University of Bergen
Department of informatics

Mitigation of Identity Theft in Online Banking

Author: Kaja Alexandra Dey

Supervisor: Håvard Raddum

UNIVERSITETET I BERGEN
Det matematisk-naturvitenskapelige fakultet

June, 2019

Acknowledgments

I would like to thank my main supervisor, Håvard Raddum at Simula@UiB, for providing guidance and support. I would also like to thank Sbanken, with special mentions of Lars Nestås and Vidar Drageide, for collaborating with us on this thesis.

My friends and peers from the study hall have supported me, both through good discussions and with motivational boosts. Thank you to Ragnhild Aalvik and Kristian Rosland, for being excellent study partners during my years at UiB. You are the reason I came to love informatics. Finally, I want to thank my family for their encouragement, backing, and help with proofreading this thesis. I am ever grateful.

Abstract

Identity theft in online banking can cause significant economic and psychological damage to the victims. Traditionally, there has been a strong focus on detecting and hindering fraud, but not nearly as much focus on any identity thefts that may have been involved in these fraud cases. This is beginning to change, but we still know very little about how identity theft happens in online banking, or how to stop it. We have, together with Sbanken, looked closely at a set of reported identity theft cases. Statistics about these cases have been analyzed to see if some subsets of the population are more at risk. Starting with 10 hypotheses, we have worked together to make several functions that should warn of possible identity thefts in the bank.

Our main hypothesis is that it is possible to detect identity theft by analyzing the metadata of each individual user account. Identity theft happens to individuals, and every case might seem like it is unique. Still, we have found some shared patterns between several of the cases, which can be used in detecting future occurrences of identity theft.

Our new functions will need further testing on the customers in a bank in order to measure their full effectiveness. Nonetheless, the process has been started, and we have some preliminary findings that indicate that our hypothesis is correct.

Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

2FA	two-factor authentication.
FN	False Negative.
FP	False Positive.
IRS	Internal Revenue Service.
KBA	Knowledge Based Authentication.
NIN	National Identification Number.
NorSIS	The Norwegian Center for Information Security.
OTP	one-time password.
TN	True Negative.
TP	True Positive.

Chapter 1

Introduction

Identity theft is not a new phenomenon, but the increasingly fast digitalization that has happened during the last two decades has brought with it some very damaging side effects. The laws regarding this type of crime, and the organizations that are supposed to protect the Norwegian citizens, have not evolved fast enough to keep up with the consequences of modern identity theft.

M. E. Kjørven explains about the "victims of digitalization" in a feature story published by NRK [1]. Identity theft can happen to anyone, but as Kjørven says, it often affects the weakest people in society.

This thesis will examine what identity theft is, how it can happen, and what banks can do in order to better detect and stop it. We have analyzed data about reported identity thefts and used this as a basis for creating specific checks that can help in the mitigation of identity theft in online banking. As the consequences of identity theft can grow devastating for the victims, it becomes all the more important to find ways of stopping the thefts from happening in the first place.

1.1 Definitions

Merriam Webster's online dictionary defines identity theft, first observed in 1964, as "the illegal use of someone else's personal information (such as a Social Security number) especially in order to obtain money or credit" [2]. The Norwegian Center for Information Security (NorSIS) defines *identity theft* as the acquiring, transition or display of personal information that is not your own, with the intent of committing fraud or another criminal act. *Identity* *fraud* is the use of the stolen identity to gain money, goods or services, or to avoid an obligation by presenting oneself as someone else using the fake or stolen identity [3].

This paper will examine identity theft in the context of Norwegian online banking. Identity theft will be defined as the illegal act of using personal information not belonging to oneself, to gain access to online banking services. This will include both creating a user account by using another person's credentials, the takeover of an existing account, and any subsequent use of the account.

An account in the context of banking can mean both user account or bank account. A person can only have one user account in each bank, but this user account can have several bank accounts connected to it. In the rest of this thesis, unless specified otherwise, all mentions of an account will refer to a user account.

1.2 Context

1.2.1 Online Banking

Norwegians first gained the ability to use online banking in 1996, when Sparebanken Hedmark launched Europe's first online banking system [4, p.8]. Much has happened since then, and a large part of the daily life of Norwegian citizens now occurs online. Online banking has become an effective way to pay bills, get increased insight into one's own economy, and it promotes freer access to goods, nearly independently of one's geographical location. Numbers from Statistics Norway from August 2018 show that 96% of the population (age 16-79) have used the Internet during the last 3 months, and 92% reported that they use the Internet for banking purposes [5]. The 2018 edition of DIBS' yearly statistical report detailing Norwegian e-trade, found that Norwegians used 124.2 billion NOK on online services, goods, and travels in 2017 [6]. This is a substantial amount for a population that in 2017 only amounted to 5.26 million [7].

1.2.2 Identity Theft in Online Banking

With such massive sums of money changing hands online, security becomes crucial in hindering illegal or unwanted activity from happening. Weak security in online banking can facilitate fraud, money laundering, and theft. The question of *identity* becomes vital, as the banks endeavor to assert who sits on the other end of a digital transaction. This job becomes all the more difficult when we consider the implications of *identity theft*, which enables a nefarious third party to impersonate an innocent person whom the banks normally would have no problem with.

Even before the time of the Internet, identity theft could be used for actions such as acquiring illegal admittance to a country or to receive a work permit. Nowadays, as the use of online banking become more widespread, identity theft can very easily give almost instant access to credit card loans and a plethora of online stores. Many people share personal or sensitive information online via services such as social media, which only makes it all the more easy for criminals to conduct identity theft.

A report from 2017 done by NorSIS and The Norwegian Tax Administration concluded that while the number of identity thefts reported during the last few years has stayed mostly the same, the consequences of these thefts affect more people, and have a wider scope [8]. The consequences of identity theft can quickly become devastating for the victim. There is often a substantial sum of money involved, and it is difficult to prove the victim's innocence since the victim's name and personal information are registered as the guilty party. This can lead to years of clean-up work, and having to respond to new incoming claims from creditors.

A "lucky victim", if one can say such a thing, may receive a phone call, or an email, asking for confirmation on a sale or loan. Otherwise, it might take several weeks, or even months, before the victim begins receiving mail informing them of outstanding bills that need to be paid. If the criminal is closely related to the victim, then they might get rid of these warning signs, and years might go by without the victim knowing that their identity is being abused. When they finally become informed of the situation, then they might be millions in debt, and find out that they, the victim, have become legally responsible for paying back the entire sum.

1.3 A Gap in the Literature

There is currently a lack of published literature about identity theft in online banking. Some articles and studies touch upon these themes, but their focus is mainly on fraud detection, such as misuse of stolen credit cards.

The published sources that we did find, most of them from before 2012, were already outdated. This literature, as well as the newer works that we found, were from the point of view of social sciences, such as business administration, law, and criminology. One such study is a study of the correlation between identity theft reports and state-level characteristics such as high residential mobility [9]. Another is the "Aftermath" study done by the Identity Theft

Resource Center, which explores the consequences for individuals that have become victims of identity theft [10]. Because of the rapid evolution of online systems, we need updated information about how identity thefts happen today, in order to have a chance of stopping them.

A study on the experiences of fraud investigators, by J. L. Leiner, highlights some very useful points on how we need more preventive measures in combating identity theft and fraud, and that we need ways to educate the public about how someone can steal or misuse their personal information [11]. The focus has been on shining light on how the thieves manage to trick users into giving up personal information, which the thief then can use to commit their crimes, but nothing about how the banks can detect that someone else is in control of the account.

In Chapter 3, we present statistics about identity theft in Norway and look into how identity theft happens here. Companies can choose wildly different ways for authenticating their users, based on what kind of service they offer, and what customer base they cater to. This means that data from other countries are of limited use to us. Still, we can use the information as a guide on what types of authentication is more often misused, and to what degree identity theft affects their society.

In 2016, Navarro and Higgins researched statistics about victims of familial identity theft, using data from an American crime survey from 2012 [12]. In Chapter 4, we present our own findings about Norwegian identity theft victims and compare these to the American cases.

Little is known, or at least publicly known, about how banks can detect identity theft. Most companies will claim that they have a high focus on security and that their customers are safe. Unfortunately, identity thefts still happen, and the criminals manage to steal money. So why does it happen? Is it that the banks don't care, or is identity theft just too complicated for the banks to detect?

1.4 Focus and Value of Study

This thesis is written in collaboration with Sbanken, formerly known as Skandiabanken. They are a Norwegian bank focused solely on online customers. Sbanken has provided us with real data of their customers and transactions, as well as a set of reported identity theft cases. We have not been able to identify studies based on similar premises, and the opportunity for academia to work so closely with a bank to find new methods of detecting and hindering identity theft is, as far as we know, unique.

Norwegian banks currently have quite strong fraud detection systems, as they are required to by law. But identity theft cases have somehow slipped past their defenses. Statistics Norway first began reporting identity theft cases in their crime statistics in 2015, which indicates that analyzing identity theft is a somewhat new thing to do. In Chapter 3.3, we will look more closely at what consequences identity theft causes today. An important aspect of why identity thefts still happen might be that there are no obvious economic gains for a bank to invest even more money into stronger detection of identity theft. As we will see in Chapter 3, it is mostly the victim, or maybe an insurance company, that is left with the economic responsibility of the identity theft. Still, one can say that the banks have a social responsibility to protect all of their customers to their greatest capacity.

Before writing this thesis, Sbanken had mechanisms made for detecting different types of fraud, but usually only became aware of any identity theft after they had been contacted by the victim. Sbanken believed that they could improve this, and create systems that would be able to give early detection of identity theft. They also wanted to identify any security mechanisms that could be improved upon, to stop identity thefts from ever taking place. This led to the collaboration on this master thesis.

