denoted in its name.

Table 3.2 Listings and Companies, Initially and after Selection, and Transactions of Selected Companies.

Category	Number of Listings		Number of Companies		Number of transactions of selected companies
	Initially	Selected (percentage of all)	Initially	Selected	
Accounts	1067	711 (66.6%)	121	22	30,038
Vouchers	887	524 (59.1%)	196	16	8,816
Pirated Software	317	288 (90.9%)	23	7	7,121
Loyalty programs	154	94 (61%)	29	7	800

Table 3.2 presents the number of companies and the number of listings observed in the dataset after the processing phase. It is apparent that the majority of the listings and sales are in the two categories: Accounts and Vouchers, while Software and Loyalty programs garner significantly less attention. Similarly, the number of companies follows the same pattern: accounts and vouchers are represented by more than one hundred companies each, whereas the other two categories have records related to a few dozen companies. The greatest variety in targeted businesses is within the vouchers category with 196 entries, although the number of listings in vouchers is with around 20 percent less than in accounts.

After the selection process described in the previous section, the number of companies was significantly reduced, although still retaining about 60 – 90% of the listings, depending on the category. Consequently, there was a high amount of companies appearing only a few times on the markets, while several products were immensely popular. The number of initial listings and companies and those selected are further illustrated in Figure 3.3.

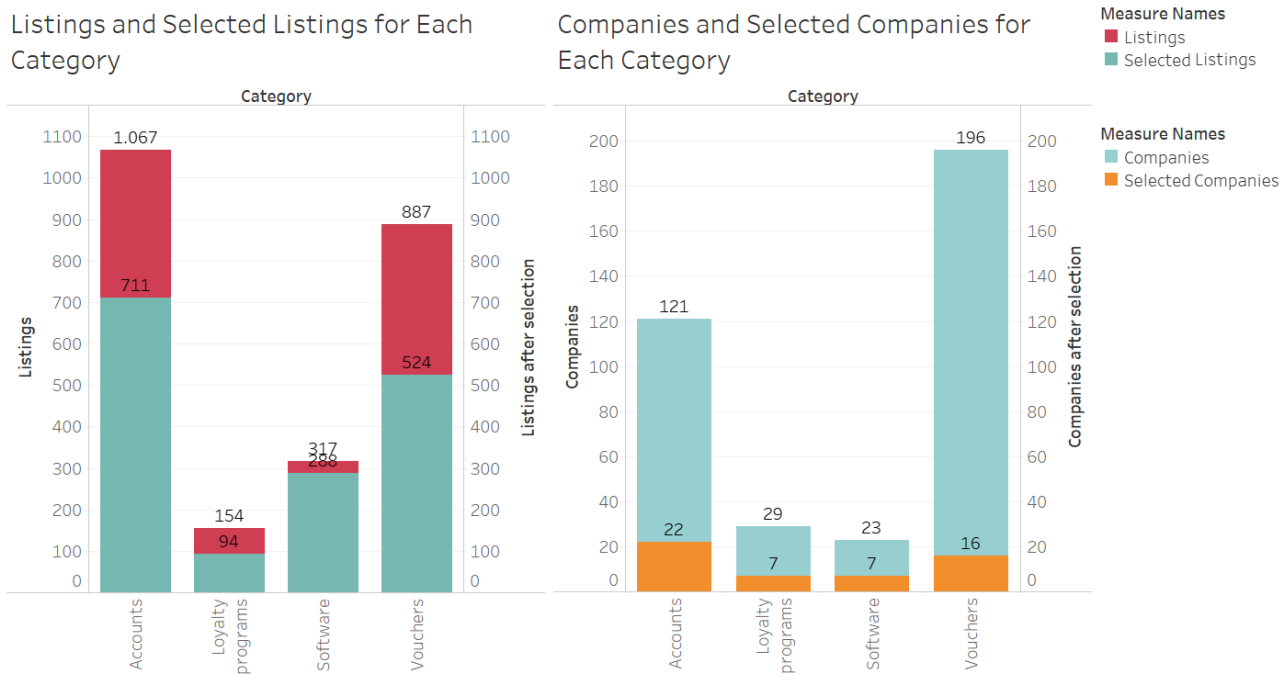


Figure 3.3 Listings and Companies, Initially and after Selection, per Category.

3.1.1 Complementary Costs Dataset

To be able to fully examine the impact of service fraud, additional information related to the affected parties was needed. As the financial losses caused by the activities on the dark web are a focal point of the research, information on the official pricing for each service examined is necessary in addition to the price data from the underground markets. This information was gathered manually by exploring the websites of the various selected companies and services and estimating the demanded price in relation to the product offered on the dark web. More details over the estimation process are provided in the next chapter.

Summary of Chapter

This chapter presented the initial dataset used in this thesis. It consists of transaction data obtained from eight underground markets, in the period 2011-2017. The dataset includes two tables: the first one comprises information on the following variables: product category, marketplace, date of transaction, order title, vendor. These variables were used to filter and categorize the listings, resulting in four categories of service fraud: accounts, loyalty programs, pirated software, and vouchers. The most often targeted companies from this list were selected for the next phases of the analysis, yielding 47 companies, representing 60 to 90% from the listings in each category. Based on this selection, data on order amount in USD, and feedback values per listing was extracted from the second table in the dataset for each company and merged with the other data.

In order to be able to estimate losses incurred by each of the selected companies, additional information was collected from company websites about official price of the illegally traded products. An illustration of the data preparation process is given in Figure 3.4:

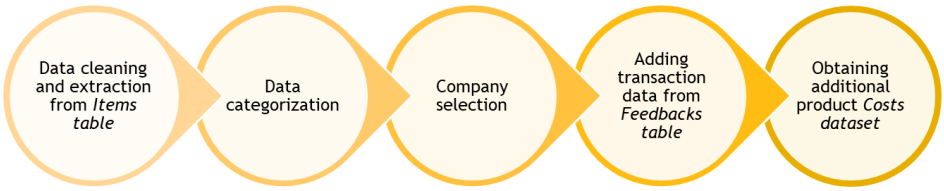


Figure 3.4 Initial Data Preparation Process

4 Descriptive analysis

In the previous chapter the research dataset was introduced, along with the selection process for the descriptive analysis. The main parameters which are to be used in it were outlined: 1. the targeted company; 2. number of sales; 3. number of listings; 4. price per listing; 5. timestamp of each listing; 6. feedback left per transaction.

As part of the research question concerns the estimation of revenues made on the underground markets from service fraud, and its financial impact on the affected companies, the following general descriptive analysis aims to provide some insights on the matter.

4.1 General Analysis

The results after classifying and filtering the listings are summarized in Table 4.1. The data is based on the selected top companies, and is therefore an approximate representation of the values, as the entries with low level of listings are not accounted for. The first column of the table representing the total revenue made on the markets, shows a definite advantage of the Vouchers category over the other three in accumulated revenues. This could be due to the fact that there is a relatively large number of listings in this category having a significantly higher price than in the others.

In order to examine the price distribution in each category, further in Table 4.1 are listed the price ranges per listing: the mean, median, minimum, maximum values and the standard deviation. It is evident that the categories of accounts and software have considerably lower mean and median prices than the listings in voucher and loyalty programs. The vouchers category also contains the most expensive listings and the greatest variety in pricing with offers going up to $1400 per listing, closely followed by loyalty programs. In contrast, the pirated software category has consistently low prices over most listings. A possible explanation for this could be that the listings in the accounts and software categories are predominantly offers for a relatively similar service or identical accounts with constant pricing. Whereas, the prices of the loyalty programs accounts are based on the amount of points accumulated in each account, which can differ widely. Similarly, the vouchers and gift cards offered for sale have different face values, ranging from $10 to over a thousand dollars per voucher, leading to the high variability.

The final measure reported in Table 4.1 is the lifespan of the listings, again represented by mean, median, minimum and maximum values. It is presumed that hijacked accounts and stolen credentials have fairly low lifespan, as they become less usable with time, often after the first user has taken advantage of the product, as that can alert the company or the rightful user of the account (van Wegberg et al., 2018; Shape Security, 2018). However, the median lifespan of most listings is quite high, with the majority being on the market for more than a month. This could be explained by vendors supplying distinct products to different customers, while keeping the same advertisement. The one category with lower lifespan is loyalty programs. It is curious to note the maximum lifespan in the accounts category, where some listings remain active for over 700 days, or approximately 2 years. This supports the notion that criminals update the listings, which can also be noticed from the descriptions of the ads, offering ‘lifetime warranty’ and substitutions.

Table 4.1. Transactions, Listings, Revenue and Lifespan per Company

Company	Number of Transactions	Number of Listings	Total Underground Revenue ($)	Price Per Listing ($)					Lifespan in Days			
				Mean	Median	Min	Max	SD	Mean	Median	Min	Max
A1	7866	172	29140.85	3.7	2.99	0	220	5.05	86.15	36	0	709
A2	7283	114	30674.27	4.21	3.94	0	33.09	4.11	102.88	45	0	732
A3	2739	57	8763.49	3.2	2	0	15	1.89	97.9	34.5	0	575
A4	2510	36	14139.36	5.63	4.99	0	15.34	2.81	166.79	115	0	740
A5	1978	36	15324.56	7.75	7.89	0	40.12	3.58	161.26	52	0	737
A6	1555	22	3353.22	2.16	1	0.5	30	2.06	119.17	51	0	375
A7	1329	29	9356.58	7.04	8.99	0	15	3.33	165	82	0	754
A8	932	23	8997.62	9.65	9.99	0	35	5.3	102.2	24	0	527
A9	846	17	3962.42	4.68	3	0	11.44	3.15	144.31	56	0	736
A10	741	21	5661.77	7.64	8.99	2.99	25.12	2.68	159.07	21	0	604
A11	725	13	3655.06	5.04	3	0	26.4	3.18	183.14	128	0	611
A12	320	12	2186.93	6.83	4.99	0	14.99	2.57	173.9	112.5	6	688
A13	239	18	562.02	2.35	1.05	0	35.86	2.8	102.56	40	0	617
A14	226	16	1223.51	5.41	4.5	0.01	26.63	3.08	115.2	49.5	0	540
A15	200	22	2820.47	14.1	6.99	0	462	48.65	76	49	0	318
A16	190	28	510.72	2.69	1.8	0	15.23	2.64	50.15	20	0	407
A17	152	17	638.37	4.2	4.96	0	5.99	1.56	81.33	21	0	493
A18	76	15	859.03	11.3	5	0.1	200	27.11	49.2	14.5	0	228
A19	41	11	477.11	11.64	12.19	4.36	22.62	4.78	29.5	17.5	0	88
A20.1	37	10	305.93	8.27	4.17	0	29.18	7.03	51.18	34	0	255
A21.1	33	13	4705.25	142.58	5.17	0.01	647.34	266.06	11.1	4.5	0	56
A22	20	9	215.46	10.77	10	6.74	22.9	3.88	29.1	4.5	0	223
L1	234	18	3220.03	13.76	2.5	2.5	293.65	34.69	45.74	0	0	257
L2	201	12	15818.95	79.49	35	23.95	425	80.24	37.58	14.5	0	128
L3	148	19	3854.5	26.04	20	7.5	225	28.51	135.28	0	0	640
L4	99	14	4125	32.23	15	7.5	300	42.88	169.2	0	0	629
L5	55	10	786.8	14.31	10	4.3	80	12.64	214.11	0	0	650
L6	33	9	2199.32	66.65	74.91	12.46	170	29.75	18.86	0	0	104
L7	30	12	2512.5	83.75	90	7.5	240	61.76	105.56	16.5	0	554
S1.1	2394	95	15242.24	6.37	5	0	223.96	7.99	93.8	30	0	528
S2	2071	108	7692.26	5.48	4.99	0	95	5.9	128.39	65.5	0	597
S1.2	1755	49	4482.66	5.72	4.94	0	69.14	6.63	105.15	76	0	440
S3	413	13	2695.05	5.42	5.02	0	70	4.32	121.3	44	0	315
S4	242	17	436.09	4.69	4.88	0	25	3.77	102.82	14	0	361
S5.1	166	13	513.79	2.31	2.2	0	16.18	1.77	123.91	29	0	464
S5.2	80	21	392.96	2.99	2.06	0	15	2.79	90.38	76	0	213
V1	3775	123	203903.4	54.01	50.4	0	543.75	35.09	20.63	6.5	0	123
V2	925	50	12433.32	13.44	12.5	2	50	6.45	68.29	24	0	326
V3	703	42	23746.74	33.78	30	0	709.61	62.63	23.26	6	0	168
V4	505	26	9086.08	17.99	17.5	2	312.75	14.37	38.2	16	0	179
V5	433	16	4492.44	10.38	9.9	0	35.72	5.12	88.31	102	0	208
V6	369	22	942.91	2.56	2.3	1.99	10	1.15	58.2	21	0	373
V7	342	25	17287.72	50.55	35	8.25	1400	107.22	31.71	15	0	144
V8	260	15	13737.9	52.84	80	0	100	40.14	38.75	12	0	154
V9	230	15	1603.2	6.97	5	0	35	5.41	81.57	54.5	0	375
A21.2	206	14	64051.05	310.93	285	50	2100	142.24	71.73	68	0	218
V10	183	14	3436.61	18.78	7.7	2	306.75	25.51	65.33	9	0	559
A20.2	102	27	9718.71	95.28	25	0.77	2725.8	283.4	33.17	0	0	230
V11	80	28	1445.42	18.07	12.8	4.44	50.43	12.84	16.93	3	0	80
V12	74	14	1133.12	15.31	17.5	2	17.74	4.36	16.85	3	0	96
V13	55	21	3992.62	72.59	67	2.66	247.5	75.11	5.82	0	0	40

Furthermore, analyzing the price distributions, it seems most categories have a relatively uniform distribution, with their mean and median having similar values. Hence, it can be concluded that there are few highly differing entries in the data and most listings of one kind had comparable prices. However, since the mean values for some of the examined companies seemed to have been affected by outliers, the median values are used in the following analysis.

The standard deviation (SD) for each company is also calculated. The lowest SD can be observed in the software and accounts categories, with most values being under 10; while loyalty programs have the highest overall SD, and vouchers have the highest variability with values ranging from 1.15 to 283. This confirms that products sold in the accounts and software categories have the least variation in prices, while the other two categories vary more significantly.

One company in the accounts category has considerably higher SD – A21.1, with a SD of 266, more than a hundred times higher than the rest in the category. Upon examining this case, it was established that 10 orders, or around one third of the completed orders, are under the value of $0.10, whereas one product which was sold 7 times, has a price of $647.34, bringing the SD to a comparatively much higher value than the rest in the category.

In summary, the total revenues made from sales of various fraudulent accounts and vouchers are relatively low, considering the length of the period. Vendors make anywhere between $1 and $80 per transaction, while the majority of the prices are around $10-20. Generally, listings stay on the markets for around a month, though some may be up for more than a year.

4.2 Total Revenue, Sales, Price per Company

4.2.1 Accounts Category

The number of companies having more than ten listings in the accounts category are 22. It can be noted that the majority of the targeted companies are media services providers, such as video and music streaming, telecommunications and sports. Among the top targets there is also a transportation company (A6), an educational service (A11), a video game (A16), and a few e-commerce websites.

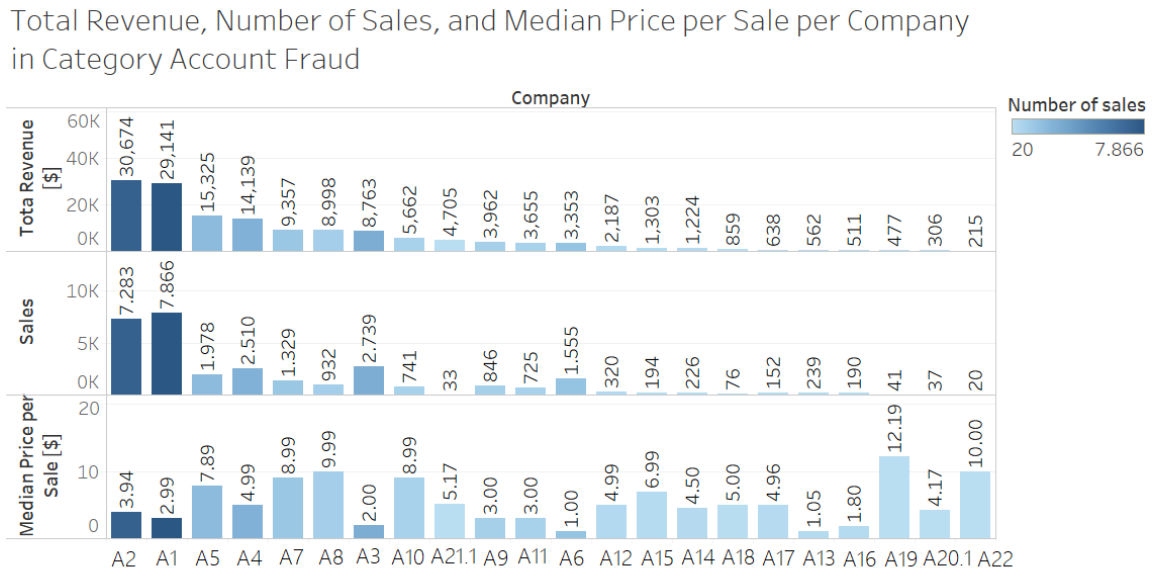


Figure 4.1 Total Revenue, Sales, and Median Price per Sale per Company in Category Account Fraud.

There are two companies which have realized significantly more revenues for the criminals on the underground markets than the others: A2 and A1, which are popular music and video streaming providers, with $30,674 and $29,141, respectively (Figure 4.1). This is nearly twice as much as the next two entries in the list: the telecommunications companies A5 and A4, amassing $15,325 and $14,139, in turn. The rest of the companies on the list have led to less than ten thousand dollars of revenue for their sellers, with the seven least profitable making less than a thousand each.