We were given insight into how a Norwegian bank worked with detecting and hindering fraud. We used this opportunity to examine the data that they had available, and research if their existing data could be used to get better results in detecting identity theft cases. Our focus has been on collecting information about what characteristics distinguish identity theft victims from the rest of the customers, and what patterns we can look for to detect identity theft.

A common factor among most identity theft cases is that early detection is crucial in keeping the ramifications of the theft as small as possible. Our goal has been to find ways to detect patterns that indicate identity theft, so that future victims may be warned instantly, and might avoid the economic and psychological burden they might otherwise be exposed to.

An important specification that needs to be made, is that identity theft is seldom something that only happens once, in one bank. If a victim has had their identity stolen, then this same identity will most likely be used to gain access to money from several different institutions. A compromised user that is denied a loan in one bank, might receive loans in other, less strict banks. This is why detecting the presence of identity theft is important, regardless of whether the customer managed to get a loan or not, since it might help the victim in stopping other transgressions.

1.5 Specific Research Aims and Objectives

The goal of this thesis has been to use the reported identity theft cases provided by Sbanken, to look for patterns that can enable Sbanken to recognize future identity theft cases for early mitigation. This goal has been re-formulated in Hypothesis 1:

Hypothesis 1. *It is possible to detect identity theft to a greater extent by looking at the metadata of each user account and its transactions.*

Several sub-goals have been used to guide this thesis. These sub-goals have been the basis for our work, and have been used as a way of measuring how well our security mechanisms and checks will work in the bank. They were also used to formulate 9 other hypotheses, which are more specific than our sub-goals. These hypotheses will be explained in detail in Chapter 4.2. Here are our sub-goals:

1. Understand how and why identity theft happens. Research why identity theft is still a problem, and why we should care enough to stop it from happening.

2. Understand the different nuances of identity theft, and why there can be significant differences between individual cases of identity theft. This includes researching if these differences can be explained by looking at the relations between the victim and the criminal, and if there has been an account takeover or creation of a new account.

3. Looking at reported identity theft cases to see if there are some underlying patterns and if some groups of people are more susceptible to identity theft than others.

4. Analyze log files and user activity in order to find parameters that will indicate the presence of identity theft.

5. Find areas of interest and start collecting information about these.

6. Analyze how customers of the bank behave and create methods for detecting changes in these patterns.

To understand how a criminal can bypass authentication mechanisms when performing an identity theft, we first had to see what authentication methods are available, how these work, and what methods are typically used in online banking. Understanding how different types of authentication work and how they can be abused, has given us the insight we needed to develop some new methods that can be used for detecting identity theft in the bank.

We closely examine what identity theft is, how it works, and how wide-spread it is today. To better understand the evolution of identity theft in online banking, we have had to examine what consequences identity theft can have for both the victim, the culprit, and society as a whole.

1.6 Structure of the Thesis

We have given a short introduction to the background and motivation for this thesis, and will now describe the structure of the rest of the thesis:

Chapter 2: Authentication and Online Banking
This Chapter describes what authentication is, and how different types of authentication scores in terms of security and usability. We describe how authentication of individuals is used in online banking in Norway, and what authentication mechanisms Sbanken uses today.

Chapter 3: Identity Theft
Here we look closely at the nuances of identity theft. We present different types of identity theft, and how they can happen. We end the Chapter by explaining what consequences identity theft has for both the criminal, the victim, and for society as a whole.

Chapter 4: Case Study
We have performed a case study based on identity theft cases provided us by Sbanken. We start by presenting our hypotheses, explaining how we created them, and how we worked towards testing their validity. We present our findings for each hypothesis, and what detection systems we have made for detecting identity theft in the bank.

Chapter 5: Discussion and Conclusion
This chapter contains a summary of our findings, and discusses their meaning and importance. We present our ideas and plans for future work.

Appendix A: Norwegian Laws
Norwegian laws that have been used in convicting the victims in identity theft cases, presented and used in Chapter 3.3

Chapter 2

Authentication and Online Banking

2.1 Authentication of Individuals

"Individual Authentication is the process of establishing an understood level of confidence that an identifier refers to a specific individual." [13, p.2]. Authentication in and of itself can only give us a *level of confidence* that a claim is correct. We cannot know with absolute certainty that we have the right person, but it is understood that the method is as good as it gets. Different situations require varying levels of authentication. Some situations, such as gaining entrance to a night club, have more or less negligible consequences for faulty authentication. On the other end of the spectrum are authentications that can have disastrous consequences if they were to fail, such as getting through US Customs and Border Protection.

The way individual authentication works is that an issuer generates some credentials that assert an identity, which can then be presented and validated. An example of this is the use of a passport. The Norwegian police *issue* passports to Norwegian citizens. This citizen can *present* their passport to a controller, who *verifies* that the passport is genuine and representing the correct person. A person is authenticated by looking at identifiers that are connected to that person. These can be divided into strong and weak identifiers. *Strong identifiers*, such as a social security number, act as a unique mapping from that identifier to a single person. *Weak identifiers*, like age or name, can be correctly mapped to more than one person. A collection of weak identifiers that combined can correctly identify an individual, can be viewed as a single strong identifier [13, p.42].

Secure and reliable individual authentication is imperative in order to hold the correct person accountable for their actions. On websites that do not handle financial or personal information, it might be enough to identify an *identity*. That is to say that it is enough to make sure that the user is linked to an identity, such as an email address, without knowing which specific *individual* owns the account. In our context of online banking, this will not be satisfactory; it is necessary for the identity to be linked to a person, and not only an organization or IP address.

Authentication should not be confused with *authorization*, which is to determine what actions an individual is allowed to do. An example of authorization is to differ between the rights of an admin and a normal user. It should also not be confused with *identification*. Identification is the process of using observed identifiers to determine who an individual is, while authentication happens after a person actively presents some claim to who they are, which is then approved or rejected.

Let's look at an example to make this point clearer. A policeman with a sketch of an assailant will use identification to find the correct individual in a database full of pictures, in order to identify who the criminal is. Authentication, on the other hand, is a one-to-one check of whether the credentials that are presented match the credentials connected to that specific person.

2.2 Types of Authentication

Individual authentication credentials are generally divided into three classes; *something you know*, *something you have*, and *something you are*. We will now present each of these groups, and give some pros and cons regarding ease of use, cost of implementation, and their level of security.

2.2.1 Something You Know

"Simple, password-based authentication is the most common form of initial, one-way authentication used in information systems" [13, p.107]. Such passwords are static (non-changing) strings of characters that a user can present as verification, thus showing that they are the correct user of the account.

Most users are already experienced with the use of static passwords, which means that there is no need for additional training costs to get customers or employees familiar with using them. They are also fairly cheap to implement securely.

A downside with the use of static passwords is that humans are involved; people often choose guessable passwords in order to be able to correctly remember them, and reuse the same passwords in multiple locations. This can make it possible to hack the account by simply using educated guesswork or a list of common passwords to brute-force their way inside.

Using static passwords for authentication makes one vulnerable to a long list of attacks, in addition to the more common brute-forcing of bad passwords. Techniques such as social engineering, phishing, and shoulder surfing can be used to steal the password from the user. We will return to these methods in Chapter 3.

The user can be tricked into installing a keylogger on their device, or security weaknesses within the system can be exploited to steal the password. Examples of such weaknesses can be to transmit passwords in clear text, or not mitigating against man-in-the-middle attacks. A man-in-the-middle network attack is when the communication between two systems is intercepted by an outside party. One type of such an attack is the replay attack; first, a nefarious third-party will sniff a password or hash over the network, before sending this information to the server whenever he should want to, pretending to be the actual user of some account.

2.2.2 Something You Have

This class covers individual authentication based on a physical token, which has been made in such a way that it is difficult to forge or alter.

ID Paper An ID-paper is something that is given to the user, which they can use to authenticate themselves. The ID-paper is often valid over a long period of time. This makes it very important that it should not be feasible to copy or alter the information it contains. There are two reasons for this; we don't want someone else to be able to copy the information from a real card to make a forgery, and we don't want a user to be able to change the information that is displayed about themselves.

The Norwegian driver's license uses a combination of several mechanisms such as hologram and reliefs in order to make the license cards more difficult to forge [14]. It is not possible to make them 100% secure, but the level of confidence is evaluated as sufficiently high. This premise is also what is used when a company gives its employees key-cards that they must use to get admission to specific areas.

OTP Device A one-time password (OTP) is, as the name suggests, a password that can only be used once. There are several different ways of both making and accessing OTPs. *Hard tokens* such as code devices or password calculators are commonly generated based on a timestamp and a shared secret key, which are then sent through an algorithm to get a number; the OTP. This code is generated locally by the code device, and not sent over the internet, or by some other signal, so there is no chance for it to be intercepted before it reaches the intended user.

An OTP can also be sent to the user each time it is needed, e.g. by *SMS*. This type of OTP is by default not as secure, as it can be intercepted by a third party. Despite becoming a much used method for sending out OTPs, using SMS to receive important information has turned out to be an extremely bad idea. We will explain why this is so in Chapter 2.2.4. There also exist *authentication apps* that can be downloaded on a smartphone to use as a code device, which will generate the needed OTP for the user.

When we enter the OTP into a web-page, it is checked server-side to see if it matches the code generated there. OTP works in the same way as a session key, where the code or password will only be valid for a short time frame, before it becomes invalid.

A positive aspect of using OTP's is that they are not susceptible to a lot of the security vulnerabilities that static passwords suffer from. Since OTPs only work once, a replay attack is not possible. Of course, there is still the possibility that the code can be snapped up in another man-in-the-middle attack where the criminal will instantly use the OTP to log into the service before the user. The short *time-frame* in which the code is valid makes this harder for the criminal, especially if they need more than one OTP.

This same feature may also cause some difficulty in the making and the usage of the OTP, as it can be difficult to find a good length of the active time-frame; if it is too short, then it may fail constantly due to bad internet, lacking connection or users taking too much time for entering their code. A too long window may make the service vulnerable to attacks.

For hardware tokens, it is a crucial step to properly *synchronize the internal clock* in the code device against that which is used server-side. This clock can also go out of sync, in which case a new token has to be sent to the user. Securely sending the hardware tokens out to the users naturally has a cost connected to it. Using scratch cards to hide the OTPs on a printed plastic card will have a cheaper production cost than electronic devices, but it will have to be sent out more often due to being used up quickly.

Using a mobile phone as an OTP device has become increasingly more popular in recent years. For the companies delivering a service, using an existing authentication app or SMS service means a huge cut in production and delivery cost, since the phone is already in the hands of the user and they know how to use it. It is also practical for the user, as most of us always carry our phone with us everywhere, so we will not be without the OTP device if we should need it.

2.2.3 Something You Are

This type of individual authentication works by "observing physical characteristics of the body of the individual" [13, p.47], and are consequently often called biometric authentication systems. In old spy movies, biometric authentication such as checking fingerprints or iris-scans, could often be observed as part of the security measures when someone wanted to enter a top-secret lab or restricted area. Today, these same types of technologies have become part of our every-day life. Many modern smartphones, as well as Windows 10, have the option of using a fingerprint or facial recognition in order to unlock the device.

Validation of Biometric Authentication

Static passwords and OTP's are easy to validate; either they match, or they don't. The case is not as simple with biometric authentication. When checking e.g. an iris scan, there is bound to be some *noise*, or small disturbances, in the data, making it not an exact copy of the data saved in the database that the measurement is checked against. This gives us the challenge of deciding how much noise we want to accept. It is important that the right person is recognized, but if we allow too many discrepancies in the data, then we might erroneously give access to the wrong person.

		Observed values	
		Positive	Negative
True values	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Table 2.1: This confusion matrix is used to evaluate the relationship between predicted values and real-world results. It is also called an error matrix because of the way the values FP and FN show us where the observed data differ from our predicted/wanted result.

True Positive (TP) and True Negative (TN) means that the authentication was successful; we have correctly authenticated the right person and rejected users that should not be admitted. It follows that False Positive (FP) and False Negative (FN) means that the authentication has failed.

FP means that the wrong person was allowed entrance to the system. This is called *False Acceptance Rate* and is something that ideally should never happen. Unfortunately, it is not without problems to achieve a score of zero FP-cases; if the system is very strict, then the system might reject many true users. This is known as FN, or *False Rejection Rate*, and happens when someone who has authorization to access the system is rejected by the biometric scan.

Cross-Error Rate is a way of measuring the balance between FN and FP that gives us the value where FP and FN are equal. This is used as a way of comparing different biometric authentication methods; the lower the cross-error-rate, the better the method is.

Usually, FP and FN are not of equal importance when looking at each method of authentication, and we prefer to lower one of them even if this means that the other will be higher. The challenge then becomes to find the correct balance between these two values. The dilemma of finding this balance can perhaps be better understood by looking to the juridical system; judges are faced with the possibility that they might wrongly convict an innocent person, or they might let a criminal go without punishment. FN and FP have a negative correlation; low FN will cause a high FP, and vice versa. In the juridical aspect, it is regarded as a better option to allow some criminals to go free than to risk a miscarriage of justice. We find the same dilemma in authentication, but instead of making sure no innocent party gets punished, it is more important to make sure no criminals are allowed into the system.

It can be an annoyance for the user that is faultily refused access, but this does not necessarily have devastating consequences for the company or system. But, if the system becomes too strict, then the user might be locked out of the system completely, and will eventually give up.

Types of Biometric Authetication

There exist numerous types of biometric access control systems. When evaluating a type of biometric authentication, it is important that it fulfills some characteristics; universality, distinctiveness, permanence, and collectability. In short, this means that every person should have the characteristic to be measured, it should be unique for each and every person, it should not be susceptible to significant changes over time, and it should be possible to efficiently determine and quantify it. All possible biometrics are scored according to these criteria.

Visual Biometrics This type of biometric authentication uses image sensors to observe visual characteristics of a person, ranging from ocular-based recognition, such as retinal and iris scan, to facial recognition, and even ear recognition. Visual biometrics are vulnerable to copying or forgery, e.g. by using a printed picture of a face or stealing a fingerprint using sticky-tape. Both retina and iris scan have a very low rate of both FP and FN when a real eye is presented, which is very good. Unfortunately, it is all too easy to make a believable forgery that can trick the scanner. A German hacker collective made a video in 2017 showing how easy it is to fool the iris scanner in a Samsung Galaxy S8 [15]. All they needed was a printed, medium-distance picture of the face of the actual user, and a contact lens used to achieve a 3D effect.

Vein Scanning Fingerprint scanning can be divided into a visual part, where one looks at the ridges and valleys (minutiae) of the fingerprint, and vein recognition, of either a finger or the entire palm. Vein scanning works by placing the finger between a near-infrared light and a monochrome CCD camera. The hemoglobin in the blood will absorb the light, and make the veins appear as dark lines. In comparison, the surrounding flesh and bone will appear lighter. Every vein-pattern is distinct, and if something were to happen to one finger, there would be 9 others to take its place as an identifier. Vein scanning is more secure due to the fact that the veins lie inside the body. The reading is contact-less, making it a very good hygienic choice. It is also impossible to steal the vein-pattern from a dead person, as the lines are only visible in a finger with a constant blood flow [16].

Behavioral Behavioral biometrics measure specific patterns in human activities. Old-fashioned handwritten signature verification is a type of behavioral biometric check that has existed for a very long time. In 1667, the Parliament of England passed a law regarding fraud, which stated that certain contracts had to include a signature in order to be valid. This led to the question of determining the authenticity of the signature. Today, signature software can be used to look at the contours and the movements used to make the signature, as well as the speed of each stroke. There are also numerous other systems that can analyze behavioral characteristics such as individual gait, or looking at the keystrokes dynamics that appear when humans type on a keyboard. Using behavioral measures makes it easier to observe when a pattern changes, which might indicate that we are observing someone other than the intended user.

Voice Authentication Voice authentication is a form of speaker recognition. Speaker recognition is the recognition of *who* says something, and should not be confused with speech recognition, which is recognizing *what* is being said.
Even though each person has a unique voice, the systems must accept some errors in the measurement to allow for a mild cold or background noise. This allowance of errors can make it possible for a close family member to trick the system, as seen when the twin of a BBC reporter in 2017 managed to bypass the speaker recognition of the bank used by his brother, and authenticate himself in his brother's name [17].

2.2.4 Multi-factor Authentication

Multi-factor authentication is the combination of two or more authentication methods—from different classes—in order to more reliably authenticate a person. This means that having two passwords does not mean that one has two-factor authentication (2FA); the user must use methods belonging to two different classes, such as a password, and a device that he owns. Many systems today have started to use 2FA to increase their security. Google lets its users choose from a growing number of possible ways of using a second authentication method, in addition to the standard password. It is possible to set up a verification code via SMS, voice call, a code generated by an authentication app such as Google Authenticator, or via a USB-based security device such as YubiKey. These USB-devices can be password protected to mitigate against a thief getting access to all your passwords, but this makes it imperative that the password is strong, and not forgotten by the user.

Sending an OTP by SMS is a well-used way of implementing 2FA, as it means the user must have access to a physical device; their mobile phone. As most people today carry their smartphone with them wherever they go, it seems like a good alternative to the old OTP generators or code cards. Unfortunately, it has been established that this is a method that is easy to trick. There have been numerous attacks against telecommunications companies to trick them into sending out SIM cards or switching the user's service to a new phone. This type of attack is now known as *SIM Swap Fraud*. When the new SIM activates, the fraudster can very quickly use the SMS-feature to log into banking services. They also have the ability to reset passwords, and in that way quickly gain access to the user's numerous accounts, and even block the user out of their accounts.

Displaying incoming SMS' or email-content on the start screen of a cell phone means that the phone does not need to be unlocked in order to steal a password or other sensitive information. This can give the false impression of having another level of security than one actually has. The same can be said if one sends the security code to an email address that is accessible via an unprotected phone.

Some verification methods are inherently stronger than others, and we usually get a trade-off between security and convenience. Using e.g. a YubiKey, or an authentication app such as Google Authenticator, might be the best options available today. Unfortunately, this will require a bit more effort than e.g. receiving codes by SMS, so it will take time and effort to convince the general public that it is worth the effort to properly secure their digital life.

As a rule of thumb, it is better to implement some type of multi-factor authentication whenever possible, but it is important to be aware of any possible weaknesses in the chosen method. The user needs to be aware of what can happen if someone were to get hold of the physical device used in the verification. Also, if one uses a single account as a recovery-function for all other accounts, then this might get exploited to bypass all additional levels of security.