Furthermore, what is evident from Figure 4.1. is that there is not much of a relation between the median price per transaction and the total amount of money collected from sales for each company affected. The high values for A2 and A1 are mostly due to the large amount of sales made on the accounts for both, as the price they were sold for is on average less than most others in the category: $3.94 for A2, and $2.99 for A1. However, the sales of such accounts are considerably more than for any other service across all categories, with more than 7000 transactions made for each. This could be explained both with the ubiquitous use of the two services, leading to their high demand in the retail underground trade, as well as the wide availability of potential accounts for hijacking.

From Table 4.1 it is clear that the accounts for media services providers and video streaming (A1, A2, A3, A4, A5, A7), especially for sports services (A8, A9, A10, and A12), are among the most widespread items for sale, while accounts for e-commerce websites and online retailers are the least offered (A19, A20.1, A22). Retail accounts usually include previously accumulated amounts to be spend or other financial information, such as credit card details, which could explain the prices they fetch on the dark web markets of around $10, considering these services are officially offered for free. This is especially the case with A21.1, which has made an estimated $4,705 from merely 33 sales.

4.2.2 Loyalty Programs Category

Compared to the accounts category, there are significantly less targeted companies in the loyalty programs category: their overall number is 29, while those which had more than ten listings are seven. Four of these are commercial airlines (L1, L4, L5 and L7), whereas the remaining three are hospitality companies (L2, L3, L6). As can be seen from Table 4.1. the presence of loyalty program accounts in the examined period is relatively low compared to the other categories, and so are the number of listings and the revenues made from them.

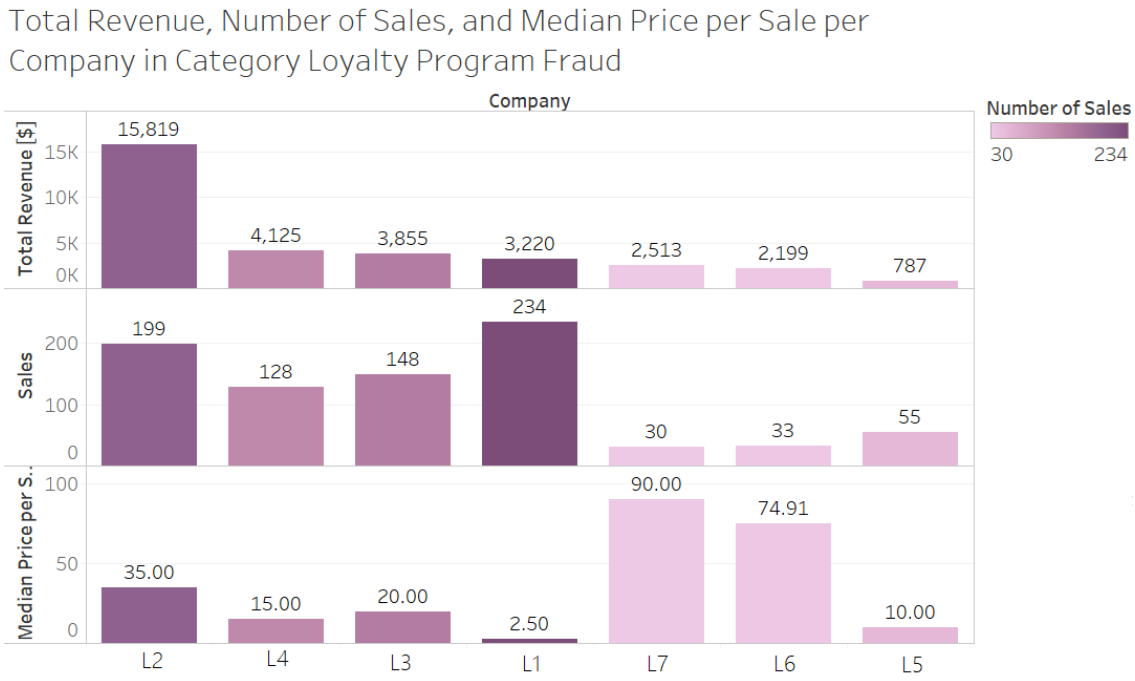


Figure 4.2 Total Revenue, Sales, and Median Price per Sale per Company in Category Loyalty Program Fraud

The most targeted company in the category is the airline L1, and its frequent-flyer program, which has registered 234 sales, made on 18 listings. However, the airline is far from being the most profitable on the underground markets, having brought around $3,220 to those attacking its customers, which is considerably less than the highest realized profit, made from exploiting the hotel chain L2: $15,818. This could be explained with the low value of L1 accounts: their median value is $2.5, the lowest in the category. On the other hand, L2 accounts have a median value of $35, and one of the highest averages of nearly $80. The only low-cost airline featured in the category: L5, has a few accounts, which are being offered for prices lower than the rest, and consequently has accumulated the smallest amount of money: an estimated $787.

Figure 4.2 displays the total profit made on the examined underground markets, and the median prices for each of the selected companies. As can be seen from the graph, the amount of money made on account of each company is primarily related to the number of transactions made, with the top four companies which have the most listings being at the upper half of the list. The two least common accounts: those of the airline L7, and the hotel chain L6, have achieved similar revenues to the top earners despite their low levels of sales, owing to their very high price per

product. There have been around 30 transactions made for each, though at median values as high as $90, and $74.91, respectively.

Furthermore, unlike the majority of the products traded in the other categories, none of the loyalty programs accounts have been sold for free, all having positive minimal values. They have also consistently achieved high maximum amounts, of a few hundred dollars each.

4.2.3 Pirated Software Category

Similarly to loyalty programs, there was a more limited profitability in the pirated software category, despite that the offering was higher. A wide variety of products were sold on the dark web, though the greater part were issued by a few companies, for instance, the listings for S2 include offers for their several products. For this reason, the offers for the various products are grouped under their corresponding publisher. The listings for S1.1 and S1.2 are the only ones listed as separate entities, due to their different nature. The two products are also the most widely traded on the markets, and along with the S2 software significantly outperform the rest in the category by sales and accumulated revenue.

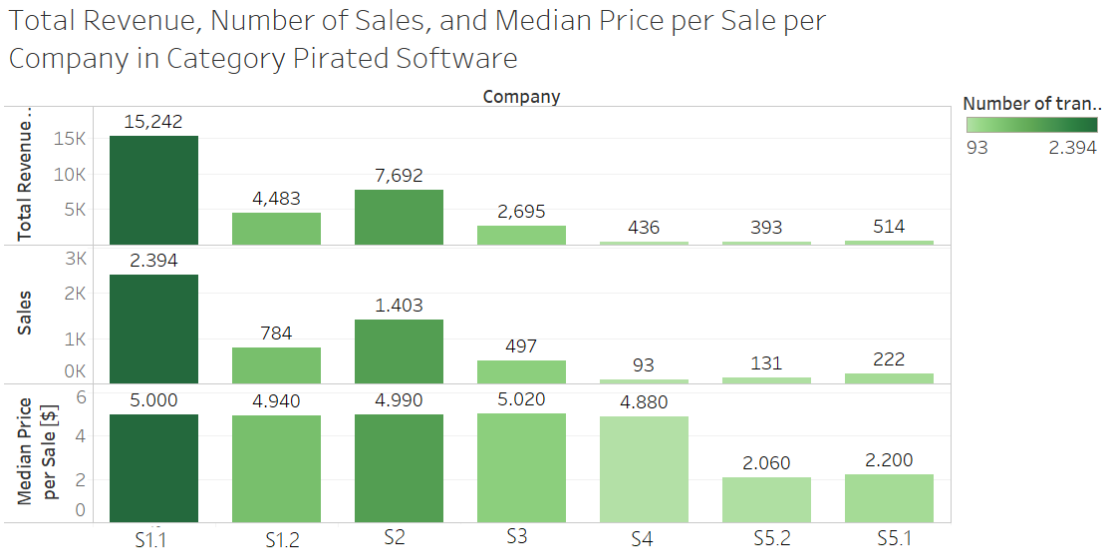


Figure 4.3 Total Revenue, Sales, and Median Price per Sale per Company in Category Pirated Software

What can be noted by exploring the prices of the listings in this category are the very slight differences between them: most of the products were sold for a median price of around $5. The exceptions to this are the listings for the antivirus software S5.1, which sold for $2.2, and the computer tool S5.2, owned by the same corporation, retailing for $2.06. Because of this uniformity, the differences in the money made on account of each company are strongly dependent on the number of sales made for each. Thereby, S1.1 takes the lead with 2,394 successful transactions, contributing to $15,242 of profit for its vendors. The second most targeted company: S2 has made approximately half that revenue with $7,692 from 1,403 transactions, while the third product: S1.2, has accumulated a little under $4,500 from 784 sales. The fourth most targeted item in the

list: the educational software S3, has been sold nearly 500 times, but due to its relatively high price has made $2,695 to its vendors.

It should be noted that computer software, including the products of the here represented companies, has long been pirated and distributed illegally over various forums and sites on the surface net. However, most software products were sold as separate items which the buyer acquired ownership of and could use indefinitely. Nevertheless, new business models such as the subscription and freemium, became the norm; thus, forcing individual and corporate customers who desire to use the latest versions of the products and their offered support to purchase subscriptions, making their accounts more valuable for the cybercriminals.

4.2.4 Vouchers Category

The final category examined in this research is that of voucher fraud, which has also brought the most profit to underground vendors. The majority of the sixteen selected companies in this group are retail stores or chain restaurants, plus a few e-commerce websites.

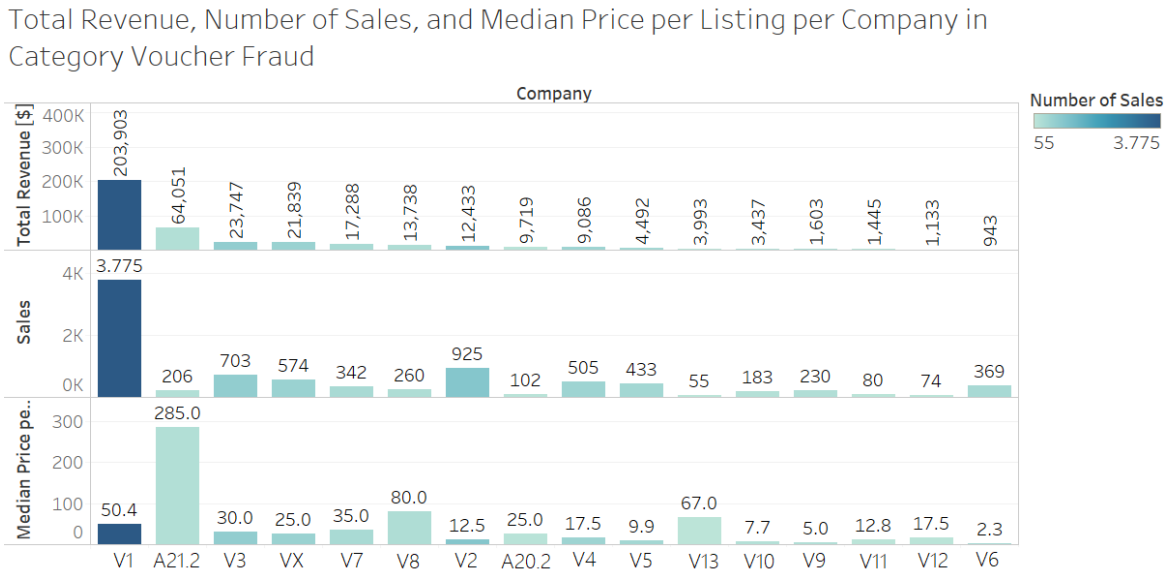


Figure 4.4 Total Revenue, Sales, and Median Price per Sale per Company in Category Voucher Fraud.

Upon examining the results in Table 4.1, it is evident that one company greatly surpasses the rest in offering and realized sales: the supermarket chain V1. Their vouchers, which sold on the underground markets have face values ranging from £50 to £1000, eventually bringing a cumulative profit of $203,903 for their vendors. Except for V1, the criminals on the dark web sold vouchers in significantly lower numbers – of no more than a thousand of each of the other companies. However, as the items in this category have on average the highest prices per listing of all the categories, the total profits are also relatively large.

As can be seen from Figure 4.4, the highest median price per listing is on vouchers of the e-commerce website A21.2: $285, which is also the biggest value across all. The reason for this extremely high value is that 70% of their items are $500 gift cards, thus leading to high average and median prices. The second highest price of $80, which is more than three times lower than

that of A21.2, is on vouchers for an American retail chain store V8, followed by $67 for gift cards for the online retailer V13. Other widely sold items were gift cards for the restaurant V2, with a comparatively low median price of $12.5; and the coffee store V3, averaging a value of about $30. The lowest price was observed on vouchers for the restaurant chain V6, valued at $2.3.

The variety in pricing of the different vouchers, and the relatively high prices, could be attributed to the relation between their list value and their selling price, as the latter is directly dependent on the former. Unlike the listings in the other categories, which offer similar products, such as accounts for the same service, and therefore are expected to have similar prices, gift cards can vary significantly, as observed earlier. Furthermore, gift cards for large amounts would normally command higher selling prices as well, which could bring about for the bigger prices on average.

4.3 Revenue and Sales over Time

In order to assess how the examined types of cyber fraud have evolved, it is useful to examine the accumulated revenue over time. The following figures show the total revenue per month, per company, for the four categories. The general trend is of growth, though, as can be expected, the revenue is strongly dependent on the functioning of the underground markets themselves. The lower values between 2011 and 2013 represent entirely the offering on the first anonymous market Silk Road 1, in the initial stages of the underground trade, and the slight decrease when it was shut down. The following steep rise observed from 2014 to 2015, marks the highest point in sales over the whole period for all the categories, except loyalty programs, which have their peak in the summer of 2016. This increase coincides with the appearance of several new markets: Evolution, Silk Road 2, Hydra, and Agora. Furthermore, the following plunge in the beginning of 2015 is closely related to the disappearance of these markets for various reasons from police take-downs to exit scams (van Wegberg et al., 2018). The subsequent gradual rise in generated revenues is due to the emergence of the Alphabay market, which during its existence until mid-2017, drew in considerable traffic.

The one category which does not follow this pattern, is loyalty programs, as can be observed from Figure 4.6. It should be noted that there are no records for such fraud before 2014, when L6 accounts were first offered for sale. This could suggest that the offering of stolen loyalty program accounts on the anonymous markets appeared a few years later than the interest in the other categories.

The peak in the revenues made from loyalty accounts is in 2016, after a bulk of high-value L2 accounts were traded on Alphabay. This could be related to a large data breach reported by the same company in 2015, when more than half of the company's locations were impacted, leading to the exposure of their customers' personal information and payment card data. Moreover, the

highest point in sales of loyalty accounts is also in 2016, although it is linked to the increased offering of L1 and L4 accounts, which sold for less than those of L2.

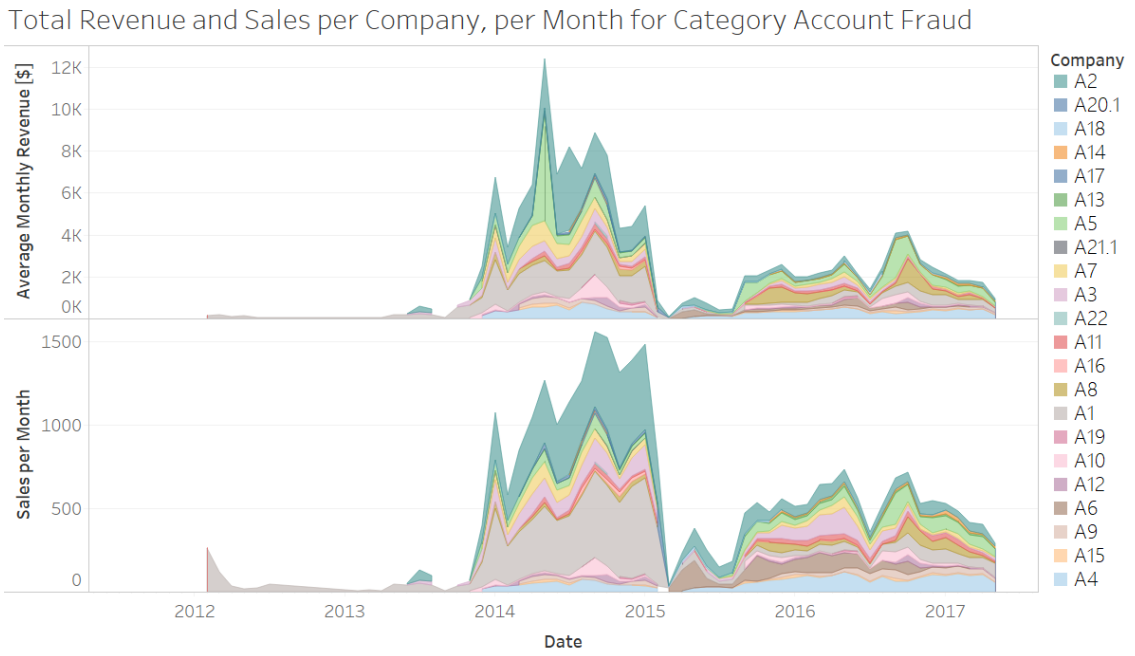


Figure 4.5 Total Revenue and Sales per Company, per Month for Category Account Fraud

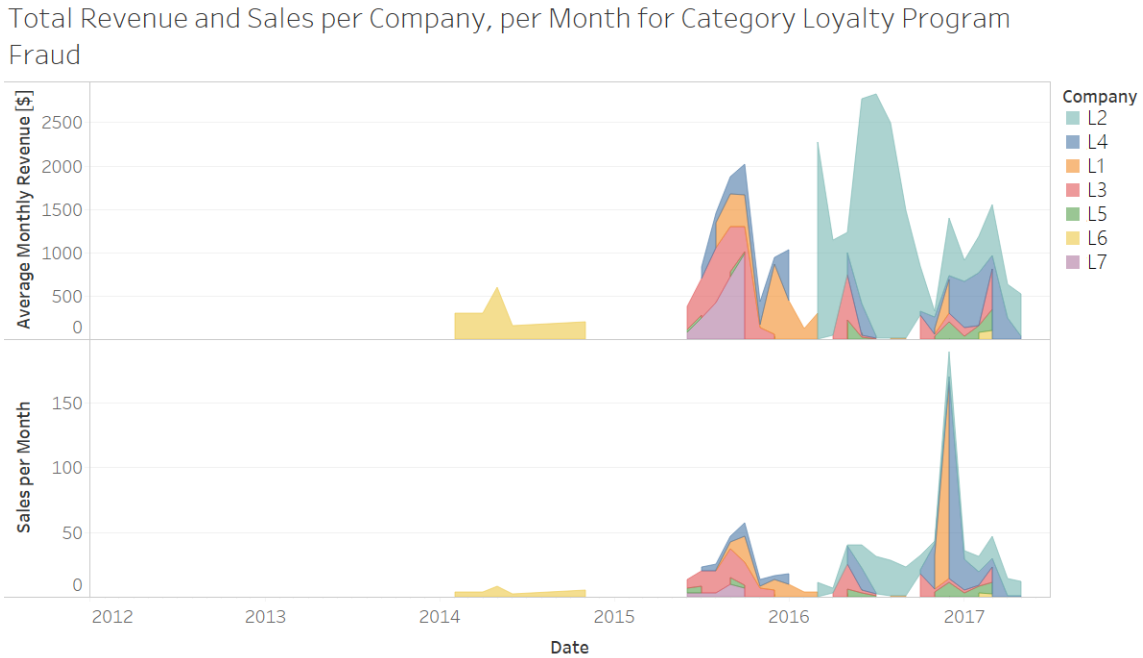


Figure 4.6 Total Revenue and Sales per Company, per Month for Category Loyalty Program Fraud.

What can also be observed from the figures is the duration of the active periods of sale of products for the various companies in the different categories. This could serve to determine whether the retail trade on the dark web markets is a continuous threat for businesses or more of a momentary occurrence.

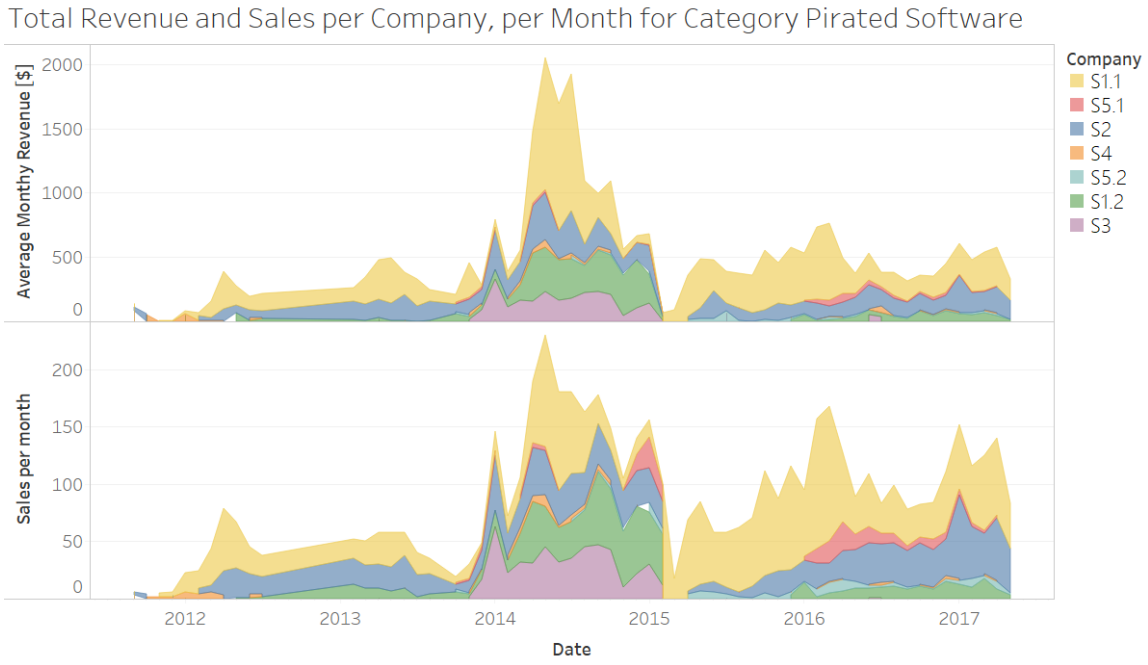


Figure 4.7 Total Revenue and Sales per Company, per Month for Category Pirated Software

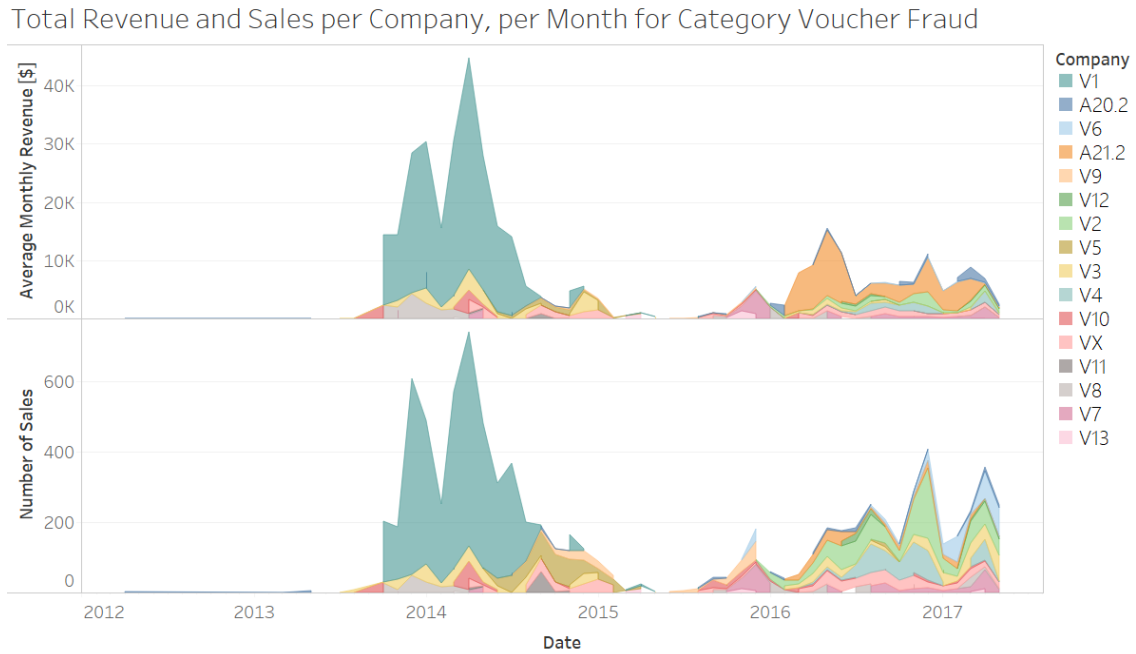


Figure 4.8 Total Revenue and Sales per Company, per Month for Category Voucher Fraud.

As can be seen from Figure 4.5, items from the accounts category were generally offered throughout the larger part of the examined period. An exception to this are clothing retail stores accounts, such as A22, and A19, which were available for a short interval of a month or two. Similarly, the accounts for A21.1 were sold for a limited time, though the amounts they made for a single listing in 2014 were substantial.

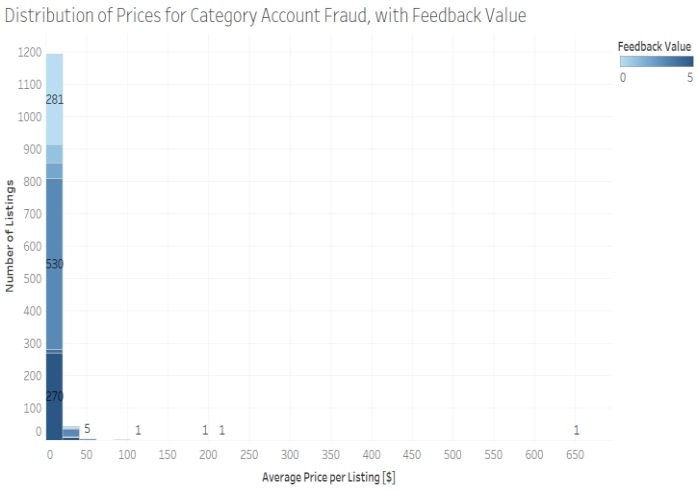
Loyalty program accounts were offered more sporadically, for a few months at a time, likely after data leaks have occurred, as discussed in the case of L2. The hotelier's accounts were also the most consistently sold on the markets: from 2016 until the end of the examined period.

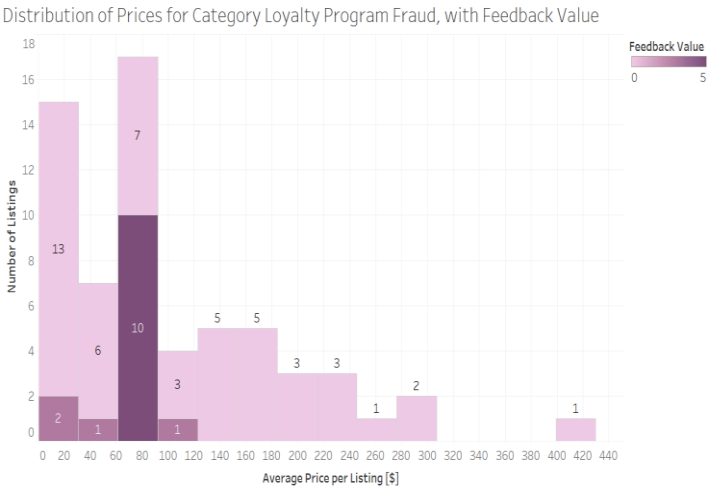
Looking at the pirated software category (Figure 4.7), what makes an impression is that almost all examined products were traded continuously over the span of the whole period. The exception to this are items of the educational software S3, which appeared solely during the peak of the trade in 2014-2015. Additionally, S1 and S2 products seem to hold a stable share throughout the entirety of the period.

Several observations can be made about the distribution of vouchers over time (Figure 4.8): first, similarly to the accounts category, vouchers for retailers, such as V13, and V11, were offered for rather short time intervals. Second, the vouchers for V1, which were the most profitable item on the examined underground anonymous markets, were distributed during the initial uprise in the trade, until 2014, and did not turn up again. Finally, a considerable portion of the coupons for restaurants only appeared on Alphabay, after 2016, suggesting a rising interest in this type of vouchers.

4.4 Price Distribution

Figure 4.9 represents the distribution of the prices of listings for each of the four fraud types. It is clear that three of the categories: accounts, pirated software, and vouchers, follow the same pattern: the greater part of the listings in these categories have prices in the lower end of the range. This is especially true for accounts, where the number of listings priced over $20 is merely 53, compared with 773 valued under $20. The case with vouchers is similar, only the prices are higher: the majority of the vouchers were offered for values under $100, and around 15% of the products were listed for greater amounts. The items in pirated software have a similar distribution, though not as strongly skewed to the left of the graph. In addition, in all categories there are singular listings being offered for exceedingly high amounts, which could be considered exceptions to the majority.





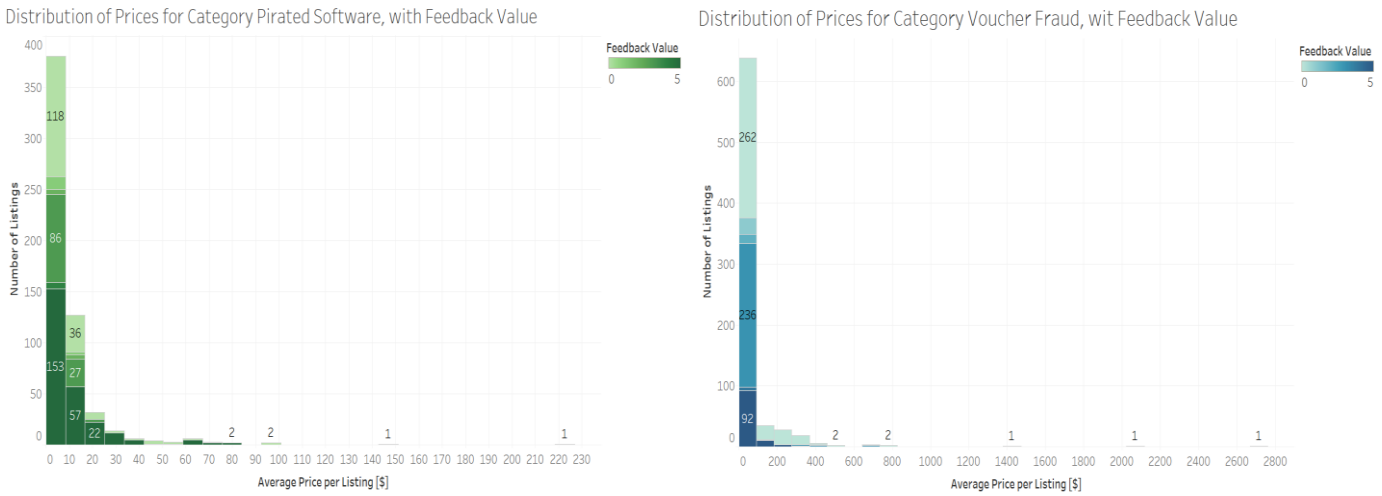


Figure 4.9 Price Distribution, with Feedback Values, per Category. Top left: Category Account Fraud; top right: Category Loyalty Programs. Bottom left: Category Pirated Software; bottom right: Category Voucher Fraud.

The category of loyalty program accounts again shows a different behavior than the others: prices of listings are more evenly distributed over the whole range, proving the higher variety in supply in the loyalty program accounts.

As previously mentioned, with the completion of each sale the customer leaves a feedback for the transaction. These feedback values are also displayed in Figure 4.9. In general, there does not seem to be a relation between customer satisfaction as reflected by feedback ratings and the price of listings or the amounts of sales. There are varying feedback values in each price group, including a big portion of zero values. Considering a lot of the sales are made at relatively low prices, it could be assumed that customers do not place such great importance on ratings for these items. Furthermore, even in the category of loyalty fraud where the prices are higher, feedbacks are predominantly neutral.

4.5 Sales Across Markets

To distinguish whether some products were more popular on certain markets or others, a comparison of the number of sales in the four categories across the examined markets is made. The sales over time for the various products can be observed in Figure 4.10. What is evident is that products from all the categories were offered on the majority of the markets, except for loyalty programs, which, as already mentioned, appeared on the markets at a later time than the other three. Loyalty programs were primarily traded on Alphabay, with a negligible amount of sales made on Agora and Evolution.

The offering in the three categories of accounts, pirated software and vouchers was significant on Alphabay, Silk Road 2 and Evolution, while there was a minimal amount of sales made on Black Market Reloaded, and especially on Pandora.

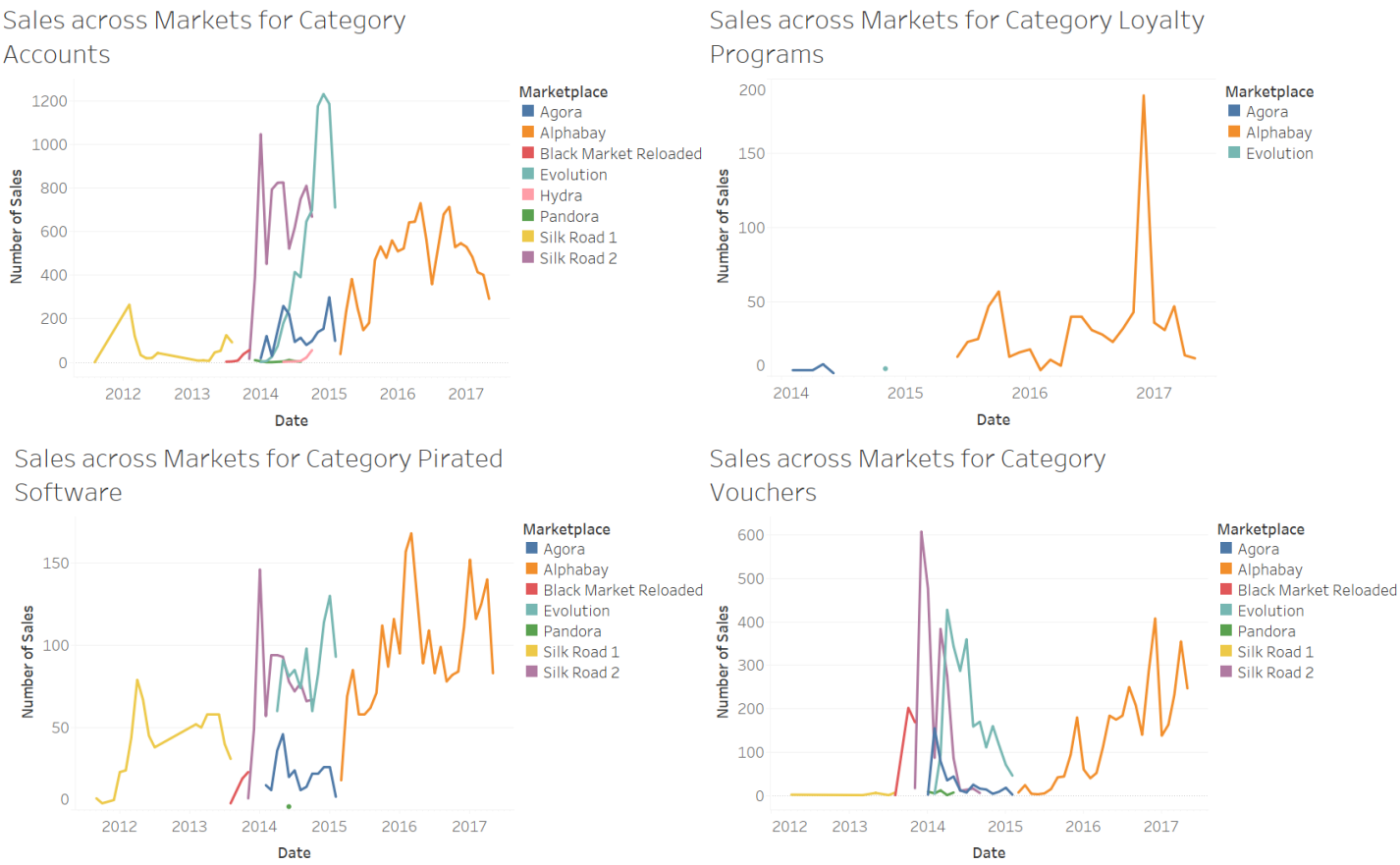


Figure 4.10 Sales across Markets per Category. Top left: Category Account Fraud; Top right: Category Loyalty Programs. Bottom left: Category Pirated Software; Bottom right: Category Voucher Fraud.