2.3 Authentication in Norway

2.3.1 National Identification Number

The Norwegian National Identification Number (NIN) is called Fødselsnummer, which literally means "birth number". It is an 11-digit personal identifier given to everyone registered in the Norwegian National Registry [18]. The NIN has the following format:

$$X_1X_2X_3X_4X_5X_6I_1I_2I_3C_1C_2$$

The first six digits represent the date of birth, on the form DDMMYY. Then follows five numbers that constitute the personal number. The personal number is again composed of two parts; the first three numbers are called the *individual numbers* I_1, I_2, and I_3, and the last two are *control numbers* C_1 and C_2. The Individual numbers are given sequentially to everyone who shares the same day of birth. The third number, I_3, indicates gender; even numbers for women and odd for men. Each of the two control digits follows a set formula based on a combination of the first 9 digits in the NIN:

$$C_1 = (8X_1 + 4X_2 + 5X_3 + 10X_4 + 3X_5 + 2X_6 + 7I_1 + 6I_2 + 9I_3) \mod 11$$

$$C_2 = (X_1 + 10X_2 + 9X_3 + 0X_4 + 9X_5 + X_6 + 3I_1 + 6I_2 + I_3) \mod 11$$

C_1 and C_2 both have to be one-digit numbers. As a result, if either C_1 or C_2 becomes 10, then it is invalid, and we try again with the next possible combination of individual numbers.

The Norwegian Government proposed a new way of making the NINs in 2017. The reason for this proposition is that Norway will not have enough NINs for all citizens around the year 2040. As of 2018, the proposal is still listed as "Under behandling", meaning that no final conclusion has been reached [19]. The Department of Finance proposed to use this new system from the year 2032, but with a possible later introduction in 2036. Several suggestions have been made as to how the new NIN should be structured. The one proposed by the Department of Finance will remove the gender identifier, and change the algorithm used to calculate one of the control digits. All existing NINs will remain as they are.

2.3.2 ID-porten

The Norwegian Agency for Public Management and eGovernment has created a common login system that can be used to log into Norwegian public services. "With ID-porten you can log in to more than 1000 different services from Government agencies." [20]. ID-porten has five types of electronic ID: MinID, BankID, BankID on Mobile, Buypass, and Commfides. These login methods are divided according to their level of security.

Medium-High Level MinID is the only service that has a medium-high security level. The user logs in by entering their NIN and a 2FA consisting of a static password and an OTP sent via SMS, or from a PIN code letter. This level gives access to *Samordna Opptak*, a service that allows Norwegian students to apply for higher education, and *Statens Lånekasse*, where the students can apply for student loans. MinID can also be used to log in to www.nav.no, which is administered by the Norwegian Labor and Welfare Administration.

High Level The services BankID, BankID on Mobile, Buypass, and Commfides give electronic ID authentication with a high level of security. This service also uses the same 2FA as in the medium-high level, with the added security that the OTP token can only be obtained after a physical meeting with the correct authorities has taken place. This gives a much lower risk of the OTP device falling into the wrong hands. High-level security is needed to access online banking services, to sign public documents, and other services that process a lot of sensitive personal information, such as health journals.

2.4 Authentication in Norwegian Online Banks

2.4.1 Historical View

The requirements for authentication in the early days of online banking were not very strict. One method used for registering new users in 2003, required the user to enter their name and NIN. A lookup was made in the National Registry on that person's registered home address, and a static pin of four digits was sent by registered mail. The problem with this model was that it was susceptible to brute force attacks; with only $10^4 = 10.000$ pin codes, and a chosen group of possible NINs, it was possible for an attacker to keep trying for a match for the selected users until a match was found. Since Norwegian NINs follows a strict pattern, it is possible to generate a group of probable NINs to test. Hole, Moen, and Tjøstheim describe, in a case study on online banking security from 2006, how such an attack can be done [21]. The NINs could also be collected in other ways since the NINs are not regarded as sensitive information. Unfortunately, some organizations and companies treat NINs as passwords, so that knowing someone's NIN is enough for authenticating as that user.

2FA was introduced later on, in the form of OTPs sent by post or SMS. This increased the security, but still allowed for directed attacks toward a specific user if someone had already found a working combination of NIN and pin. The technology to intercept an SMS was available, and if a criminal were to order a pin, then they would know when it would arrive in the victim's mailbox, and could then simply go pick it up.

This was made easier by exploiting other Norwegian services such as "Statens Lånekasse". They, in an attempt to make their web-pages more user-friendly, allowed auto-completion of a lot of personal information by entering a NIN. That means that if a criminal had found a matching pair of NIN and pin, then they could very easily get access to that person's name, address, and other personal information. Obtaining this also allowed the criminal to try and change the registered information by calling the bank, pretending to be the correct user, and ask them to change the registered mobile number or home address. The use of social engineering could give the criminal a better chance of correctly answering any control questions.

2.4.2 Today

We will now present how a user can become a customer of a Norwegian bank, and what methods they have for logging into their account. Our main focus will be on how this is done in Sbanken, as this is the bank that we have worked with.

Registration as a New User A BankID, or BankID on Mobile, is usually needed to become a customer in a Norwegian online bank. Some Norwegian banks allow a person to become a customer without having a BankID, but this is a multi-step process that involves filling out a detailed customer profile. The bank might request even more information and documentation if they see this as necessary as a countermeasure against money laundering.

A BankID can only be obtained through a Norwegian bank. They are unique and personal IDs that can be obtained by anyone who fulfills these criteria: they have to be older than a set age limit, have a Norwegian NIN, a residence registered in the Norwegian National Registry, and an e-mail address and phone number on which they can be reached.

The age limit varies greatly between the different banks. Several banks have an age limit as low as 13 years. Most Norwegian banks are offering a BankID from age 15, which is the age of liability in Norway. The age of majority in Norway is 18, and all minors have to be accompanied by a guardian that can give their consent.

As the very first step of authentication, a physical meeting where the customer presents their passport, or other approved documentation, is needed. If someone already has a bankID from one bank, this will work in all other Norwegian banks as well. Of course, some security mechanisms will be run in the background, as a way of combating fraud.

Sbanken, as a bank without a physical visitor's office, has a more roundabout way for their users to obtain an OTP device; a registered mail is sent to a post office close to that person's registered home address, so that the OTP device can be retrieved by presenting their passport to the post office worker. Sbanken only allows new users to register by using a BankID, BankID on Mobile, or authenticating themselves by passport.

Login Banks today offer multiple ways of logging in to the bank, in order to make it simpler for the user to choose the method that they find most practical. Sbanken offers five methods for logging in from a web browser, as well as the possibility of logging into the Sbanken-app on a smartphone. These methods are listed in Table 2.2, together with what types of authentication they require.

BankID uses NIN, a personal static password, and an OTP token from another bank, e.g. a code generator from DNB.

A **Code Card** is the code device that is used by Sbanken. The card is a scratch card with several codes hidden on it. When logging in, the user will scratch a specific field on the card, which is indicated by a number provided on the web page. When logging in, the user needs their NIN, a static password, and the code from the code card.

SMS can be used to receive the OTP. Otherwise, this will work in the same way as receiving the OTP from a code card or code device from another bank. Unfortunately, as we saw in Chapter 2.2.4, using SMS to receive codes makes one vulnerable to SIM swap attacks.

A **QRcode** can be viewed in the web browser when the user has entered their NIN. This can then be scanned with the Sbanken app, to do authentication in the app with a 4–8 digit pin code. The user is then successfully logged in, in the web browser. An extra authentication level can be added if the user protects their phone by FaceID or TouchID, but it is up to each individual user to do so.

BankID on Mobile lets the BankID (NIN and static password) be securely stored on the user's SIM card, so that the user only needs to enter a self-made 4–8 digit code each time they want to log in. This code might be replaced by a biometric scan on the phone, such as fingerprint scanning. The BankID is saved on a specific SIM card, which means that this method is not susceptible to SIM swap attacks, as changing the SIM card requires activation of the new SIM card in the online bank.

In order to activate BankID on Mobile, the user first has to log into the bank using a Code Card. If they do not have a code card, then this needs to be ordered from inside the web bank. This can be done by logging in using a BankID from another bank.

When installing the **Sbanken app**, it also becomes necessary to authenticate oneself the first time using BankID. After that, the user creates a 4–8 digit code which they can use to access the app, as well as having the option of adding access via either FaceID or TouchID.

Method	Know	Have	Is	How
BankID	yes	yes	no	password, OTP-device, no option
Code Card	yes	yes	no	password, code card OTP, no option
SMS	yes	yes	maybe	password, phone, FaceID/TouchID
QRcode	yes	yes	maybe	password, phone, FaceID/TouchID
BankID Mobile	yes	yes	no	password, phone, FaceID/TouchID
App	yes	yes	maybe	password, phone, FaceID/TouchID

Table 2.2: Factors of authentication for the different types of login options used in Sbanken. Some of the methods can be set up to use biometric authentication, e.g. to unlock the phone. This indicates that we could achieve 3 factors authentication, if not for the fact that all of the biometric checks may default to requiring static passwords instead, meaning that we only have 2FA. FaceID and TouchID represent types of biometric scanning.

Chapter 3

Identity Theft

We will now present the basic categories that identity theft falls into, before explaining how identity theft can occur. The chapter will end with a look at how identity theft can be detected, how it is punished, and consequences for the victims and society as a whole.

3.1 Types of Identity Theft

Identity theft can be divided into categories based on how the criminal obtains the account and the relationship between the victim and the criminal. These categories are illustrated in Figure 3.1

Existing Account Takeover This category contains cases where the criminal takes over and uses an already existing account. This means that someone has legally registered as a user of some service—they might even have used it for many years—then somebody obtains their login credentials, or in any way gets control over the account, and manages to successfully authenticate themselves as the correct owner of that account.

New Account Fraud The creation of a new account by using personal information that does not belong to oneself, is in itself an identity theft. Usage of this new account, to obtain either products or services, is classified as a new account fraud.

Synthetic Identity Theft This can be seen as a sub-category of new account fraud, where the process of making a new account is done by fabricating some of the information or stealing information from more than one person. This has become a fast-growing problem in

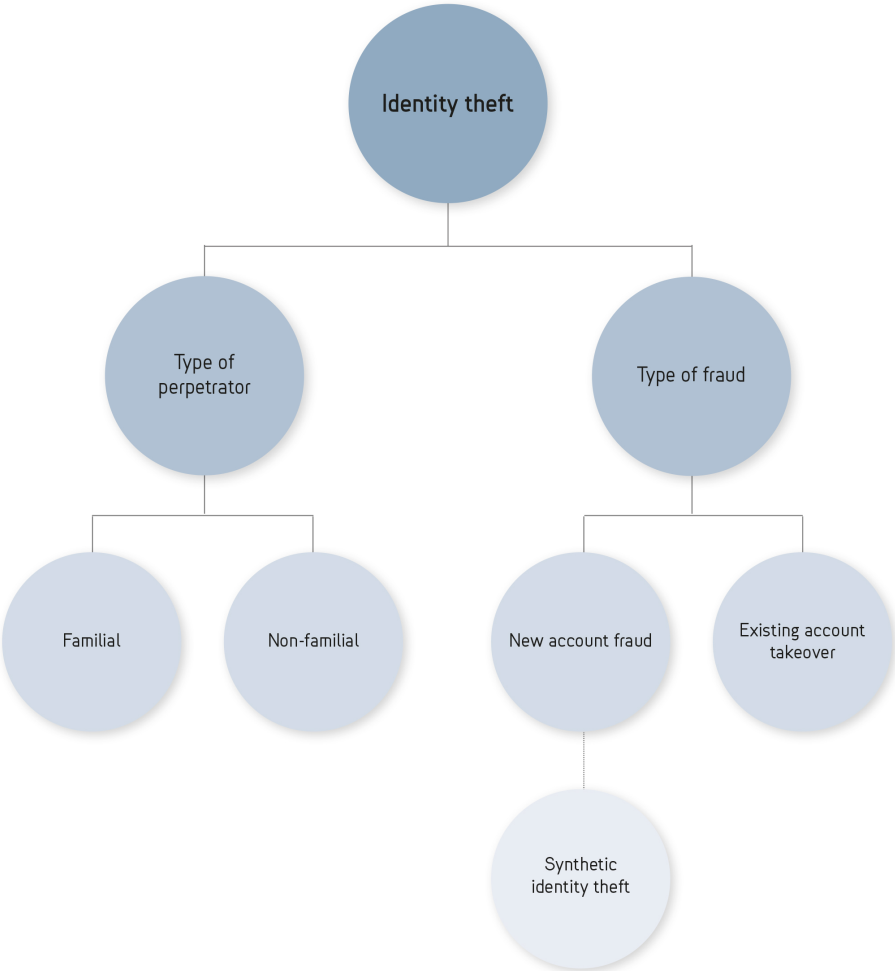


Figure 3.1: Types of identity theft

countries such as the U.S., where the criminal will combine a stolen social security number (SSN), and a fake name, to make a fictitious identity [22].

Since the culprit can use an address belonging to another random person, or a post box they have access to, there is a much greater chance of the victim never realizing that someone has stolen their SSN. As long as the proper security checks are in place, this should not happen in Norway, since the Norwegian National Register creates a link between a person's NIN and their name.

In 2013, NAV uncovered 74 fake identities in the National register [23]. These fake identities were of non-existing children and had been used by their registered "parents" to gain over one million NOK in welfare benefits. When so many important systems depend on getting their data from the National Register, it becomes especially important that we can rely on the data being real, and up to date.

A total renovation and modernization of the National Register is currently in progress. It was begun in 2016 and is scheduled to be completed in 2020. The government granted a sum of 536 million for this project in 2014 [24].

Familial Identity Theft This category consists of cases where the victim of identity theft has a strong personal connection to the perpetrator of the crime. The perpetrator in this type of identity theft is a parent, sibling, child, guardian, or a partner. The criminal usually knows the victim's personal information and NIN, and has easy access to their OTP-device. The crime becomes possible when more weight is placed on the amount of trust we have in the people close to us, instead of considering the consequences if this trust is misplaced.

This type of identity theft can be especially trying for the victim; in addition to being the injured party, they often have to take into consideration the consequences for the criminal. There is a high probability of there being a considerable amount of unreported crimes in this category. We will come back to this later.

Non-familial Identity Theft This entails all cases that do not fall under the definition of familial identity theft, which means that it will also include all cases where the perpetrator cannot be identified.

Child Identity Theft This type of identity theft differs from the ones mentioned so far, as it identifies identity theft by the type of victim, as opposed to the type of perpetrator. Child identity theft is a growing problem in countries where the NIN is not clearly connected to one person of known age. In the US, anyone with access to the child's NIN can use this to establish a line of credit, or take on other loans. Since the ID is stolen from a child, who naturally do not regularly monitor their credit score, the crime can go undetected for many years.

In Norway, it is illegal to take up loans before the age of 18. Since Norwegian NINs contain the year of birth of the person it belongs to, as well as being specifically connected to that person in the National Register, it is easy to check that a person is old enough to take up a line of credit.

3.2 Attack Vectors

The presence of identity fraud means that authentication has failed, and the criminal has gained access to services in another person's name. We saw in Chapter 2 how important it is to use secure authentication, and listed a few possible attack vectors that are used against different authentication methods. A heavy responsibility lies on keeping all user information private and secure, both on the side of the company or organization that offers a service, and on the user that uses it. There are many ways for the criminal to obtain the needed information to commit identity theft, too many to list in this thesis, but we will now briefly explain some of the more common attacks.

3.2.1 Phishing and Social Engineering

Phishing attacks are often done on a massive scale, using very basic techniques, in the hope that at least a few individuals will fall for it. It can e.g. be used to trick the customers of a bank to log into a fictitious website that seems identical to the website of their bank, thus giving the criminal instant access to the user's login credentials.

The users have a responsibility for evaluating what information they give away online, and make sure that the pages they use are the correct ones. A case handled by the Norwegian Financial Services Complaints Board (FinKN), stated that being the victim of email phishing is considered as gross negligence on the part of the user [25]. We will look more closely at the laws followed in this case later in this chapter.

Recent years have seen an increase in more sophisticated attacks aimed at a selected handful of individuals. Such *spear-phishing attacks* tend to follow a pattern of gathering information about the company or single individual, developing a relationship, exploiting any known vulnerabilities, and then executing their attack.

This targeted manipulation is known as *social engineering* or *people hacking*, because it exploits the human aspect of authentication. The goal is to trick the victim into giving away sensitive information, or find information about a single individual, in order to take over their bank account. Nowadays, many people share a considerable amount of personal information about themselves and their company online, making it easier for an attacker to understand whom they can target, and how to best do so.

People need to be aware of whom they share their information with, and what it can be used for. It is important to ask questions before disclosing personal information. If someone receives a strange call or email from e.g. their bank, then it is important to verify that this is truly from the bank, before giving out information. As seen from the case handled by the FinKN, it is the users' responsibility to make sure that they are communicating with the correct entity.

Many businesses use *control questions* in order to verify the identity of their customers. Back in the days before social media sites, this could be a decent way of limiting the chance of someone misusing the system. Today, when information such as a home address, your mother's maiden name, teacher's name, or favorite food, is something everyone with internet access can find on services such as Facebook or in public records, this becomes dangerous. With some research prior to the attack, the criminal can easily respond to such control questions.

3.2.2 Theft

A stranger might *steal* your wallet or purse, and obtain ID cards, credit cards, and other personal information. Someone planning to steal your ID might *monitor your mailbox* to snatch up a bank statement, an incoming OTP device, or other important information regarding your person or finances. Dumpster diving, the act of physically looking through trash cans not belonging to oneself, is another method for getting access to personal information. When the pilferer is someone with access to your house, it becomes important to keep bank statements, credit cards, and passports, in a secure location.

If one uses a mobile phone for banking purposes, then it becomes especially important to secure it as well as possible. There are many ways of doing this, such as having a strong password-lock on the screen and setting up a SIM-lock. Biometric authentication, as discussed in Chapter 2, can make every-day use of the phone simpler for the user. This type of authentication will also prevent *shoulder-surfing* if a user authenticates themselves using their phone in public. Shoulder-surfing is when someone standing close to you can "lean over your shoulder" to see what passwords you type in. Since the user does not have to enter their actual passwords as often as without biometric authentication, it also means that they can utilize longer passwords. This will give increased security.

It is important to properly secure all our digital devices if they are used to log into services such as a bank or an email account. If the device gets stolen, and it is not password protected or in any other way securely locked, then the thief will get instant access to all accounts that are currently having an active session on that device. For example, if the user has signed into an email app on their phone, and the phone gets stolen, then the thief may use this account to send "forgotten password"-requests to multiple services, and in that way get access to the user's other accounts.

3.2.3 Attacks Against Companies and Organizations

It is imperative that the entire communication between the user and the service is end-to-end encrypted, and that the customer information is stored in a secure way by the company. Data breaches have happened before and will happen in the future. There is a huge market on the dark web for stolen information such as NINs, healthcare information and username-password combinations. The stolen data could in itself be part of a multi-stage attack, or the criminal might put it up for sale to the highest bidder. Employees in a company can be targeted in the same way as any other individual, in order to find a way for the criminal to enter the system. There is also the possibility of an insider in the business willfully helping the attacker.

Far-Reaching Consequences In 2015, The U.S. Internal Revenue Service (IRS) became the target in a massive identity theft fraud. They have admitted that at least 724.000 American citizens had their tax data stolen via a web page hosted by the IRS [26]. The attack was possible due to weak authentication. What happened was this:

A service launched by the IRS in 2014 allowed taxpayers to register themselves online in order to see their transcripts from previous years. The IRS used what is known as Knowledge Based Authentication (KBA). The KBA used by the IRS was a collection of multiple-choice questions about the person that tried to log in. These questions were things that the IRS already knew the answers to, such as a previous address, birth dates, loan amounts etc. Common for most of this information was that it was possible to find it on social media sites, on Google maps, or even buy it online.