Sales in the accounts category grossly outnumber transactions in the other three categories. Accounts were also the only ones offered on all eight markets, including Hydra. The offering had its peak in 2014-2015 with the greater part of sales being made on Silk Road 2 and Evolution. After the closure of these markets, and the establishment of Alphabay, there was a considerable amount of accounts traded there, although the volumes do not reach those achieved previously.

The retail of vouchers follows a similar pattern to that of accounts, though on a smaller scale: while the highest amount of accounts sold on one market was over 1200 on Evolution, the most vouchers that were traded on a single market was slightly over 600 on Silk Road 2.

The retail of pirated software was consistent over time and markets, with Alphabay reaching the highest number of sales made on a single market, though the number is still slightly behind the highest point of sales made cumulatively in 2014-2015.

Overall, it can be seen that items from all categories, except loyalty programs, were consistently offered across the majority of the markets and over the entirety of the examined time period.

4.6 Sales Across Vendors

It is interesting to explore the level of vendor competition in the different categories, so as to determine whether there is a variety of sellers supplying the products, or just a few big vendors. Figure 4.10 displays the distribution of sales among vendors: each plot shows the portion of sales made by a vendor as percentage of all per category. It is clear from the charts that there are a few

dominant vendors in each category which have made the greater part of the sales. This is especially true for the accounts category, in which one of the 183 vendors has been responsible for over half of the sales. The rest is more or less divided between numerous other sellers, with only three or four managing to reach a 5 – 10% market share.

The loyalty programs category sees even less differentiation, though this could also be attributed to the overall much lower number of sellers in this category (18). Nearly 85% of the sales are generated by three vendors, with one making close to 50% of the sales. There have been significantly more vendors active in the pirated software category - 119, though similarly half the market is held by two vendors, who have almost equal market shares.

The one category which has more diversity in terms of vendors, is vouchers. Nevertheless, the stronger presence of three – four vendors among the 117 sellers is still evident, though they have reached around 17% market share at most.

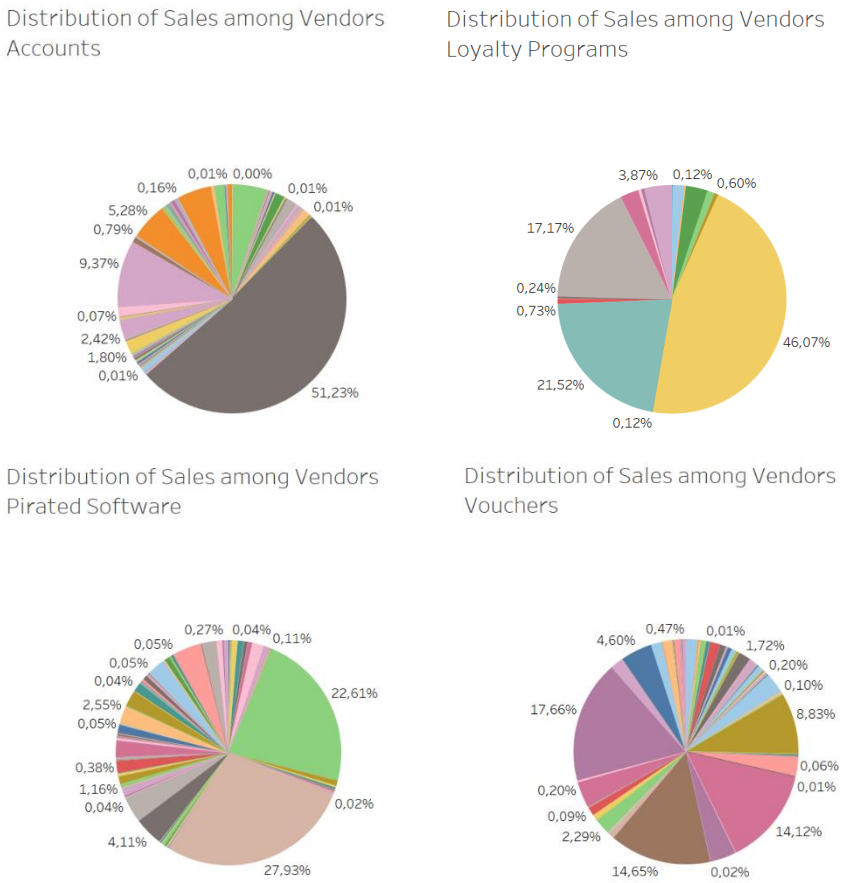


Figure 4.11 Market Share per Vendor, per Category. Top left: Category Account Fraud; Top right: Category Loyalty Programs. Bottom left: Category Pirated Software; Bottom right: Category Voucher Fraud.

This points that a small portion of vendors are responsible for a considerable number of the sales in each category, and consequently - the revenues. Furthermore, the majority of the sellers in each category apparently earn a negligible revenue, hence the earnings from the sale of fraudulent products is probably not the reason they participate in the underground trade.

4.7 Company Losses

4.7.1 Official Price Estimation

The process for determining the official selling prices of the various products followed different routes for the categories:

- For service accounts and subscriptions such as A1, or A2, as well as software subscriptions, the official website usually supplied several options ranging from subscribing for various time periods (monthly, quarterly-, semi- or annually), to signing up for basic, or premium services. In order to avoid overestimations, the price for the most basic service offered was selected, at the most favourable terms, which is usually the annual subscription. Considering most items offered on the underground markets are listed as lifetime accounts, the official prices for the service were annualized, not accounting for their value in perpetuity.

- In the case of loyalty programs, the offers on the dark web usually listed the amount of points accumulated in the accounts. Thereby, the website of the respective company was examined so as to determine the monetary equivalent of a point, and subsequently calculate the actual value of the account on offer. When this information was not made officially available, or the access to it was restricted to only members of the service, secondary sources were consulted, such as other websites providing relevant information; and FAQ (Frequently asked questions) sections.

- The entries in the database in the vouchers category generally have a value for the discount they offer stated in the description, which was the one used for the analysis. In the case the discount was in the form of a free item, the respective website was consulted so as to estimate the worth of the offered item, again selecting the most affordable option.

4.7.2 Costs

The estimated prices the items on the list officially retail for are detailed in Table 4.2, along with the projected financial costs incurred by the targeted companies from the sale of each item, and in total. The cost is calculated by comparing the retail price of each item and the median price it was sold for on the dark web.

Considering the accounts category, there are a few insignificantly low values: for accounts from A6, A20.1, A12, A21.1. However, these are mostly for services which are in fact offered for free, hence they are possibly bought for any accumulated amounts or payment information which could be linked to the account.

Apart from these, the costs are considerably higher across all categories, on occasions reaching a few hundred dollars. The greatest differences are noted within the loyalty program and the retail stores accounts. The reason for this could be in the way those accounts were valued: according to the listed amount within the account, which significantly raises the calculated losses.

There are a few companies in the vouchers category, which did not have full pricing information for the offered coupons, such as V8, V10, V13. Since the information could not be acquired from a secondary source, the final selling price was estimated from the available data. Therefore, the values for these companies may not be fully reliable and the projected losses were assumed to be zero.

Table 4.2. Estimated costs per Item, and Total, per Company

Company	Median Underground Price ($)	Lowest Official Price ($)	Estimated Cost Per Item ($)	Total Losses ($)	Company	Median Underground Price ($)	Lowest Official Price ($)	Estimated Cost Per Item ($)	Total Losses ($)
A1	2.99	96	93.01	731616.7	V1	50.4	163.09	112.69	425404.8
A2	3.94	120	116.06	845265	V2	12.5	38.14	25.64	23717
A3	2	72	70	191730	V3	30	97.79	67.79	47656.37
A4	4.99	360	355.01	891075.1	V4	17.5	54.15	36.65	18508.25
A5	7.89	420	412.11	815153.6	V5	9.9	25.51	15.61	6759.13
A6	1	0	0	0	V6	2.3	7.56	5.26	1940.94
A7	8.99	180	171.01	227272.3	V7	35	182.09	147.09	50304.78
A8	9.99	31	21.01	19581.32	V8	80	73.34	0	0
A9	3	84	81	68526	V9	5	29.14	24.14	5552.2
A10	8.99	184	175.01	129682.4	A21.2	285	528.58	243.58	50177.48
A11	3	300	297	215325	V10	7.7	46.99	39.29	7190.07
A12	4.99	10	5.01	1603.2	A20.2	25	177.93	152.93	15598.86
A13	1.05	120	118.95	28429.05	V11	12.8	2126.25	2113.45	169076
A14	4.5	240	235.5	53223	V12	17.5	24.89	7.39	546.86
A15	6.99	780	773.01	154602	V13	67	603.17	536.17	29489.35
A16	1.8	27	25.2	4788					
A17	4.96	60	55.04	8366.08		Subscription based, annual payment based on monthly rates			
A18	5	0	0	0					
A19	12.19	2661.63	2649.44	108627		Subscription based, annual payment			
A20.1	4.17	0	0	0					
A21.1	5.17	0	0	0		Based on average gift card value, account value unknown			
A22	10	2183.24	2173.24	43464.8					
L1	2.5	1433.89	1431.39	334945.3		Based on available balance in accounts			
L2	35	220.6	185.6	37305.6					
L3	20	654.56	634.56	93914.88		Based on price per point estimation			
L4	15	1676.25	1661.25	164463.8					
L5	10	911.11	901.11	49561.05					
L6	74.91	4625	4550.09	150153					
L7	90	5467.17	5377.17	161315.1					
S1.1	5	32	27	64638					
S2	4.99	240	235.01	486705.7					
S1.2	4.94	250	245.06	430080.3					
S3	5.02	180	174.98	72266.74					
S4	4.88	200	195.12	47219.04					
S5.1	2.2	80	77.8	12914.8					
S5.2	2.06	25	22.94	1835.2					

4.8 Concluding Remarks

In this chapter, the data from nearly 30,000 listings and over 300,000 transactions across eight underground marketplaces was analysed, with the purpose of arriving at some insights on the retail trade of fraudulent digital items and answering the third research sub-question.

The main types of service fraud categorized in the previous chapter: accounts, loyalty programs, pirated software, and vouchers were examined. It was determined that the most profitable category was that of vouchers, while accounts make up the greatest part of all transactions which have taken place. Listings remain on the markets typically for about a month, though some were observed for longer periods of more than a year.

In general, most products retail for relatively low prices, which are consistent across all listings. It is the higher transaction volumes of some products that put them ahead of the rest. The one category which has a fairly different price distribution, is that of loyalty programs, which had a higher variability in selling prices. That is also the category with estimated highest caused losses, due to the large amounts of points stored in the stolen accounts. However, the availability of such accounts is also limited.

It was established that items from all categories, except loyalty programs, were consistently offered across the majority of the markets and over the entirety of the examined time period. Furthermore, in each category there are a few sellers which hold the most of the market share, and consequently make the majority of the profits from this trade.

In conclusion, despite the overall low amount of money made by criminals on the underground markets from service fraud, when the losses for the respective businesses are taken into consideration, the costs become more significant. Furthermore, these amounts only account for the direct financial losses but not for any reputational damage or loss of customer trust, which is often described as more costly, especially in the long run.

5 Data Collection and Preparation

The previous chapters outlined the metrics to be used in the explanatory model and the main targets to be analysed. This section details the collection of the required additional data, as well as the following preparation procedures: its transformation and the replacement of missing data.

5.1 Overview of Additional Datasets

Company characteristics datasets: these datasets include the gathered additional data in order to define each selected variable, and comprise the following information:

1. Company size: as there is no open-source dataset providing financial information for all companies and all indicators, the selected variables defining this metric were manually collected from officially published annual reports and financial statements for the examined period from the companies' corporate websites or government institutions. In the case such were not public, third party sources such as marketing databases were consulted.
   Nevertheless, data on some indicators is still lacking, since not all gathered official information was complete. Another limitation in the compiled dataset is introduced by the inconsistent terminology used across different countries. For instance, reporting in the USA follows one structure, while companies operating under other jurisdictions follow different guidelines, which can lead to discrepancies in the use of financial terms. These differences in terminology were examined manually to arrive at a more uniform interpretation, although it is a time-consuming process which is bound to introduce some errors.
   Variables: Revenues, Net Income, Total Assets, Total Equity, Number of Employees.

2. Domain popularity ranking: this information serves to define the Visibility aspect, as it reflects the online presence of a company. It is sourced from Cisco Umbrella Popularity List, which consists of the top 1 million most popular internet domains, measured by the number of unique client IPs visiting a domain relative to the sum of all domain requests. As it is continuously kept up to date, and historical data was not available, the list from March, 2020 is used in the analysis.
   Variable: Popularity.

3. Reputation: the variable is based on the Global Top 500 Companies list (brandirectory, n.d), which is a popular measure to evaluate companies' corporate reputation. As the differences in ranking over the selected years were minor, the list from year 2017 is selected for this thesis. However, some reserve can be expressed to the reliability of this source as it is a commercially compiled ranking, even though it is from a world-renowned organization.
   Variable: Reputation.

4. Area of Service: this information specifies whether the examined companies provide their services on a global or a local level: for instance, only in a specific country or a geographic area. This information was gathered from different data sources: preferably the company's official website, or when unavailable: a third source such as marketing databases, and is traced back to the year 2017.
   Variable: Area of Service.

Limitations of the collected dataset

Considering the majority of the data is manually retrieved from a wide variety of sources, the nature and intensity of the collection process might have led to an increase in human errors. Furthermore, the limited availability of some information concerning the time period leads to some gaps in the data. In addition, some of the used sourced are commercial rankings, and may not be highly reliable.

Summary of utilized datasets

This chapter detailed the collection process for the data to be used in the second part of the analysis. Four datasets were created in addition to the main target dataset and the company costs dataset: the company size data; domain popularity ranking; reputation data; and area of service. They are summarized in Table 5.1, along with a commentary on the extracted variables for each.

Table 5.1 Summary of Utilized Datasets

Dataset	Source	Extracted Data
Main Underground Dataset	Original dataset	Target name, fraud category, number of fraud incidents
Company Costs Dataset	Manual observation	Losses
Company Size Dataset	Manual observation	Revenues, net income, total assets, total equity, number of employees
Popularity Ranking	Cisco-Umbrella	Domain popularity
Reputation Dataset	Global 500	Reputation
Area of Service	Manual observation	Area of service of each company

5.2 Data Transformation

Since the various collected financial indicators are measured in different units and have diverse ranges, it is advisable to normalize them before any further analysis. This would ensure that none of the measures have a higher weight than the others, thus disproportionately influencing the analysis. Therefore, the company size metrics are scaled in a range from zero to one.

5.3 Data Replacement

It was not possible to obtain all required information for the company size indicators. This is especially the case for companies registered as private firms, or which have been at the time of the examined period. As such entities are usually not required to publish official financial data, such information was mostly unavailable, outside some paid commercial services. The amount of missing data in relation to all is illustrated in Figure 5.1. Out of all 47 entries, eleven have missing values in one or more of the three company size variables: net income, total assets, and total equity, amounting to 4.255% missing values overall.

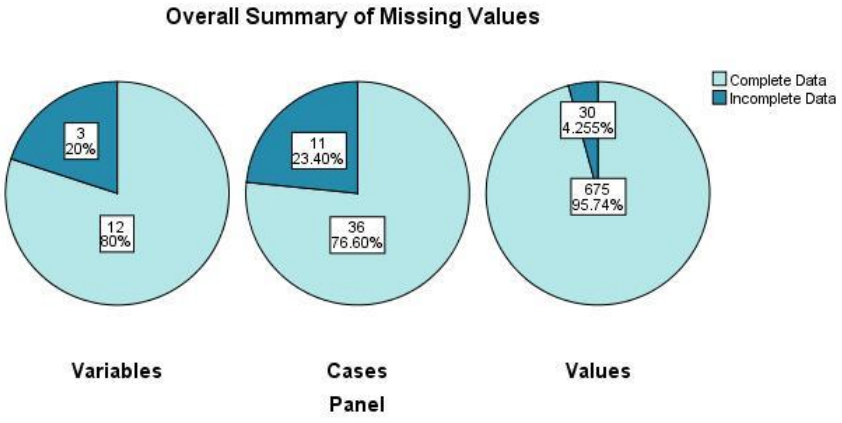


Figure 5.1 Summary of Missing Values

As the amount of missing information is less than five percent from all available data, trimming incomplete entries would generally be an option. However, considering the small size of the sample, it is preferred that as many cases as possible are retained. Therefore, the missing data is filled in using a multiple imputation technique.

The missing values should be explored, so as to determine which imputation method is most appropriate. After examining the data, it is determined that the pattern of the missing values is monotone, as data is missing in a systematic manner, which can be seen in Figure 5.2. A monotone pattern indicates that when a variable has a non-missing value, all preceding variables also have non-missing values (IBM, n.d.). All cases which have missing values in total equity, also have missing values in total assets, which in turn lack values in net income.