It is noteworthy that a large healthcare organization, Primera Blue Cross, was the victim of a large-scale attack prior to the exploitation of the IRS. The attack gave the criminals access to information on more than 11 million users, which indicates that they got massive amounts of information, including names, dates of birth, addresses, and financial information such as bank account numbers [27].

There have been many reports from American citizens that when they apply for tax returns from the IRS, this has already been done—someone else has taken their money. The thieves had re-submitted the victim's old refund-report, in the hope that the user they were impersonating would get a return on their tax that year as well. Of course, the thief would then take all the money for themselves.

Prioritizing Security The possible cost of a leak or successful attack can have massive ramifications. Security is like an insurance; it might seem like wasted money when everything is going smoothly, but when an accident or attack occurs, you would not like to stand there without it. The case actually might be even worse for security; good security is something that is not often noticed, since everything is working as it should, making it seem like it is not necessary.

It is even worse if the compromised company fears the loss of customers due to lack of credibility and because of this, *denies that an attack has ever taken place.* This will cause users to remain unaware that someone might have stolen their information. By not sharing the facts about the attack, then a security flaw that could also exist on other systems will most likely go undetected for a long time. This gives the criminals the opportunity to attack these other systems as well.

Every other year, the Norwegian Business and Industry Security Council publish a report titled the "Norwegian Computer and Data Breach Survey" [28]. They remark that it is important to get arenas where companies can share their knowledge and experiences. Their report from 2018 is their 11th such report, and has data collected from over 1500 public and private companies. 13% of the companies said they had detected attempted data-breaches, and 3% said that such breaches had taken place. 18% said that their company had encountered "phishing or other social engineering attacks". This last category had increased significantly from 2016, when it was only 8% [28, p.14-15].

67% report that the reason for the experienced security breach was "chance or bad luck", "human error" was 55%, and only 19% said that they had "Inadequate prioritization of security efforts" [28, p.20]. Interestingly, 40% report that the incidents were discovered by pure chance. This should not be the case if over 80% of the companies actually have a good prioritization of security.

3.2.4 Misplaced Trust

The most common reason why identity theft happens, is that the victim has misplaced trust in a company or person. This can either be that they trusted a scammer, the company that they trusted was not careful with their information, or that they had a personal relationship with someone who betrayed their trust. Misplacement of trust can be tied together with social engineering, where a criminal will try their utmost to get the victim to trust them. Misplaced trust is also the case in most instances of familial identity theft, or for someone else whom the user trusts enough to let them get entrance to their home or personal devices, such as PC or smartphone. That person could e.g. be a health care worker, cleaning assistant, friend, neighbor, colleague, or someone helping them with a technical problem.

3.3 Consequences and Responsibility

A paper from 2017 by Javelin Strategy Research, states that there were 15.4 million U.S. identity theft victims in 2016 [29] . This amounted to an increase of 16% from 2015.

Crime statistics reported by Statistics Norway also show an increase in reported acts of identity infringement, as can be seen in Figure 3.2. A report done by NorSIS and Skatteetaten about identity theft in 2016 found that 89% of the Norwegian population said that they would report identity theft to the police [30, p.37]. In the same year, only 32% of the victims of identity theft had reported the crime to the police, and numbers from 2017 showed that this number had sunk to only 27% [8, p.6]. This indicates that there might be a large number of unreported crimes.

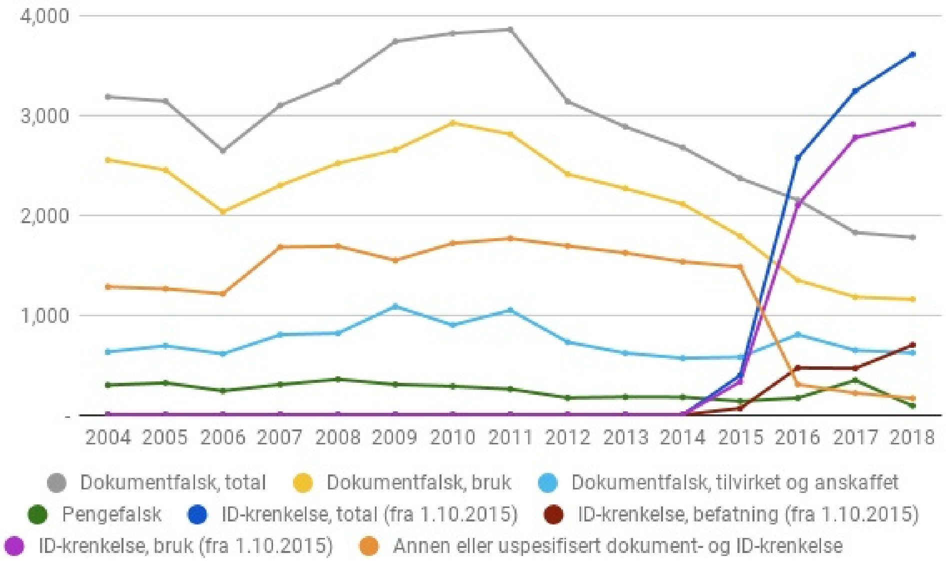


Figure 3.2: Graphical representation of data on reported crimes in Norway in the years 2004 to 2018, based on numbers from Statistics Norway [31]. ID infringement was not part of the statistics before the year 2015.

3.3.1 The Criminal

Anyone who by illegal means gains possession of someone else's identity—and uses said identity to gain an illicit economic advantage, or in any way cause a loss to someone else—can get a prison sentence of up to two years, in accordance with The Norwegian Penal Code of 2005, section 202. This section was added to the Penal Code in June 2009 (see Appendix A.2).

However, the new Penal Code of 2005 was not entered into force until October 2015. In the meantime, in December 2010, a "copy" of Section 202 was added to the old Penal Code from 1902, under Section 190a. The relatively fast introduction of this section can be seen as a testament to the importance of plugging this gap in the laws dealing with identity theft. It also shows that this was of great importance to the Government/Parliament at that time.

The use of the stolen identity often causes breaches of other sections of the Penal Code. If the criminal takes up loans, then they also have committed fraud and can be punished after the Penal Code Section 371 and 372. When there is a large financial loss involved, then Section 372, dealing with aggravated fraud, can give a penalty of up to six years in prison.

A problem when dealing with the consequence of identity theft comes from the fact that very few criminals actually get charged with a crime. The police can follow digital trails to figure out where and when the account was logged into when the identity theft took place, but it can be much harder to connect this information to a specific individual. A case from 2017 was dropped by the police, even though they had found the computer used to commit the identity fraud [32].

When the chance of discovery and punishment is low, and the possible gains can be in the millions, it is not that surprising that identity theft is still a growing problem in society today.

3.3.2 The Victim

By default, the bank is responsible for covering any loss caused by unauthorized payment transactions, such as in credit card fraud. The Norwegian law for Financial Agreements Section 35, states some exceptions to this rule.

The customer is held responsible for the entire loss if they, by gross negligence, have broken their obligations described in Section 34 (See Appendix A.1). If the transaction was done using an electronic payment instrument, then the customer is only held responsible for up to 12.000 NOK. This is unless they have deliberately failed to follow their obligations given in Section 34. If that is the case, then the customer becomes responsible for the entire sum, even when electronic payment instruments have been used.

The first such obligation described in Section 34, is that the customer must follow the terms of use set by the bank that issued the BankID and code device. A common rule for all these banks, is that the user must keep their passwords private. Knowingly sharing your passwords with someone else, e.g. a partner, is seen as breaking this agreement, and will make the user responsible for the entire sum.

Large Sums Involved The sum in question can quickly grow massive, even if one can detect the theft almost instantly. A case from 2018 shows how a criminal managed to get 1.2 million NOK in loans during a single week [33]. The thief had stolen a bank card from the victim's mailbox, and no passwords or other information were needed for the criminal to get the loans. The victim said to the newspaper VG that the criminal had applied for loans in eight banks, and had been granted loans in four of them. The case is still ongoing, so it is unknown whether the victim or the bank will be held accountable for this sum.

Hard to Detect The victim that noticed the identity fraud after only a single week was very lucky; a car seller had called him to confirm the deal that the criminal had made in the victim's name. If this had not happened, then the victim would probably not notice a single thing before the creditors started demanding to be paid back.

Another case has a 22-year old woman as the victim when her mother took up loans valued at approximately 80.000 NOK [34]. The daughter was judged as responsible for paying back the entire sum to the banks. When the girl was only 16 years old, her mother had helped her create a BankID. Since the daughter had not changed her password afterwards, as this should be kept a secret, this was seen as a violation of the obligations set by the bank.

It becomes especially hard to detect the theft when it is a family member who steals your identity. Another victim had her boyfriend, whom she had lived together with for 13 years, take up 2,1 million NOK in loans in her name [35]. Since they were living together, it was easy for him to be the first one to check the mailbox and remove any incriminating mail. He has been convicted of identity theft and aggravated fraud, and she has been declared innocent. Still, this woman is now responsible for the entire debt. The reason for this is that she let her boyfriend use her phone, which he then used to take up 14 loans, using her BankID. She had not given him the password, but neither had she been actively hiding it from him when she logged into the bank.

Extensive Consequences Millions in debt, and betrayed by someone you trust; it is easy to see that this will affect the victims for the rest of their lives. If the criminal is not caught or stopped, then the problems may never stop, as new claims might keep on appearing. Thus, the victim will continuously have to prove their innocence. This can take a massive amount of time, as seen in a case of identity theft from 2014. The victim started to have his post stolen from the mailbox in 2010, including a bank card. The criminal was later found and brought to trial. Still, the victim spent three years of continuously having to battle incoming claims. He said in an interview with TV2 that he had not paid a single bill, so there was no monetary loss, but that he had used an "unbelievable amount of time" on this case [36]. He also said that he had spent almost every day after school to get an overview of all incoming claims from creditors. In this case, the criminal was caught, but if they never get stopped, or the victim's identity gets sold on the web, then the problems may keep popping up again for the rest of their lives.

3.3.3 The Norwegian Society

The consequences that identity theft have on society as a whole, are quite complicated. This sub-chapter will touch upon a few of these aspects.

Each victim of identity theft will most likely struggle with the ramifications of the theft for a long time. Their view of society might be drastically changed as a result of their experience, especially if they did not receive the help that they expected to get. The feelings and opinions that the victim is left with, can very likely spread—like a ripple effect—to everyone around them. NorSIS published another survey in 2018 about Norwegians and their knowledge and habits around digital security. One finding here was that as much as 39% did not trust the police to help them if they were the victim of cybercrime [37, p.21]. 47% were completely or partially sure that the police would help them, and 14% did not know. This lack of trust in the police being able to help victims of digital crime, can become a serious problem in a society where digitalization is as prominent as it now is in Norway.

Surveillance All Norwegian banks are required to do their part in stopping criminal activities such as money-laundering or terror financing. This means that extra security checks must be in place, e.g. checking that the recipient in a transaction is not black-listed. More information will lead to higher confidence in the checks. The problems emerge when the authorities try to set a line between what is accepted surveillance for protecting their citizens, and what becomes a breach of privacy. The Norwegian society is heavily centered around trust. Institutions such as banks have a massive amount of data on all their users. It becomes very important to have secure, ethical guidelines when it comes to the storage and the processing of this information, so that the customers can continue using these services without fear of their privacy being compromised in any way.

3.4 The Situation Today

Identity theft and fraud can be tricky to detect. The very obvious cases, such as someone transferring their entire life savings into a foreign, black-listed bank account, are instantly stopped, and a warning or a request for confirmation is sent to the owner of the account. Sometimes, the actual owner of the account is in full control and wants to proceed with the transaction. This has been observed multiple times when it comes to dating fraud, where the "victim" ignores all warnings from the bank, even when the bank says that the recipient of the money is a known fraudster that has tricked others before. It is the owner's right to manage their money in whatever way they want. The exception is if the bank suspects that the transaction is part of a money laundering scheme, or in any way used to finance terrorism. If that is the case, then the bank will stop the transaction regardless of the accountholder's wishes, and turn the case over to the correct authorities for further investigation.

The normal course of events when it comes to identity theft, is that the victim in some way becomes aware of the theft, and then contacts the bank. If the victim has an active account in that bank, then the bank will ask the victim to report the crime to the police. If this does not happen, then the case is closed, and the victim is responsible for any outstanding amounts. If the bank receives a report of the crime, they will start an investigation to detect what has happened, and how much money has been lost. This can take a considerable amount of time and resources.

This was the situation when we started to work on this thesis.

Chapter 4

Case Study

4.1 Sbanken

This thesis is written in collaboration with Sbanken, a purely digital Norwegian bank formerly known as "Skandiabanken". It was launched in April 2000 as the Norwegian branch of the Swedish Skandiabanken AB. The Norwegian Skandiabanken became a separate company with stock market launch in 2015, and changed its name to Sbanken in August 2017.

Sbanken has given us access to some of their customer data, with the objective of looking for patterns that could be used to better detect and stop identity theft in the bank. We wanted to take a closer look at what information they already had, to see if it could be of increased use. In cases where the data was not enough to detect identity theft, then we would look at potential ways of collecting this information, or create other methods that could help them do this in the future.

4.2 Research Questions

The main goal of this thesis is to look at the available information regarding the users of online banking and their transactions, and use this to find a pattern that can help us identify identity theft. This formed the main hypothesis that has guided this thesis. The hypothesis was first mentioned in the introduction, but for convenience' sake, we will repeat it now:

Hypothesis 1. *It is possible to detect identity theft to a greater extent by looking at the metadata of each user account and its transactions.*

To help us research whether Hypothesis 1 was correct or not, several other hypotheses were made as well. We will now introduce all the subsequent hypotheses that have guided this thesis. This will include an explanation of the process we went through in making these hypotheses, as well as how we checked them in order to see if they could be confirmed or disproved.

Four Types

Since identity theft is such a complex problem, one of the first things we looked at was if we could break our cases into different categories to make them easier to analyze. We have focused on the four main categories of identity theft mentioned in Chapter 3; new account fraud, takeover of an existing account, familial, and non-familial identity theft. These four categories were combined into four types, and our hypothesis was that these types would be recognizable in our set of identity theft cases.

Hypothesis 2. *Identity theft can be divided into four types, which will have different characteristics and means of detection.*

	familial	*non-familial*
existing account	*Type A1*	*Type A2*
new account	*Type B1*	*Type B2*

This would mean that there is a way to look at the characteristics of an individual case, and test if it matched some patterns indicating the presence of any of these four types of identity theft. Our belief in this hypothesis stemmed from some simple observations of how identity theft takes place, e.g. that a stranger and a close family member would have access to different information. The continued relationship between a criminal family member and the victim was believed to have some impact on how the criminal used the account. Thus, it would most likely yield poor results if we were to try identifying identity theft as a consequence of one single type of criminal activity, as opposed to four different patterns.

New Account Fraud

In order to create a new account in the victim's name, the criminal needs access to either an OTP-device that is already registered on that user in another bank, or credentials that can be used to retrieve said OTP-device from a bank. We theorized that there was a high probability of there being one type of account creation that was used more often. Some banks allow users to take up loans without secure authentication, but as this is not the case in Sbanken, we had no way of looking closer at that type of identity theft.

Hypothesis 3. *There is a preferred registration method used by fraudsters in the illegal creation of a new account.*

Existing Account Takeover

In the same way that we wanted to look at what methods were used in new account fraud, we believed that we could detect some pattern in the ways that the criminals conducted a takeover of an existing account. We also believed that by looking at the existing use pattern of the victim, then we could detect any changes in this pattern, which would give an indication that identity theft had taken place. This led to two new hypotheses:

Hypothesis 4. *There is a preferred login method used by fraudsters when they take over an existing account.*

Hypothesis 5. *An existing account that has been taken over will show some deviation from the normal use-pattern.*

Personal Contact Information Regarding the Account

Personal contact information, such as a phone number or email address, can be used as extra steps in authentication or account recovery. It can also be used as an extra security step by sending the user information regarding their account, such as sending an email telling the user that they have logged in from a new location. Thus, a criminal would often want to change this information, both to give himself access to other services, but also to keep the true user unaware of the theft for as long as possible. We came up with two hypotheses; one dealing with the sudden change of personal information, and the other dealing with reuse of information, such as 2 accounts both having the same phone number, even though the accounts belong to two different persons.

Hypothesis 6. *Changes to personal information, especially if several changes have taken place at the same time, is an indication of identity theft.*

Hypothesis 7. *Reuse of personal information from other accounts in the bank, such as a phone number or email address, is an indication of identity theft.*

Credits and Loans

An identity theft is often not noticed until someone actually steals some money, either by transferring savings from an existing account, or using the account to take up loans. We came up with three new hypotheses regarding the actions of the user around the time that they take up a loan, or execute a large transaction.

Hypothesis 8. *Requesting a loan without first having visited the financing-web, is an indication of identity theft.*

Hypothesis 9. *There is an increased likelihood of identity theft if the first thing the customer does after receiving the loan, is to send the money out of the account.*

Hypothesis 10. *Identity theft cases have an abnormally high amount of unsecured credits.*

4.3 Methods

4.3.1 Testing the Hypotheses

The different research questions introduced in Chapter 4.2 resulted from looking at how identity theft could be detected by the bank. This led us to look at how these hypotheses would come into play, and how the bank could measure the results. Figure 4.1 visualizes in what setting we wanted to test each hypothesis.

Our work has mostly been divided into two main areas of research:

- Identifying some characteristics of the existing identity theft victims.
- Finding ways that the bank can both detect and prevent future identity thefts.

First, we looked at the data that we had collected about all the reported identity theft cases. Then we looked more closely at how the log files were structured, and what data was available to us. This included figuring out what database parameters might prove useful to us in the future.

The rest of this chapter will be used to explain how we structured our research, why we made the choices that we did, and present our findings.