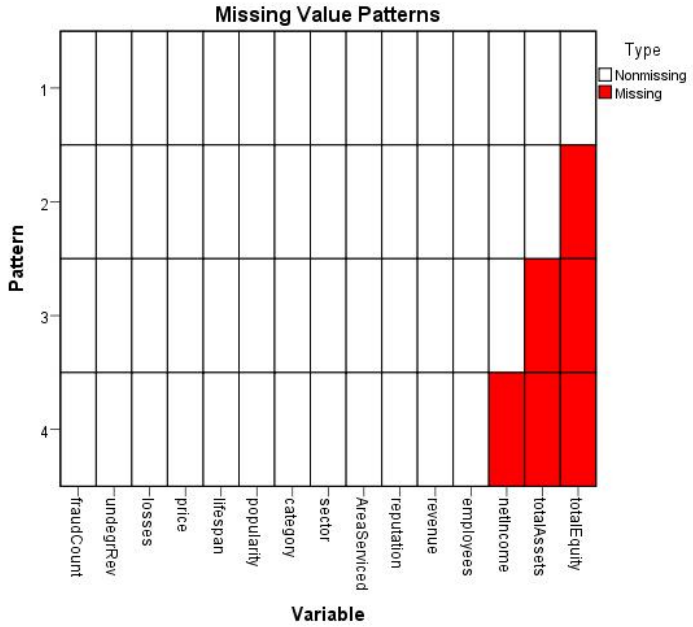


Figure 5.2 Missing Value Patterns

Thus, the data is transformed using a monotone imputation method. This technique uses all preceding variables in the model as predictors to impute the missing values for the fitted dependent variable through a linear regression model (IBM, n.d). The imputation procedure is completed using the Multiple imputation functionality in SPSS.

Finally, after transforming and processing the required data from the collected datasets, it is merged based on the previously assigned identification number, hence forming one new dataset.

Summary of Chapter

The collection and preparation process of the company data required for the explanatory analysis was described in this chapter. Four additional datasets were created: one for the company size metric, comprising data on Revenues, Net income, Total assets, Total equity, and Number of Employees; and one dataset for the remaining three metrics: Reputation, Domain Popularity, and Area of Service.

As data was unavailable on some of the company size indicators, a data imputation technique was employed to replace the missing values. This concerns 30 values in three of the variables, representing a little over 4% of all collected data. The values are missing in a monotone pattern; therefore, a suitable for this purpose imputation method was applied.

The final step before performing the explanatory analysis was merging the collected and processed data for each company.

6 Explanatory Analysis

The influence of the various collected characteristics on the target selection of companies is studied through an explanatory analysis, and more specifically by multiple regression. Two types of regression models are generated: a negative binomial regression and a linear regression.

Initially, the variables used in the models are described, followed by some additionally required data processing, and the analysis itself.

6.1 Variables for Regression Analysis

The dependent and independent variables to be used in the regression models are elaborated on in the following sections.

6.1.1 Dependent Variables

The dependent variable for the negative binomial regression is fraud count. It describes the number of fraud incidents suffered by a specific company, and is a count variable.

The linear regression model employs the continuous variable total losses, representing the losses an entity has suffered, as calculated in the previous chapter.

6.1.2 Independent Variables

The regression models would attempt to explain the relationship between target selection and the outlined characteristics. As such, the independent variables are based on the selected metrics:

- Company size: comprises the normalized continuous variables: Revenue, Net income, Total assets, Total equity, and Number of employees.
- Reputation: a Boolean variable, based on whether the company is featured on the Global top 500 list; has a value of zero when the company is on the list, and one if it is not.
- Popularity: a continuous variable, based on the domain popularity ranking in the top 1 million list. It is inverse ranked: the highest ranked entity receives the highest score; in the case a company domain is not featured in the ranking, a zero is assigned.
- Average underground price: a continuous variable reflecting the average price a company product was sold for on the underground markets, z-transformed.
- Area of service: Boolean variable, which receives a value of one when a company operates worldwide, and a zero when it is locally based.

Control variables

The following variables serve as control variables in the models, with the purpose to limit possible bias resulting from a certain category being too popular on the markets, or the time an item has spent on the market:

- Category: the type of fraud, represented by the four categories: account, loyalty program, pirated software, voucher. Dummy variables are created to model the different categories.

- Mean lifespan: continuous variable, reflecting the average time a company item has been for sale in the dark web markets.

6.2 Data Analysis

In order to determine the most appropriate regression model for the analysis, the dependent variables were examined. The fraud count variable is an observed count, following a negative binomial distribution as seen in the descriptive analysis, suggesting a Negative Binomial Regression model would be most suitable.

A count variable could also be modelled using a Poisson regression, however that model assumes that the data follows a Poisson distribution and therefore, the mean equals the variance. By observing the standard deviation of 1658.619 of the fraud count variable, it is established that the mean (949.47) is significantly smaller than the variance. This indicates that the data is over-dispersed and a Poisson model would be unsuitable.

It is expected that some of the independent variables could correlate with each other: specifically, the company size metrics, since they are representative of one company characteristic, as well as other variables, such as reputation or popularity. It is normal that a company which is more established would be larger and would have greater visibility. Hence, the first assumption to be checked is whether the independent variables show high correlation with each other, which could lead to inaccurate coefficient estimations and incorrect interpretations. For this purpose, all independent variables are examined through a correlation matrix, which is supplied in Appendix A. Since the explored data does not have a normal distribution of values, Spearman's rho coefficient is used to assess the relationship between the variables. All five company size metrics are moderately to highly correlated, except for net income and number of employees. Reputation seems to share a slight correlation with domain popularity and price, as well as with the company size metrics, as expected.

One way to reduce multicollinearity is by carrying out a feature selection analysis. It is performed by examining the Variance Inflation Factor (VIF) values of the independent variables and filtering those which have VIF values higher than 10 one by one, until the multicollinearity is reduced. However, the majority of the company size variables had similar VIF values, and this method did not provide satisfactory enough results. The collinearity diagnostic with VIF factors for each variable is displayed in Appendix B.

Another option for decreasing multicollinearity is by reducing the five company size measures into a fewer number of dimensions through a Principal Component Analysis (PCA), which is the technique applied in this study. The analysis is carried out in SPSS.

6.2.1 Principal Component Analysis

PCA transforms the initial correlated variables into principal components: uncorrelated linear combinations of the original set of variables, weighted by the portion of the variance they explain in the dataset (Abdi & Williams, 2010). The first factor explains the highest portion of the variance, and each following factor represents the most of the remaining variance. The total amount of variance captured by a given principal component is represented by eigenvalues.

The Principal Component Matrix is presented in Table 6.1. It is evident that the first two components explain the majority of the variance, with the first representing about 75%, and the second – 20%, while the remaining three have negligible values. Considering the first two components explain such a high percentage of the variance, it can be assumed that there are no latent factors explaining the variance.

Table 6.1 Principal Component Matrix

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.770	75.407	75.407	3.770	75.407	75.407	2.856	57.125	57.125
2	1.033	20.657	96.064	1.033	20.657	96.064	1.947	38.939	96.064
3	.137	2.747	98.811						
4	.050	.991	99.802						
5	.010	.198	100.000						

Extraction Method: Principal Component Analysis.

There are several ways to select the optimal number of components from the PCA analysis. One method is by choosing those components, which have eigenvalues greater than one. Another criterion is to select a number of components which in total explain between 70% to 80% variance. In this case, the first two components cumulatively capture around 96% of the variance, and both have eigenvalues above one, making them an obvious choice. In order to confirm the selection, a scree graph plotting the eigenvalues and the components is examined (Figure 6.1). The point where the graph makes the largest drop, or "elbow" joint, marks the moment when little variance is explained by the next components. Therefore, the scree plot also suggests the use of two components.

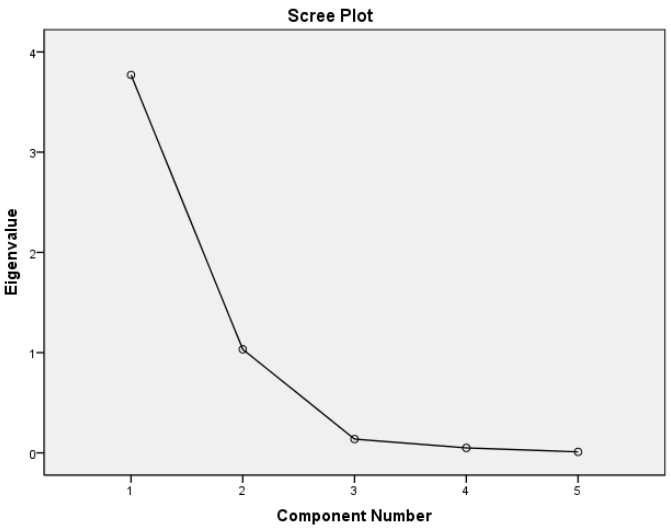


Figure 6.1 Scree plot PCA

In order to better fit the data and improve interpretability, a factor rotation can be applied. A rotation method which assumes that the factors are independent and not correlated with each other is orthogonal rotation. A commonly used type is varimax, which maximizes the variance of the loadings among the factors, while maximizing the differences between high and low loadings on a given factor, making it easier to link the rotated factors with the variables. The varimax rotation is applied in this case.

Table 6.2 Rotated Component Matrix (PCA)

The component loadings and explained variance after the rotation can be seen in the third section of Table 6.1. The rotated component matrix is displayed in Table 6.2. It can be concluded that the first three variables: net income, total equity, and total assets, load on the first component, while number of employees and average revenue per year are loading on the second. Consequently, these two rotated factors are going to be used in the regression analysis.

Rotated Component Matrix[a]

	Component	
	1	2
Net Income	.961	.148
Total Equity	.939	.309
Total Assets	.893	.341
Number of Employees	.139	.985
Average Revenue per Year	.483	.862

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

6.2.2 Assumptions for Linear Regression

The second model to be used is linear regression, which is going to take the dependent variable Total losses.

Before a linear regression can be carried out, it has to be assessed whether the data meets several assumptions: normality, linearity, homoscedasticity, and absence of multicollinearity. The presence of multicollinearity was reduced through the PCA procedure; thus, it can be concluded that this assumption is met. The following graphs provide insight whether the data adheres to the other assumptions.

The first assumption to be checked is that of normality of the error distribution. The residuals should be approximately normally distributed, which is examined through a normal Predicted Probability (P-P) plot. The assumption is met when the values are distributed along the diagonal normality line in the plot. Judging from Figure 6.2, it can be assumed that the data has some deviation from the diagonal line and is not normally distributed. However, after natural log transformation, the results appear to have improved.

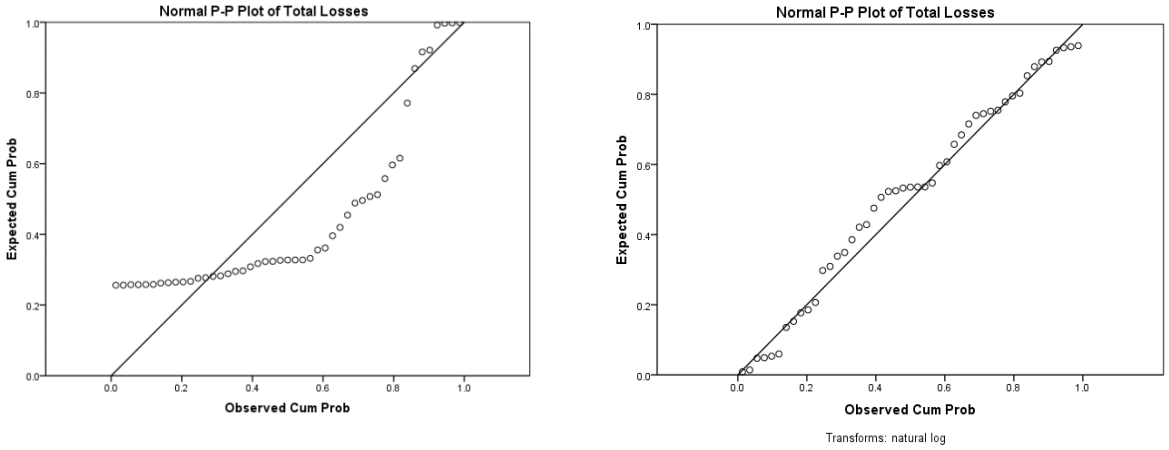


Figure 6.2 P-P Plot of Total losses. Left: not transformed. Right: After ln Transformation

The Residuals versus Predicted values scatter plots in Figure 6.3 can be used to check the assumptions of linearity and homoscedasticity. The data points should be symmetrically distributed around a horizontal line with an approximately constant variance, if the relationship between the dependent and independent variables is linear. Furthermore, the data should not show a particular pattern if it is homoscedastic. Homoscedasticity presupposes there is constant variance of the errors: the error term is similar across all values of the independent variables. The plot before transformation shows clear pattern. Apparently, the results have improved after the natural log transformation as seen on the right side of Figure 6.3.

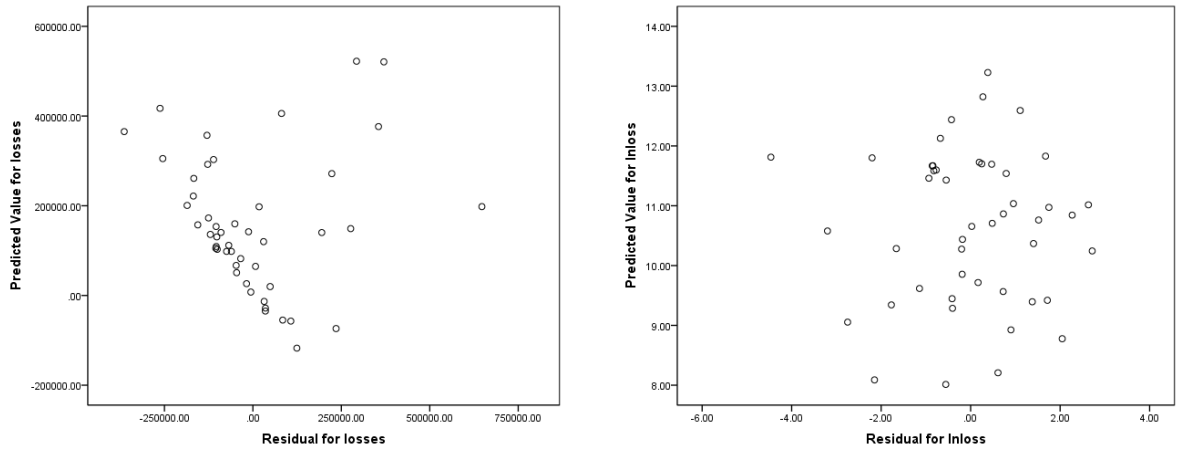


Figure 6.3 Residuals vs Predicted Values. Left: not transformed. Right: After ln Transformation

6.3 Regression Models and Results

The two created models are presented in the following section: the negative binomial regression, which tests the likelihood of a company to be targeted based on its characteristics; and the linear regression, used to examine how these characteristics influence losses incurred by the companies.

6.3.1 Negative Binomial Model

The performance of the model is going to be examined to determine whether the reported results are reliable enough. The first measure to be checked is the deviance. In table 6.3 are displayed several measures of the goodness of fit, the one that is going to be used to evaluate the model is the Pearson Chi-Square. Its Value/df measure should be more than 0.05 so as to determine that the model fits the data well. In this case the value is 1.966, which is well above the minimal threshold.

Table 6.3 Goodness of Fit Negative Binomial Regression

Goodness of Fit[b]

	Value	df	Value/df
Deviance	52.384	36	1.455
Scaled Deviance	52.384	36	
Pearson Chi-Square	70.784	36	1.966
Scaled Pearson Chi-Square	70.784	36	
Log Likelihood	-347.505		
Akaike's Information Criterion (AIC)	717.011		
Finite Sample Corrected AIC (AICC)	724.554		
Bayesian Information Criterion (BIC)	737.363		
Consistent AIC (CAIC)	748.363		

Dependent Variable: Fraud Count

Model: (Intercept), AreaServiced, reputation, popularity, sizeC1, sizeC2, lifespan, price, category

b. Information criteria are in small-is-better form.

Table 6.4 Omnibus Test Negative Binomial Regression

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
43.493	10	.000

Dependent Variable: Fraud Count

Model: (Intercept), AreaServiced, reputation, popularity, sizeC1, sizeC2, lifespan, price, category

a. Compares the fitted model against the intercept-only model.

The other measure to be examined is the Likelihood Ratio Chi-Square as displayed in Table 6.4, omnibus test. This measure tests how the model compares to the 'null' model, or a model without any independent variables. The reported value is 43.493, with the associated p-value of 0.000. As this value is below 0.05, it can be concluded that the model is a significant improvement over the model with only intercept.

The negative binomial model with all parameters is presented in Table 6.5. The models generated for each independent variable are included in Appendix C. The table contains the regression coefficients for each independent variable, with their standard errors, Wald Chi-Square values, and p-values.

Table 6.5 Results Negative Binomial Regression

			Hypothesis Test		
Parameter	B	Std. Error	Wald Chi-Square	df	Sig.
(Intercept)	6.125	.5946	106.132	1	.000
[AreaServiced=0]	.520	.4368	1.419	1	.234
[AreaServiced=1]	0[a]	.	.	.	.
[reputation=0]	1.107	.3877	8.149	1	.004
[reputation=1]	0[a]	.	.	.	.
popularity	-.034	.0141	5.710	1	.017
sizeC1	-.338	.2252	2.251	1	.134
sizeC2	-.089	.2234	.160	1	.690
lifespan	.004	.0049	.702	1	.402
price	-.338	.2722	1.542	1	.214
[category=1]	.527	.5850	.811	1	.368
[category=2]	-1.694	.6386	7.035	1	.008
[category=3]	.220	.8062	.075	1	.785
[category=4]	0[a]	.	.	.	.
(Scale)	1[b]				
(Negative binomial)	1				

Results Negative Binomial Regression

Company size

One of the factors being evaluated is that related to firm value, which in this case is represented by the two components of the company size metric. The results of the regression point that this characteristic is not significant enough to be able to explain why some firms are more targeted than others. The same result is observed in the models generated for each of the two size metrics. This is in line with the statement that, regarding the RAT framework, the aspect of Value does not seem to have much significance (Leukfeldt & Yar, 2016).

Reputation and Domain popularity

Reputation appears to have the strongest influence on the likelihood of being targeted. It has a positive coefficient in the model, which indicates that the more esteemed and well-known a company is, the higher the chances it may become a victim of service fraud. The other characteristic related to a company's visibility factor: domain popularity, also seems to have an influence on the amount of fraud suffered, even if it is not as pronounced. Companies seem to experience less fraud the more their domain popularity increases. When comparing with the separate regressions for each of the two variables, the results are similar, although popularity seems to lose some of its significance when the other input variables are included.

Location

Similarly to company size, the location of a firm does not seem to be able to explain whether it would be more targeted. Considering the majority of the items traded on the underground markets can be redeemed online, or are products offered in multiple areas, this result was to be expected.

Relative price

The price an item was traded for does not explain why a company may be victimized more often in neither the complete model with all variables, nor the model regarding only price as an independent variable. One possible explanation is that there is demand for all types of items: those which cost less like streaming services accounts, as well as the more expensive products, such as vouchers and loyalty programs with remaining balances.

6.3.2 Linear Regression Model

The linear regression model aims to uncover any potential relations between the selected company characteristics and the amount of total losses experienced by businesses.

The performance of the model is determined by the F-value, which is 2.603, with a p-value of 0.017, as can be seen in Table 6.6. This is lower than the significance level of 0.05, therefore, the independent variables reliably explain the differences in the dependent variable. The R-Square, which is used to estimate what part of the variance in the dependent variable can be justified by the model, is 0.420, indicating that 42% of the variance in total losses is explained by the company characteristics. However, the R-square value is sometimes considered not accurate enough measure as it can have higher values when more predictors are included in the model due to chance and not because they actually explain the variance. Which is why the adjusted R-square is also examined, which in this model is 0.258. Due to the way it is calculated, when there is a small number of observations, but a large number of input variables, the difference between the two values is higher. Despite that the value of the adjusted R-squared is lower, it is still high enough to explain some of the variance.

Table 6.6 Parameters Linear Regression

Tests of Between-Subjects Effects

Dependent Variable:lnloss

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	75.270[a]	10	7.527	2.603	.017
Intercept	430.447	1	430.447	148.841	.000
category	19.023	3	6.341	2.193	.106
AreaServiced	22.706	1	22.706	7.851	.008
reputation	5.134	1	5.134	1.775	.191
lifespan	5.615	1	5.615	1.942	.172
price	.718	1	.718	.248	.621
popularity	6.824	1	6.824	2.360	.133
sizeC1	14.158	1	14.158	4.896	.033
sizeC2	1.350	1	1.350	.467	.499
Error	104.112	36	2.892		
Total	5508.564	47			
Corrected Total	179.381	46			

a. R Squared = .420 (Adjusted R Squared = .258)

The results of the complete linear regression are displayed in Table 6.7. The regressions for each variable are in Appendix D. The table shows the coefficients for each variable; the standard errors associated with each coefficient; and t-test and significance, used to test the null hypothesis that the coefficient is equal to zero.

Table 6.7 Results Linear Regression

Parameter	B	Std. Error	t	Sig.
Intercept	8.780	.937	9.367	.000
[category=1]	1.255	.775	1.618	.114
[category=2]	2.343	1.039	2.254	.030
[category=3]	2.073	1.057	1.960	.058
[category=4]	0[a]	.	.	.
[AreaServiced=0]	1.982	.707	2.802	.008
[AreaServiced=1]	0[a]	.	.	.
[reputation=0]	.860	.645	1.332	.191
[reputation=1]	0[a]	.	.	.
lifespan	.009	.006	1.393	.172
price	.189	.379	.498	.621
popularity	-.043	.028	-1.536	.133
sizeC1	-.663	.300	-2.213	.033
sizeC2	-.191	.280	-.683	.499

Results Linear Regression

Company size

Upon examining the values for the two components of the size metric, it is established that the first component has a negative influence on the losses, while the second component is not significant enough to explain any variance in company losses. Hence, when the first component, which represented mainly the variables net income, total equity and total assets, has lower values, companies suffer higher losses. It is possible that the larger companies have better security controls in place, or that the losses they incur from this fraud are negligible compared to their financial position.

Reputation and Domain popularity

Unlike with target frequency, both reputation and domain popularity do not seem to have an effect on the total losses. A possible interpretation is that while companies which are more visible are targeted more often, the losses suffered from these attacks are not that significant, which resonates with the results obtained for the company size variable.

Location

The location of a company appears to have a relation with losses. The interpretation of the results from Table 6.7 is that when a company operates locally, it has a higher chance to suffer greater losses from service fraud. As the definition for local relates to large geographical areas comprising whole continents, and a big portion of the examined companies are based in the USA, this result could indicate that companies located there are more affected.

Relative price

The relative price of items does not have significant enough influence on total losses, similarly to the effect on target frequency.

Reflection on Models

It should be noted that the presented models can serve as an indication of the issue but are in no way conclusive, as they are bound by several limitations. There are various other factors which could be examined, but due to time constraints and data limitations, this has not been done in the current study. These factors could potentially have a greater influence on target selection than those explored in the regression models. Furthermore, a drawback of the analysis is the small number of examined cases, which could limit the significance and generalizability of the models. Finally, the data was largely collected and processed manually, which would inevitably produce some inconsistency.

Summary of Chapter

The explanatory analysis performed with the variables collected in the previous chapters was described in this section. Two regression models were generated: a negative binomial regression examining the frequency of fraud incidents; and linear regression, exploring total losses. As independent variables in both models were used: company size, reputation, popularity, average underground price, and area of service, while category and mean lifespan served as control variables.

It was established that the company size variables correlated with each other, thus a principal component analysis was performed. This resulted in the reduction of the five correlated factors into two new components. The assumptions for linear regression were examined, leading to the natural log transformation of the total losses variable.

The results of the first regression showed that reputation and domain popularity managed to explain some of the difference in the number of fraud incidents. The second model indicated that a larger company size, as well as a wider area of service, have a negative effect on company losses.

7 Discussion

This chapter discusses the answers to the research questions set in the beginning of this thesis, by reviewing the conclusions reached in the previous chapters. An outline is given of the limitations concerning the used data and the employed research method, along with five suggestions for possible future research.

7.1 Conclusion

This section is going to summarize the various theoretical and analytical work completed in previous chapters and provide answers to the examined research questions by reviewing the outcomes of the analyses.

The main research question this thesis aims to answer is:

What company characteristics influence the likelihood and financial impact of service fraud conducted on the underground markets in the dark web?

Clarity on the main question will be achieved by providing answers to the four sub-questions.

1. *What is the current underground digital fraud landscape and how does service fraud fit into it?*

The first question served to provide the basis for the rest of the research. Taking into consideration the novelty of the topic, a detailed review of the relevant concepts and themes was presented so as to create a fuller picture of the issue:

- Main concepts were introduced: the notions of cyberspace and cybercrime; an outline of the prominent theories in cyber security management; models regarding cyber security investment and estimating the cost of cybercrime.

- The theoretical foundation of this study was represented by briefly exploring relevant theories from criminology and economics, and outlining the application of the Routine Activity Theory to the examined case.

- The role of the underground markets in the digital fraud landscape was detailed from the first attempts at illegal online trade, to the specifics of the current markets and the main factors influencing market dynamics.

- The notion of service fraud is introduced and elaborated on, including its high impact on businesses and organizations; the prevalent attack vectors through which criminals acquire stolen credentials: credential spills, malware infections, and phishing campaigns.

- A criminal business model for service fraud is suggested along with the five parts of its value chain: collection of stolen credentials; development of malicious software; distribution of an attack; taking over of compromised accounts; and cashing-out.

- Main types of service fraud as encountered in the literature are described: fraud affecting online retailers comprising stolen accounts and vouchers; and fraud with hotel and airline loyalty award programs.

2. *What company characteristics can be identified as potentially able to explain differences in security incidents*?

Various factors which could contribute to the appeal of a company, and consequently the number of security incidents suffered by it, were examined from previous studies focusing on target selection. They are sorted according to the three target suitability aspects of RAT:

- Value: company size, as measured by: profits, return on assets, asset intangibility, net income, revenues, total assets, total equity, market capitalization, number of employees, number of customers; digitization of value; industry; future growth opportunities.

- Visibility: popularity; reputation; symbolic significance and criticalness; location.

- Accessibility: IT security budget; weakness of defense mechanisms; presence of cyber risk management committee.

A selection of characteristics to be used in the explanatory analysis was made based on their prominence and the feasibility of collecting the required data. The selected measurements were: company size, based on the five indicators: revenue, net income, total assets, total equity, number of employees; reputation; domain popularity and location of a company.

3. *How can service fraud be classified and what are the main targets affected by it?*

To provide an answer to the third question, a descriptive analysis was performed. Prior to that, the explored dataset had to be manually analyzed, so as to extract the meaningful listings from the irrelevant and inconsistent data, which was quite a time-consuming process given the size of the dataset. In the end the 47 most affected companies were selected for further analysis. These represented approximately 60% to 90% from the original entities, depending on the category.

Four main categories were introduced based on common characteristics of the included items: *accounts, loyalty programs, pirated software,* and *vouchers*. The descriptive analysis revealed interesting findings, among which the price distribution, revenues, average prices in the four categories, pointing that loyalty programs were the priciest item, while accounts were the most popular. The accumulated revenue and sales of accounts and vouchers per company over the whole period was examined in order to present the evolution of the fraud on the underground markets.

Supplementary data was collected for the price of every sold item, so as to estimate the average total losses each affected company had incurred. This value along with the number of fraud incidents per company were used as dependent variables in the explanatory analysis. Furthermore, relative price per listing was used as an independent variable.

The additionally required data for the characteristics of the selected companies to be examined was collected almost entirely through manual observations from various sources, based on the

selected metrics from the previous section and the chosen companies from the descriptive analysis.

4. *Which of the collected company characteristics are able to explain the differences in the number of security incidents and the amount of incurred losses by service fraud?*

Regression analysis was conducted with the aim of establishing whether the collected variables influence the two selected dependent variables: incident count and total losses. A negative binomial regression and a general linear regression were used to model the relationships between the dependent variables and the extracted independent variables: company size, reputation, domain popularity, area serviced, relative price. Mean lifespan and category were used as control variables.

Extensive procedures focusing on preparing and processing the data for the analysis were required. As the company size characteristics lacked a few data points, multiple imputation was performed on the missing values. When testing the model, multicollinearity was observed between the company size metrics. In order to deal with this, a Principle Component Analysis was done, reducing the five correlated metrics into two uncorrelated components representing the initial variables.

Overall, few of the selected characteristics were significant enough to be able to explain whether a company would be more or less targeted. These were the visibility variables: reputation and domain popularity. Reputation can be linked to a higher number of fraud incidents, while the probability to be more frequently targeted reduces with popularity. The two variables do not seem to be able to explain whether a company would sustain more financial losses.

The two components of the size metric were not significant enough in the negative binomial model, indicating target selection does not relate to company size. The first component, however, was found to have a negative effect on total losses, indicating smaller companies may suffer more from service fraud financially.

Whether a company offers its products and services on a global level or on a smaller scale does not seem to affect the frequency of fraud incidents. However, operating locally seems to relate to having higher losses according to the results of the linear regression. The relative price of items on the underground markets does not seem to influence neither target frequency, nor sustained losses.

A possible explanation why the visibility related characteristics have a higher influence on the number of experienced security incidents than the value factors, could be that visibility can be considered more of an external feature of firms. When criminals decide which companies to target, they would aim for these which could bring them more value and often businesses with better branding and more prevalent image are also seen as highly successful, although this may not be so. A case in point are some of the companies included in this study: one of them which is a well-established media services provider, in fact has not generated any profits in recent years, but has only reported net losses. Three other equally popular companies have negative equity, indicating they have sustained losses for multiple periods, and have not been able to break even. This is especially true for relatively young and expanding companies, offering a popular product, but which still do not have extensive capital. As cybercriminals would prefer to offer items which have a higher demand on the markets, it is likely they would aim for the companies supplying such products. Without acquiring deeper knowledge about a company, fraudsters would be inclined to make decisions based on their perceived rationality.

7.2 Limitations

Limitations of the study have been acknowledged in previous chapters in the data collection and analysis phases. They are summarized in this section along with an explanation of possible implications for the results.

Data quality

- The majority of the data used in the analysis was either collected or processed manually. This was undoubtedly the most time-consuming and challenging part of this study, especially filtering through and categorizing the initial database. The collection of the company size characteristics posed some difficulty as well, as numerous documents had to be manually scanned for the appropriate measures. Consequently, human errors could have been introduced, which is difficult to trace. This potentially reduces the completeness and accuracy of the data, which in turn influences the quality of the analysis results.

- Data on some of the company size characteristics was not publicly available, due to different reporting regulations for privately-held companies. The missing values were acquired from third party sources, such as marketing agencies reports, or estimated through imputation techniques, which might be influencing the model output.

- The measurement used for the reputation metric might not be representative enough, as it is solely based on a commercial ranking, which is hardly a reliable source. Nevertheless, the metric has been suggested in other research and serves to provide an initial idea about this characteristic. The use of a more comprehensive metric consisting of a wider variety of measures would certainly yield a more dependable result.

- The examined dataset comprises data from the period 2011-2017, which can be considered sufficient enough for the purpose of this research to provide an initial view of service fraud. Yet, developments in the technical and cybercriminal world are occurring at such a fast pace, that it would not be appropriate to draw overly general conclusions about the issue.

Completeness of model

- The variables included in the model are related to the value and visibility aspects of RAT's target suitability model, while accessibility, represented by: employed security controls, assigned security budget and other measures taken to counter online fraud, is not evaluated. This is due to the difficulty of obtaining such information, often regarded as confidential. Nevertheless, it could have a considerable influence on the amount of service fraud suffered by the various companies. According to (Leukfeldt & Yar, 2016), the focus of criminals is moving away from the value and visibility characteristics of a target towards its accessibility. Thus, making assumptions without including security characteristics in the model would not be conclusive enough.

- The entities included in the descriptive and explanatory analyses were selected on the basis of their popularity on the markets: listings of products which had been sold less were excluded as they did not provide enough data points for the analysis. This could introduce bias in the results, considering there are a lot of companies which have less presence on the markets, but nevertheless are targeted.

Furthermore, this study examines service fraud transpiring on the underground markets in the dark web. Taking into account the unpredictability and volatility of these markets, there could be various external factors explaining the differences in observed fraud numbers. For instance, some of the affected companies were only present in the dataset for a few years and not the whole period. This could have been caused by a one-time data breach as witnessed with one of the hospitality companies, featured in the analysis, leading to the circulation of stolen credentials for a limited period of time. However, the drop in offers could also be attributed to the seizure or disappearance of a particular market where the trade was active, and not because there was no interest in the items anymore.

In addition, the examined dataset consists entirely of underground markets operating in the English language; thus, any trade happening in more local or country-specific identical marketplaces is not considered in the study. This could be a reason why almost all affected companies in the analysis are major brands either globally or in the English-speaking world, which could potentially lead to a subjective view on the matter.

7.3 Future Research

Given the results of the performed analyses and the limitations defined in the previous section, some possibilities for future research are going to be summarized in this section.

- The current research is based entirely on quantitative methods, which despite having their own merits, cannot capture the full intricacy of the examined research topic. The initial intention of this thesis was to discover what influence service fraud has on affected companies and their business models through the use of expert interviews. Regrettably, due to lack of sufficient participation this idea had to be abandoned. However, this topic remains open to exploration and other researchers may have more success in acquiring the necessary intelligence to estimate the business effects of such fraud. Collaboration with companies offering security services could potentially extend the reach of the research and provide further insights into the global trend.

- The descriptive analysis produced various metrics, nonetheless, due to time constraints a few of them were selected for the explanatory models. A following research could look into the effects of the various unexplored factors, such as the role of the different underground markets, or the vendors in the target selection, as it was established that few vendors control the majority of the offerings. Since the focus of this thesis was on the targeted businesses, the market-specific factors influencing the fraud were not examined, providing ground for other studies to concentrate on this aspect.

- The work on this thesis could be extended by the addition of more targeted companies from the dataset. The refined dataset included a greater number of entities; however, as previously mentioned the final selection only included those most affected. Future study

could incorporate the others, and explore the differences between the two groups in terms of company characteristics, probability of being targeted, and factors influencing this likelihood.

- The inclusion of an evolutionary analysis of the data could enhance the results of the thesis. The explored dataset spans the years 2011 to 2017, while the collected company size metric was averaged to an annual measure, thus any developments over the period could not be traced. Collecting more data and performing an enhanced model could reveal trends in the underground trade with such items, as well as indicate the effect the existence of certain markets has on the online fraud ecosystem.

- There is reason to believe that cybercriminals are getting more and more reckless in their activities, and fraudulent items are increasingly being offered on the surface web, social media, and messaging apps groups. It would be interesting to examine whether this development has caused a shift in the offerings of stolen accounts and vouchers from the dark web to more easily accessible channels which do not employ the stringent security and authentication measures of the underground markets. Analyzing data from more recent years could show whether any apparent decrease in trade has taken place, and whether research efforts should be directed elsewhere.

Summary of Chapter

This chapter gave answers to the four research sub-questions, and consequently, the main research question: ***What company characteristics influence the likelihood and financial impact of service fraud conducted on the underground markets in the dark web?*** *The characteristics reputation and popularity were found to have an effect on the likelihood of being targeted by service fraud, while size and location have an influence on financial losses. It could be reasoned that cybercriminals are affected by their bounded rationality, therefore aim for companies with greater public image, expecting to realize higher profits.*

The rest of the chapter outlined the limitations of this research, regarding used data and completeness of the model. The quality of the results could have been affected by the accuracy of the data, possibly reduced by unavailability of information, and the manual collection and processing of the majority of the datasets. Taking into account the listed limitations and the time frame of the project, suggestions for future research are provided at the end of the chapter.

8 Relevance

This thesis contributes to research by paving the way in exploring an area of digital fraud which had not been previously addressed: service fraud. The few references to the topic found in the literature were all in reports from the security industry. The existing prior research into target selection focused on banking malware and its implications for financial institutions. As a consequence, there was no consistent data or general knowledge on this type of online fraud. The contributions of this work to science and society are explained in the following section, including practical recommendations for the various affected actors.

8.1 Scientific Contribution

The purpose of this study to explore a novel research topic and develop a model explaining target selection in online fraud has led to the following contributions:

- The data comprising the various transactions from the underground markets was available in a greatly fragmented and inconsistent manner, rendering it unusable for research. The manual processing of this information has resulted in the creation of a new dataset, which can be used by researchers in the field to garner new insights. Furthermore, the data has been classified into four categories based on type of fraud. This categorization could aid in executing more targeted studies in a group of interest.

- This thesis introduced various metrics to examine the relationship between company characteristics and likelihood of being a victim of service fraud. The collected data could be used in future studies employing different methods for analysis or exploring other aspects of the issue. This could ease the process of acquiring data and allow researchers to focus on expanding the topic beyond the current results.

- This project explored the effect of size, reputation, domain popularity, relative location as factors of the 'value' and 'visibility' elements of RAT. Thus, the results contribute to the discussion how applicable this theory is to crimes committed in the cyber domain, and how much influence its different elements have on target selection.

- This study enhances the still relatively under-researched area of target selection in cybercrime. The investigated victimization in digital fraud can complement the efforts of previous research which focused on various types of cyberattacks and malware, and serve to create a fuller picture of the target selection landscape.

8.2 Social Contribution

Society is gradually becoming more dependent on the Internet for various activities from online shopping to the use of entertainment and travel services, expanding the playfield for fraudsters willing to take advantage of every opportunity. In their strive to attract and retain more clients, businesses are trying to ease the customer experience and provide more services digitally, thus increasing the potential weak points in their defenses. Nevertheless, the responsibility for having a more secure society does not lie solely in the hands of companies, governments or organizations, every individual can do their part. But without having sufficient knowledge, one can hardly take the right measures.

This study aims to increase awareness on a previously unexplored issue which can affect people as well as companies. Individuals can become more conscious in their use of various services online and the credentials they are supplying, knowing there is an actual possibility this information can be misused. Organizations and security companies can gain an idea of the digital fraud landscape and the company characteristics which put businesses at higher risk. The results of the explanatory analysis, albeit limited, can point to which company aspects cybercriminals may be paying more attention to.

8.3 Recommendations

As service fraud is relatively novel compared to banking or credit card fraud, businesses which could potentially be affected have generally not implemented the high level of security measures as seen within the financial sector, while still handling products and payments. Not only does that make them an attractive target for cybercriminals but it leaves their customers at a higher chance of becoming a victim. The identification and categorization of the different types of service fraud presented in this thesis can aid companies in determining more easily which of their offered products could be exploited: gift cards, accounts, or both; and take the appropriate contingency measures.

The current process of creating an account in an online retailer or a service provider is tailored to deliver the greatest convenience to the user, thus rarely engaging customers in more complicated identity verification or authentication procedures. It is advisable that companies perform more thorough new user identifications by utilizing digital identity checks when available, or examining device data, such as geo-spatial information, biometric data, or user behaviour, in order to establish one's identity with greater accuracy. Moreover, businesses should employ more stringent authentication procedures. Presently, the login process at most affected companies constitutes a simple username and password combination. The addition of a multi-factor authentication has the potential to limit some of the fraudulent activities and provide another layer of security.

The statistical models indicated that more established companies are at a higher risk of being targeted, while the relatively smaller businesses may suffer more losses. The larger companies may have implemented more appropriate fraud detection and mitigation measures, or are better able to absorb the inflicted losses. This serves to show that cooperation should be encouraged within the affected industries, as loss of data from one company could easily damage another. The knowledge sharing process can be further supported by the security industry, as they have aggregated data available from various clients across numerous industries. Leveraging this information and experience could lead to the creation of more robust service fraud detection and prevention methods, as well as more strict consumer protection regulations, similar to those existing in the mature financial sector.

Moreover, the results of the model could support decision making and risk assessment evaluations. The majority of the cyber risk management frameworks outlined in Chapter 2 make use of expected losses and risk probability estimations to quantify the potential risks a company is exposed to. The company size metric, as well as the reputation and location characteristics, after further refinement, could be applied in such risk assessments, facilitating the cyber risk management process.

In addition to businesses, another actor which could benefit from this research are the law enforcement and government authorities investigating activities on the underground markets. The results of the model could help in identifying potential characteristics which

render an organisation more vulnerable, and aid in prioritizing potential targets. This could improve knowledge on target selection and lead to better understanding of the criminal landscape. In this regard, cooperation with research institutions would also prove beneficial to all involved parties.

Furthermore, this study has confirmed previously made observations in research focusing on the dark web markets, that the majority of the trade is held by a handful of vendors. As illustrated in the analysis, this is the case to a varying degree in all four examined categories of service fraud. Accordingly, law enforcement agencies could direct their efforts in tracing and apprehending the criminals responsible to the greatest extent for the trade, thus disturbing the criminal supply chain, and gaining insights into the main motivations behind committing this type of fraud.

Finally, since service fraud can affect directly customers both financially and psychologically, it is essential to try and raise awareness of the issue among society. Businesses can inform their customers as to why they should take better care when creating accounts, and refrain from reusing usernames and passwords. As this requirement can seem slightly abstract and unnecessary to some, it can be useful to illustrate the consequences of credential theft clearly, as is attempted in this study, so that users can put the issue in perspective.

Considering that the more prominent service provider companies, hotel and restaurant chains are among those most targeted, users should be better informed and not remain under the assumption that their accounts or credentials are well protected, simply because they are in the hands of a reputable company. Individuals should remain cautious with their personal data, and ensure they address any suspicious activity concerning their purchases, accounts, or finances.

8.4 Link to Master Program

At the core of the Management of Technology master program is the interrelation between business and technology, and how technological assets can be best used to support the growth and competitiveness of a company. Considering the wide-spread digitization of businesses, the protection of digital resources is becoming a prominent issue. Appropriate cyber risk management practices can increase the value of a company by providing sufficient protection to its assets, or improve its image and increase customer trust. This thesis has empirically investigated which characteristics of businesses make them an appealing target for cybercriminals engaging in fraud with stolen accounts, vouchers, loyalty programs and software. An increased understanding in this area can aid security managers in making decisions on which controls to implement based on the type of risks they are exposed to. Moreover, the insights from this study can be used to perform better risk assessments depending on the particular business environment a firm operates in. Finally, the work in this thesis involved an understanding of various areas such as cyber risk management, corporate finance, analytical and modelling tools, all of which are part of the master program.

References

Abell, P. (1992). Sociological Theory and Rational Choice Theory, 1–22.

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary*

reviews: computational statistics, *2*(4), 433–459.Accenture. (2017). *2017 Cost of Cyber Crime Study*.

Akerlof, G. A. (1978). The market for "lemons": Quality, uncertainty and the market mechanism. *Uncertainty in Economics*, 235–251.

Aldridge, J., & Decary-Hetu, D. (2014). Not an "eBay for drugs": the cryptomarket "Silk Road" as a paradigm shifting criminal innovation. https://doi.org/http://dx.doi.org/10.2139/ssrn.2436643

Anderson, R. (2001). Why Information Security is Hard – An Economic Perspective. In *Seventeenth Annual Computer Security Applications Conference, 10-14 Dec. 2001, New Orleans, LA, USA, USA*. IEEE. https://doi.org/10.1109/ACSAC.2001.991552

Armor. (2018). *The Black Market Report*.

Asghari, H., van Eeten, M., & Bauer, J. M. (2016). Economics of cybersecurity. In *Handbook on the Economics of the Internet* (pp. 262–287). Edward Elgar Publishing.

Barratt, M. J. (2012). Silk Road: Ebay for drugs. *Addiction*, *107*.

Becker, G. S. (1974). Crime and punishment: an economic approach. In *Essays in the economics of crime and punishment* (pp. 1–54).

Blueliv. (2018). *The Credential Theft Ecosystem*.

Bohme, R. (2010). Security Metrics and Security Investment Models. *Advances in Information and Computer Security. IWSEC 2010. Lecture Notes in Computer Science*, *vol 6434*. https://doi.org/https://doi.org/10.1007/978-3-642-16825-3_2

Bojanc, R., & Jerman, B. (2008). Towards a standard approach for quantifying an ICT security investment. *Computer Standards & Interfaces*, *30*, 219–222. https://doi.org/10.1016/j.csi.2007.10.013

Brains, C., Willnat, L., Manheim, J., & Rich, R. (2011). *Empirical Political Analysis* (8th ed.). Boston: Longman)

Brandirectory. (n.d.). Global 500 2017. Retrieved 2020-03-18, from https://brandirectory.com/rankings/global/2017/

Buxton, J., & Bingham, T. (2015). The Rise and Challenge of Dark Net Drug Markets. *Global Drugs Policy Observatory Policy Brief*, (January). https://doi.org/http://doi.org/2054-1910

Casey, E. (2004). *Digital Evidence and Computer Crime*.

Cheung, R. (2017). *Targeting financial organisations with DDOS: A multi-sides perspective* (Master thesis, Delft University of Techology).

Christin, N. (2013). Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 213–224).

Clarke, R. (1997). *Situational Crime Prevention: Successful Case Studies* (Second Edi). New York: Harrow and Heston.

Dang, C. D., & Li, F. (2015). Measuring Firm Size in Empirical Corporate Finance Measuring Firm Size in Empirical Corporate Finance Abstract. (519).

EMCDDA. (2017). Drugs and the darknet: Perspectives for enforcement, research and policy. Luxembourg: EMCDDA-Europol Joint Publications.

ENISA. (2017a). ENISA overview of cybersecurity and related terminology.

ENISA. (2017b). *Tools and Methodologies to Support Cooperation between CSIRTs and Law Enforcement*. https://doi.org/10.2824/177198

Esparza, J. M. (2019). Understanding the credential theft lifecycle. *Computer Fraud & Security Bulletin*, *2019*(2), 6–9. https://doi.org/10.1016/S1361-3723(19)30018-1

European Commission. (2019). Cybercrime. Retrieved September 10, 2019, from https://ec.europa.eu/home-affairs/what-we-do/policies/cybercrime_en

Europol. (2018). Internet Organised Crime Threat Assessment (IOCTA) 2018. https://doi.org/10.2813/858843

Franklin, J., Paxson, V., Perrig, A., & Savage, S. (2007). An Inquiry into the Nature and Causes of the Wealth of Internet Miscreants. *CCS'07, Oct 29 - Nov 2, 2007, Alexandria, Virginia, USA*.

Gordon, L. A., & Loeb, M. P. (2002). The Economics of Information Security Investment. *ACM Transactions on Information and System Security*, *5*(4), 438–457.

Green, D. P., & Shapiro, I. (1996). Pathologies of Rational Choice Theory: A Critique of Applications in Political Science. New Haven: CT: Yale University Press.

Hayes, J. & Bodhani, A. (2013) Cyber Security: Small Firms Under Fire. *Engineering and Technology Magazine*

Herley, C., & Florencio, D. (2010). Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy. *Economics of Information Security and Privacy*, 33–53.

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

Holt, T. J. (2012). Examining the Forces Shaping Cybercrime Markets Online. *Social Science Computer Review*, *31*(2), 165–177. https://doi.org/10.1177/0894439312452998

Holt, T. J., Smirnova, O., & Chua, Y. T. (2016). Exploring and Estimating the Revenues and Profits of Participants in Stolen Data Markets. *Deviant Behavior*, (37:4), 353–367. https://doi.org/10.1080/01639625.2015.1026766

IBM. (n.d.) Method (Multiple imputation). Retrieved 25-03-2020 from https://www.ibm.com/support/knowledgecenter/en/SSLVMB_24.0.0/spss/mva/id_idd_mi_method.html

IBM Security. (2018). IBM Study: Hidden Costs of Data Breaches Increase Expenses for Businesses. Retrieved September 15, 2019, from https://newsroom.ibm.com/2018-07-10-IBM-Study-Hidden-Costs-of-Data-Breaches-Increase-Expenses-for-Businesses

ITU-T. (2008). ITU-T Recommendations X.1205.

Jones, J. A. (2005). An Introduction to Factor Analysis of Information Risk (FAIR). *Risk Management Insight*.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk.

Econometrica: Journal of the Econometric Society, 263–291.

Kamiya, S., Kang, J., Kim, J., Milidonis, A. & Stulz, R. (2018). What is the impact of successful cyberattacks on target firms?. *National Bureau of Economic Research*, Cambridge, MA

Kshetri, N. (2005). Pattern of global cyber war and crime: A conceptual framework. *Journal of International Management*, 541-562

Kroll. (2017). *Global Fraud & Risk Report*.

Lagazio, M., Sherif, N., & Cushman, M. (2014). A multi-level approach to understanding the impact of cyber crime on the financial sector. *Computers & Security*, *45*, 58–74.

Leukfeldt, E. R., & Yar, M. (2016). Applying Routine Activity Theory to Cybercrime: A Theoretical and Empirical Analysis. *Deviant Behavior*, *37* (3), 263–280. doi: 10.1080/01639625.2015.1012409

Malin, C. H., Gudaitis, T., Holt, T. J., & Kilger, M. (2017). Social Dynamics of Deception: Cyber Underground Markets and Cultures. In *Deception in the Digital Age* (pp. 125–148). https://doi.org/10.1016/B978-0-12-411630-6.00004-9

Martin, J. (2013). Lost on the Silk Road: Online drug distribution and the "cryptomarket." *Criminology and Criminal Justice*, *14*(3), 351–367. https://doi.org/http://doi.org/10.1177/1748895813505234

Mirian, A., DeBlasio, J., Savage, S., Voelker, G. M., & Thomas, K. (2019). Hack for Hire: Exploring the Emerging Market for Account Hijacking. In *WWW'19, May 2019, San Francisco, CA, USA* (Vol. 2). https://doi.org/https://doi.org/10.1145/33085583313489

Motoyama, M., McCoy, D., Levchenko, K., Savage, S., & Voelker, G. M. (2011). An Analysis of Underground Forums. In *ICM, Berlin, Germany*.

Natalius, S. (2018). Assessing the Role of Online Banking's Characteristics in the Target Selection of the Banking Malware. (Master thesis Delft University of Technology).

NIST. (2018). Framework for Improving Critical Infrastructure Cybersecurity. https://doi.org/10.6028/NIST.CSWP.04162018

Oikarinen, J., & Reed, D. (1993). Internet Relay Chat Protocol. Retrieved from https://tools.ietf.org/html/rfc1459

PYMNTS. (2018). The Rising Tide Of Loyalty Fraud And How To Stop It. Retrieved September 17, 2019, from https://www.pymnts.com/news/security-and-risk/2018/loyalty-rewards-programs-digital-fraud-prevention/

Rajnovic, D. (2012). Cyberspace - What is it? Retrieved September 11, 2019, from https://blogs.cisco.com/security/cyberspace-what-is-it

Romanosky, S., Telang, R., Acquisti, A. (2011) Do Data Breach Disclosure Laws Reduce Identity Theft? *Journal of Policy Analysis and Management*, Vol. 30, No. 2, 256–286

Rosenquist, M. (2009). Prioritizing Information Security Risks with Threat Agent Risk Assessment.

Rowe, B., & Gallaher, M. (2006). Private sector cyber security investment strategies: An empirical analysis.

Sato, Y. (2013). Rational choice theory. *Sociopedia.isa*, 1–10. https://doi.org/10.1177/205684601372

Sen, R., & Borle, B. (2015) Estimating the Contextual Risk of Data Breach: An Empirical

Approach, Journal of Management Information Systems, 32:2, 314-341, DOI:10.1080/07421222.2015.1063315

Shape Security. (2018). *2018 Credential Spill Report*.

Sonnenreich, W., Albanese, J., & Stout, B. (2006). Return On Security Investment (ROSI) – A Practical Quantitative Model. *Journal of Research and Practice in Information Technology*, *38*(1), 45–56.

Symantec. (2018). *Internet Security Threat Report* (Vol. 23).

Tajalizadehkhoob, S., Asghari, H., Gañán, C., & Eeten, M. V. (2014). Why Them? Extracting Intelligence about Target Selection from Zeus Financial Malware 1 Introduction. , 1–26.

Thomas, K., Li, F., Zand, A., Barrett, J., Ranieri, J., Invernizzi, L., … Bursztein, E. (2017). Data Breaches, Phishing, or Malware? Understanding the Risks of Stolen Credentials. In *CCS'17, Oct. 30-Nov. 3, 2017, Dallas, TX, USA*. https://doi.org/10.1145/3133956.3134067

Thomas, K., McCoy, D., Grier, C., Kolcz, A., & Paxson, V. (2013). Trafficking Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse. In *Proceedings of the 22nd USENIX Security Symposium, 2018, Washington, DC, USA*.

Thomas, K., Yuxing, D., David, H., Elie, W., Grier, B. C., Holt, T. J., … Vigna, G. (2015). Framing Dependencies Introduced by Underground Commoditization. In *In Proceedings of the Workshop on Economics of Information Security (WEIS)*.

Thomas, R., & Martin, J. (2006). The Underground Economy: Priceless. *;login:*, *31*(6), 7–16.

van den Berg, J., van Zoggel, J., Snels, M., van Leeuwen, M., Boeke, S., van de Koppen, L., … de Bos, T. (2014). On (the Emergence of) Cyber Security Science and its Challenges for Cyber Security Education. *NATO STO/IST-122 Symposium in Tallin*, (c), 1–10.

Van Moorsel, D. (2016). Target selection regarding financial malware attacks within the Single Euro Payments Area (Doctoral dissertation, Delft University of Technology).

van Wegberg, R., Tajalizadehkhoob, S., Soska, K., Akyazi, U., Ganan, C. H., Klievink, B., … Mellon, C. (2018). Plug and Prey? Measuring the Commoditization of Cybercrime via Online Anonymous Markets. In *Proceedings of the 27th USENIX Security Symposium, August 15-17, 2018, Baltimore, MD, USA*.

Verizon. (2012). 2012 Data Breach Investigation Report.

Wilson, E. (2019). Disrupting dark web supply chains to protect precious data. *Computer Fraud & Security Bulletin*, *2019*(4), 6–9. https://doi.org/10.1016/S1361-3723(19)30039-9

Yar, M. (2005). The novelty of 'cybercrime' an assessment in light of routine activity theory.*European Journal of Criminology*, *2*(4), 407–427.

Zhao, J. (2007). Significance of Risk Quantification: The Smart Decision-Making Process. In *Palisade at Risk User Conference*.

Appendix A

Correlation matrix

Table A.1 Correlation matrix

Correlations

			Relative Price	Domain Popularity	Area Serviced	Reputation	Average Revenue per Year	Net Income	Total Assets	Total Equity	Number of Employees
Spearman's rho	Relative Price	Correlation Coefficient	1.000	-.161	.045	-.333*	.329*	.119	.285	.219	.248
		Sig. (2-tailed)	.	.281	.766	.022	.024	.424	.052	.140	.092
		N	47	47	47	47	47	47	47	47	47
	Domain Popularity	Correlation Coefficient	-.161	1.000	-.398**	.407**	-.331*	-.275	-.400**	-.364*	-.021
		Sig. (2-tailed)	.281	.	.006	.005	.023	.062	.005	.012	.890
		N	47	47	47	47	47	47	47	47	47
	Area Serviced	Correlation Coefficient	.045	-.398**	1.000	-.117	-.120	-.055	.055	.003	-.206
		Sig. (2-tailed)	.766	.006	.	.432	.422	.714	.714	.982	.165
		N	47	47	47	47	47	47	47	47	47
	Reputation	Correlation Coefficient	-.333*	.407**	-.117	1.000	-.697**	-.274	-.375**	-.145	-.628**
		Sig. (2-tailed)	.022	.005	.432	.	.000	.062	.009	.330	.000
		N	47	47	47	47	47	47	47	47	47
	Average Revenue per Year	Correlation Coefficient	.329*	-.331*	-.120	-.697**	1.000	.444**	.557**	.376**	.693**
		Sig. (2-tailed)	.024	.023	.422	.000	.	.002	.000	.009	.000
		N	47	47	47	47	47	47	47	47	47
	Net Income	Correlation Coefficient	.119	-.275	-.055	-.274	.444**	1.000	.802**	.723**	.241
		Sig. (2-tailed)	.424	.062	.714	.062	.002	.	.000	.000	.103
		N	47	47	47	47	47	47	47	47	47
	Total Assets	Correlation Coefficient	.285	-.400**	.055	-.375**	.557**	.802**	1.000	.803**	.247
		Sig. (2-tailed)	.052	.005	.714	.009	.000	.000	.	.000	.095
		N	47	47	47	47	47	47	47	47	47
	Total Equity	Correlation Coefficient	.219	-.364*	.003	-.145	.376**	.723**	.803**	1.000	.199
		Sig. (2-tailed)	.140	.012	.982	.330	.009	.000	.000	.	.179
		N	47	47	47	47	47	47	47	47	47
	Number of Employees	Correlation Coefficient	.248	-.021	-.206	-.628**	.693**	.241	.247	.199	1.000
		Sig. (2-tailed)	.092	.890	.165	.000	.000	.103	.095	.179	.
		N	47	47	47	47	47	47	47	47	47

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Appendix B

Collinearity diagnostic VIF factor

Table B 1 Collinearity Diagnostic VIF

Coefficients[a]

Model		Standardized Coefficients			Collinearity Statistics	
		Beta	t	Sig.	Tolerance	VIF
1	(Constant)		3.826	.000		
	Domain Popularity	-.435	-2.039	.049	.421	2.374
	Area Serviced	-.295	-1.735	.091	.663	1.508
	Reputation	-.157	-.895	.377	.620	1.613
	Category	-.226	-1.414	.166	.750	1.334
	Average Revenue per Year	-.707	-.810	.423	.025	39.711
	Net Income	.036	.044	.965	.029	34.348
	Total Assets	-.365	-.644	.524	.060	16.781
	Total Equity	.252	.254	.801	.019	51.480
	Number of Employees	.619	.802	.428	.032	31.083

Appendix C

Negative Binomial Regression for Each Variable

Syntax call:

```
GENLIN fraudCount BY category AreaServiced reputation (ORDER=ASCENDING) WITH lifespan
sizeC1 sizeC2  popularity price
  /MODEL lifespan category AreaServiced popularity reputation price sizeC1 sizeC2
INTERCEPT=YES
 DISTRIBUTION=NEGBIN(1) LINK=LOG
  /CRITERIA METHOD=FISHER(1) SCALE=1 COVB=MODEL MAXITERATIONS=100
MAXSTEPHALVING=5
   PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-012 ANALYSISTYPE=3(WALD) CILEVEL=95
CITYPE=WALD
   LIKELIHOOD=FULL
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION.
```

Area Serviced

Goodness of Fit[b]

	Value	df	Value/df
Deviance	71.635	41	1.747
Scaled Deviance	71.635	41	
Pearson Chi-Square	88.335	41	2.155
Scaled Pearson Chi-Square	88.335	41	
Log Likelihood[a]	-357.131		
Akaike's Information Criterion (AIC)	726.261		
Finite Sample Corrected AIC (AICC)	728.361		
Bayesian Information Criterion (BIC)	737.362		
Consistent AIC (CAIC)	743.362		

Dependent Variable: Fraud Count

Model: (Intercept), AreaServiced, lifespan, category

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
24.243	5	.000

Dependent Variable: Fraud Count

Model: (Intercept), AreaServiced, lifespan, category

a. Compares the fitted model against the intercept-only model.

			Hypothesis Test		
Parameter	B	Std. Error	Wald Chi-Square	df	Sig.
(Intercept)	6.284	.3879	262.532	1	.000
[AreaServiced=0]	.126	.3812	.109	1	.741
[AreaServiced=1]	0[a]	.	.	.	.
lifespan	.000	.0042	.005	1	.945
[category=1]	.989	.5079	3.792	1	.051
[category=2]	-1.486	.5836	6.482	1	.011
[category=3]	.755	.6618	1.300	1	.254
[category=4]	0[a]	.	.	.	.
(Scale)	1[b]				
(Negative binomial)	1				

Popularity

Goodness of Fit[b]

	Value	df	Value/df
Deviance	64.242	41	1.567
Scaled Deviance	64.242	41	
Pearson Chi-Square	85.067	41	2.075
Scaled Pearson Chi-Square	85.067	41	
Log Likelihood[a]	-353.434		
Akaike's Information Criterion (AIC)	718.869		
Finite Sample Corrected AIC (AICC)	720.969		
Bayesian Information Criterion (BIC)	729.969		
Consistent AIC (CAIC)	735.969		

Dependent Variable: Fraud Count

Model: (Intercept), lifespan, category, popularity

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
31.636	5	.000

Dependent Variable: Fraud Count

Model: (Intercept), lifespan, category, popularity

a. Compares the fitted model against the intercept-only model.

			Hypothesis Test		
Parameter	B	Std. Error	Wald Chi-Square	df	Sig.
(Intercept)	7.385	.5147	205.850	1	.000
lifespan	.003	.0044	.335	1	.563
[category=1]	.091	.5955	.023	1	.879
[category=2]	-1.701	.5077	11.219	1	.001
[category=3]	-.372	.7277	.261	1	.609
[category=4]	0[a]	.	.	.	.
popularity	-.036	.0137	6.844	1	.009
(Scale)	1[b]				
(Negative binomial)	1				

Reputation

Goodness of Fit[b]

	Value	df	Value/df
Deviance	61.136	41	1.491
Scaled Deviance	61.136	41	
Pearson Chi-Square	88.978	41	2.170
Scaled Pearson Chi-Square	88.978	41	
Log Likelihood[a]	-351.881		
Akaike's Information Criterion (AIC)	715.762		
Finite Sample Corrected AIC (AICC)	717.862		
Bayesian Information Criterion (BIC)	726.863		
Consistent AIC (CAIC)	732.863		

Dependent Variable: Fraud Count

Model: (Intercept), lifespan, category, reputation

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
34.742	5	.000

Dependent Variable: Fraud Count

Model: (Intercept), lifespan, category, reputation

a. Compares the fitted model against the intercept-only model.

			Hypothesis Test		
Parameter	B	Std. Error	Wald Chi-Square	df	Sig.
(Intercept)	5.603	.3908	205.556	1	.000
lifespan	.003	.0045	.563	1	.453
[category=1]	.876	.4475	3.829	1	.050
[category=2]	-2.043	.5347	14.594	1	.000
[category=3]	.252	.6098	.171	1	.679
[category=4]	0[a]	.	.	.	.
[reputation=0]	1.063	.3335	10.166	1	.001
[reputation=1]	0[a]	.	.	.	.
(Scale)	1[b]				
(Negative binomial)	1				

Price

Goodness of Fit[b]

	Value	df	Value/df
Deviance	70.406	41	1.717
Scaled Deviance	70.406	41	
Pearson Chi-Square	88.722	41	2.164
Scaled Pearson Chi-Square	88.722	41	
Log Likelihood[a]	-356.516		
Akaike's Information Criterion (AIC)	725.033		
Finite Sample Corrected AIC (AICC)	727.133		
Bayesian Information Criterion (BIC)	736.134		
Consistent AIC (CAIC)	742.134		

Dependent Variable: Fraud Count

Model: (Intercept), lifespan, category, price

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
25.471	5	.000

Dependent Variable: Fraud Count

Model: (Intercept), lifespan, category, price

a. Compares the fitted model against the intercept-only model.

			Hypothesis Test		
Parameter	B	Std. Error	Wald Chi-Square	df	Sig.
(Intercept)	6.446	.3153	418.025	1	.000
lifespan	-.002	.0045	.211	1	.646
[category=1]	.988	.4846	4.157	1	.041
[category=2]	-1.503	.5309	8.014	1	.005
[category=3]	.699	.6060	1.332	1	.248
[category=4]	0[a]	.	.	.	.
price	-.341	.2714	1.581	1	.209
(Scale)	1[b]				
(Negative binomial)	1				

Size Component 1

Goodness of Fit[b]

	Value	df	Value/df
Deviance	71.699	41	1.749
Scaled Deviance	71.699	41	
Pearson Chi-Square	89.108	41	2.173
Scaled Pearson Chi-Square	89.108	41	
Log Likelihood[a]	-357.163		
Akaike's Information Criterion (AIC)	726.325		
Finite Sample Corrected AIC (AICC)	728.425		
Bayesian Information Criterion (BIC)	737.426		
Consistent AIC (CAIC)	743.426		

Dependent Variable: Fraud Count

Model: (Intercept), lifespan, category, sizeC1

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
24.179	5	.000

Dependent Variable: Fraud Count

Model: (Intercept), lifespan, category, sizeC1

a. Compares the fitted model against the intercept-only model.

			Hypothesis Test		
Parameter	B	Std. Error	Wald Chi-Square	df	Sig.
(Intercept)	6.349	.3146	407.295	1	.000
lifespan	.000	.0042	.001	1	.978
[category=1]	.933	.4767	3.832	1	.050
[category=2]	-1.585	.5193	9.313	1	.002
[category=3]	.730	.6859	1.132	1	.287
[category=4]	0[a]	.	.	.	.
sizeC1	-.043	.2020	.045	1	.831
(Scale)	1[b]				
(Negative binomial)	1				

Size Component 2

Goodness of Fit[b]

	Value	df	Value/df
Deviance	71.447	41	1.743
Scaled Deviance	71.447	41	
Pearson Chi-Square	85.908	41	2.095
Scaled Pearson Chi-Square	85.908	41	
Log Likelihood[a]	-357.037		
Akaike's Information Criterion (AIC)	726.074		
Finite Sample Corrected AIC (AICC)	728.174		
Bayesian Information Criterion (BIC)	737.175		
Consistent AIC (CAIC)	743.175		

Dependent Variable: Fraud Count

Model: (Intercept), lifespan, category, sizeC2

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
24.431	5	.000

Dependent Variable: Fraud Count

Model: (Intercept), lifespan, category, sizeC2

a. Compares the fitted model against the intercept-only model.

			Hypothesis Test		
Parameter	B	Std. Error	Wald Chi-Square	df	Sig.
(Intercept)	6.276	.3471	326.996	1	.000
lifespan	1.818E-5	.0042	.000	1	.997
[category=1]	1.026	.5028	4.161	1	.041
[category=2]	-1.502	.5298	8.036	1	.005
[category=3]	.780	.6375	1.497	1	.221
[category=4]	0[a]	.	.	.	.
sizeC2	.141	.2823	.251	1	.616
(Scale)	1[b]				
(Negative binomial)	1				

Appendix D

Linear Regression for Each Variable

Syntax call:

GENERAL LINEAR MODEL

UNIANOVA lnloss BY category AreaServiced reputation WITH lifespan price popularity sizeC1 sizeC2
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/SAVE=PRED RESID
/PRINT=PARAMETER
/CRITERIA=ALPHA(.05)
/DESIGN=category AreaServiced reputation lifespan price popularity sizeC1 sizeC2.

Area Serviced

Tests of Between-Subjects Effects

Dependent Variable:lnloss

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	47.030[a]	5	9.406	2.914	.024
Intercept	919.863	1	919.863	284.957	.000
category	24.646	3	8.215	2.545	.069
AreaServiced	14.051	1	14.051	4.353	.043
lifespan	4.211	1	4.211	1.304	.260
Error	132.351	41	3.228		
Total	5508.564	47			
Corrected Total	179.381	46			

a. R Squared = .262 (Adjusted R Squared = .172)

Parameter	B	Std. Error	t	Sig.
Intercept	8.537	.649	13.163	.000
[category=1]	1.166	.729	1.600	.117
[category=2]	2.384	.945	2.521	.016
[category=3]	2.247	1.054	2.131	.039
[category=4]	0[a]	.	.	.
[AreaServiced=0]	1.326	.635	2.086	.043
[AreaServiced=1]	0[a]	.	.	.
lifespan	.007	.006	1.142	.260

Popularity

Tests of Between-Subjects Effects

Dependent Variable:lnloss

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	33.629[a]	5	6.726	1.892	.117
Intercept	619.724	1	619.724	174.328	.000
category	15.723	3	5.241	1.474	.236
lifespan	2.394	1	2.394	.674	.417
popularity	.650	1	.650	.183	.671
Error	145.752	41	3.555		
Total	5508.564	47			
Corrected Total	179.381	46			

a. R Squared = .187 (Adjusted R Squared = .088)

Parameter	B	Std. Error	t	Sig.
Intercept	9.511	.788	12.063	.000
[category=1]	1.019	.769	1.324	.193
[category=2]	1.880	.959	1.960	.057
[category=3]	1.580	1.116	1.416	.164
[category=4]	0[a]	.	.	.
lifespan	.005	.006	.821	.417
popularity	-.009	.022	-.428	.671

Reputation

Tests of Between-Subjects Effects

Dependent Variable:lnloss

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	40.070[a]	5	8.014	2.359	.057
Intercept	882.108	1	882.108	259.610	.000
category	12.289	3	4.096	1.206	.320
reputation	7.091	1	7.091	2.087	.156
lifespan	3.489	1	3.489	1.027	.317
Error	139.311	41	3.398		
Total	5508.564	47			
Corrected Total	179.381	46			

a. R Squared = .223 (Adjusted R Squared = .129)

Parameter	B	Std. Error	t	Sig.
Intercept	8.880	.621	14.290	.000
[category=1]	1.024	.746	1.371	.178
[category=2]	1.528	.955	1.600	.117
[category=3]	1.650	1.050	1.572	.124
[category=4]	0[a]	.	.	.
[reputation=0]	.810	.561	1.445	.156
[reputation=1]	0[a]	.	.	.
lifespan	.006	.006	1.013	.317

Price

Tests of Between-Subjects Effects

Dependent Variable:lnloss

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	34.475[a]	5	6.895	1.951	.107
Intercept	812.073	1	812.073	229.770	.000
category	12.898	3	4.299	1.217	.316
lifespan	3.505	1	3.505	.992	.325
price	1.496	1	1.496	.423	.519
Error	144.906	41	3.534		
Total	5508.564	47			
Corrected Total	179.381	46			

a. R Squared = .192 (Adjusted R Squared = .094)

Parameter	B	Std. Error	t	Sig.
Intercept	9.243	.570	16.201	.000
[category=1]	.939	.784	1.197	.238
[category=2]	1.663	.985	1.689	.099
[category=3]	1.634	1.076	1.518	.137
[category=4]	0[a]	.	.	.
lifespan	.007	.007	.996	.325
price	.263	.404	.651	.519

Size Component 1

Tests of Between-Subjects Effects

Dependent Variable:lnloss

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	37.347[a]	5	7.469	2.156	.078
Intercept	911.796	1	911.796	263.201	.000
category	17.572	3	5.857	1.691	.184
lifespan	1.532	1	1.532	.442	.510
sizeC1	4.368	1	4.368	1.261	.268
Error	142.034	41	3.464		
Total	5508.564	47			
Corrected Total	179.381	46			

a. R Squared = .208 (Adjusted R Squared = .112)

Parameter	B	Std. Error	t	Sig.
Intercept	9.227	.564	16.356	.000
[category=1]	1.305	.784	1.665	.103
[category=2]	1.832	.940	1.948	.058
[category=3]	1.914	1.075	1.781	.082
[category=4]	0[a]	.	.	.
lifespan	.004	.006	.665	.510
sizeC1	-.327	.291	-1.123	.268

Size Component 2

Tests of Between-Subjects Effects

Dependent Variable:lnloss

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	33.448[a]	5	6.690	1.879	.119
Intercept	907.959	1	907.959	255.092	.000
category	14.455	3	4.818	1.354	.270
lifespan	2.277	1	2.277	.640	.428
sizeC2	.469	1	.469	.132	.718
Error	145.933	41	3.559		
Total	5508.564	47			
Corrected Total	179.381	46			

a. R Squared = .186 (Adjusted R Squared = .087)

Parameter	B	Std. Error	t	Sig.
Intercept	9.330	.587	15.881	.000
[category=1]	.988	.790	1.251	.218
[category=2]	1.785	.962	1.855	.071
[category=3]	1.635	1.094	1.495	.142
[category=4]	0[a]	.	.	.
lifespan	.005	.006	.800	.428
sizeC2	-.107	.294	-.363	.718