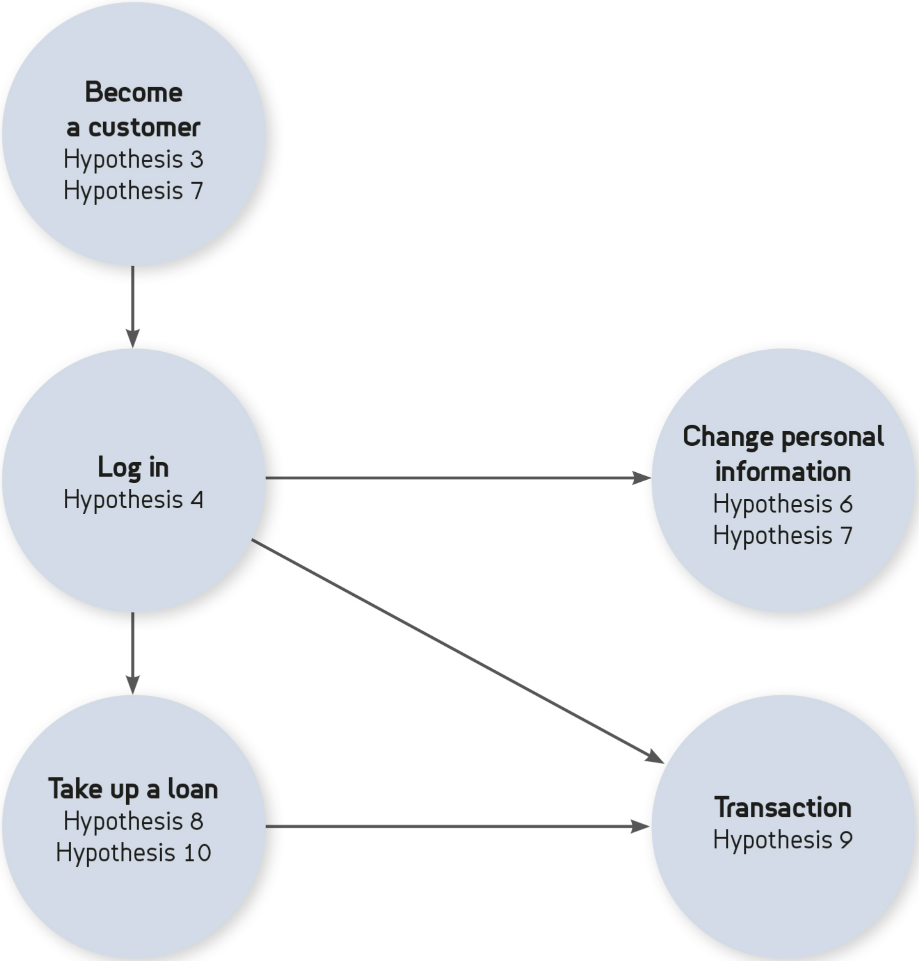


Figure 4.1: Connection between the hypotheses, and how we visualized them coming into play in a customer's user-activity. Hypothesis 2 and 5 are missing from the diagram. This is because they are more general, and do not have a single point of measure.

4.3.2 Limitations

Limitations of the identified cases This study is based on a set of reported identity theft cases. We have no way of knowing if this is all the identity thefts that have happened in the bank in this period, or how large a percentage they represent. In the same manner, we do not know if any missing cases will match any pattern of our reported cases, or if there might be another form of identity theft that we have yet to discover.

Another aspect is that all our cases are *claims* of identity theft. They may be valid, or they may be attempts at fraud. One example could be a user who does not want to take responsibility for their actions, claiming that someone else is to blame. We did not have access to information about the cases beyond their initial report, in addition to parts of their customer history. We did no attempt at filtering the cases, based on what we thought were "real" identity thefts, but used all the reported cases that were provided us.

Constraints Concerning Time and Privacy As we were not employed in the bank, there were some constraints on what we were allowed to see and do. The biggest hindrance was that in many instances, we could not control the systems ourselves, but had to sit next to someone who could "push the buttons" and show us the systems and log files that we needed to look at.

Luckily, several employees have helped us a great deal in finding the information we needed for our work. Still, a great amount of time was spent waiting.

4.4 Characteristics of an Identity Theft Victim

We started our work by collecting data about the known identity theft cases. The number of cases given to us will not be mentioned in this thesis, by request from Sbanken.

The information we extracted included how the victim had reported the case and when it was reported. This information was used to label every case as either familial or non-familial identity theft. Their age and gender were extracted from their NINs.

4.4.1 Gender

	Male	Female
Familial	14%	27%
Non-Familial	31%	28%

Table 4.1: Distribution of identity theft victims, based on gender and type of identity theft. Together, the four cells summate to 100%.

41% of the identity theft victims reported that the perpetrator was, or had been, a close family member. 10% reported that the offender was a neighbor or a close friend. These 10% have been regarded as non-familial, as they do not fulfill the criteria that we set for familial identity theft in Chapter 3.1.

55% of the victims are female, and the difference in the percentage of male and female victims of familial identity theft is particularly striking. It is important to note that we do not know the gender of the perpetrator. Still, we saw that several of the women that had reported the crime, also claimed that their husband had at some time taken full control of the account, or even created several accounts in their wife's name.

The distribution of male and female customers in Sbanken is actually almost a perfect match of the opposite of the distribution of all identity theft victims, meaning that even though they have less than 50% female customers, there are still more than 50% female identity theft victims. This indicates that the imbalance of male and female victims is actually much larger than what we first assumed.

When looking at familial identity theft, the balance is as bad as 1/3 male and 2/3 female victims. The balance is more even when looking at non-familial identity theft, and male victims of non-familial identity theft is our largest category, with 31% of all the cases. If we take into consideration the imbalance of male and female customers, then the number of non-familial identity thefts indicates an almost 50/50 distribution of male and female victims in this category.

4.4.2 Age

	Age
Mean	43
Median	41
Minimum	19
Maximum	88

Table 4.2: Age distribution of identity theft victims

As we can see from table 4.2, identity theft is something that affects people of all ages. The mean and median values almost perfectly mirror the values for all customers in Sbanken, but are slightly higher. This could be because the bank has customers below the age of 18, and we have found no victims below the age of 19. We cannot know for sure why it is so, but we see many possible reasons. A person below the age of 18 cannot take up loans, nor do they often have a lot of money for someone to steal. This makes them less attractive as victims of identity theft. It might also be because they do not realize that they are the victim of a crime, and that they might find out and report the crime when they become older.

4.4.3 Type of Identity Theft

We wanted to test if there are any differences between the four types of identity theft which we introduced in Hypothesis 2. Our findings on this can be seen below, in Table 4.3:

	Familial	Non-familial
New Account Fraud	29%	46%
Account Takeover	13%	12%

Table 4.3: Distribution of identity theft victims, based on Hypothesis 2.
Together, the four cells summate to 100%.

All acquaintances, such as friends or neighbors, are still marked as non-familial, in accordance with our definition of familial from Chapter 3.1.

The numbers shown in Table 4.3 can give us an indication of who commits what type of identity theft, but the actual number might be different than what we see here. Everyone who did not include information about who had committed the theft, has been labeled as non-familial. It might be that several of the users were victims of familial identity theft, whom for various reasons did not want to report this to the bank.

Almost 50% of the reported cases were a result of a non-familial new account fraud. Adding the roughly 30% of the victims reporting new account fraud done by a family member, makes the category of new account fraud significantly larger than account takeover.

Account takeover makes up 25% of all the reported cases, with a more even balance of familial and non-familial identity thefts. A significant part of the non-familial cases in this category consisted of someone the victim knew personally, but were not a close family member. The large portion of familial theft in regards to account takeover might indicate that the criminal's close relationship to the victim makes it easier for them to get access to passwords and OTPs, which are needed to take over an existing account.

4.4.4 Comparison With American Victims

Navarro and Higgins used statistics from a 2012 National Crime Victimization Survey, to explore the differences between victims of familial and non-familial identity theft [12]. The study was made by analyzing responses from American citizens, which naturally means that there are some clear distinctions between their findings, and what we observe in Norway. One of these differences is that they found evidence of child identity theft. Another American study, done by Javelin Strategy and Research in 2018, found that "More than one million children were victims of identity theft in 2017" [38]. In Norway, however, child identity theft appears to not be a problem. We found no such cases, and, as we mentioned in Chapter 3, it is illegal in Norway to take up loans before the age of 18. The banks can check the age of each person with a NIN, and can then easily refuse all who are under the age of 18.

Navarro and Higgins report that the misuse of existing accounts is the most prominent problem, with 53.4 % of the familial identity theft victims, and 56.5% of the non-familial identity theft victims reporting that they had experienced this. New account fraud, on the other hand was only on 25.0% for familial victims, and 18.3% for non-familial victims. These victims could in theory report that they had been the victim of both new account theft, existing account takeover, as well as a few other categories that are not relevant to us.

These numbers are in clear contrast to what we found here in Norway. We found that the creation of a new account covered 75% of our cases, making it far more prominent than account takeover.

This might simply be a result of the American study being based on data from 2012. It could also be the result of lacking use of 2FA in the US, making it much easier to take over an existing account, than what it is in Norway. We do not know for sure, but it seems like many American banks do not use 2FA at all. Some of them seem to have added the possibility of setting up 2FA during the last few years [39], but it is still not the default security setting. This, combined with security breaches and information dumps, as we briefly discussed in Chapter 3.2.3, can make it easier for the criminal to take over existing accounts, because they have enough information to successfully pretend to be the real user.

Still, we also see some clear similarities; women are considered more at risk than men, and we can see that identity theft happens to people of all ages. Their study also includes other factors, such as race, household income, and marital status. We have not included these factors in our study.

4.5 Measuring the User's Login Activity

To get an overview of whether the user is behaving abnormally or not, we started researching how a normal user behaves. We were allowed to receive some customer data on when the customers log into the bank, and what device they use to do so. We got access to roughly 4 million user entries, which had been processed to contain only relevant, anonymous data points. All measurements described in this section were made by analyzing this same set of user entries.

4.5.1 Software Stack

Python Both R and Python are good programming languages to use when it comes to statistical analysis and modeling. The choice of using Python for this task stems from two main reasons. We had some prior experience in using Python for machine learning, so it was seen as the most time-efficient alternative, as opposed to learning R from scratch. The other reason was that we wanted to see whether we could use machine learning to get even better results. The deciding factor was that we had already seen how easy it was to use existing Python libraries to implement machine learning, so this could easily be applied to test different solutions. We also wanted to get some more experience with using Python for statistical analysis, which was a definite point in its favor.

Data handling We used the Python libraries Pandas and Seaborn for data processing and to present the data visually.

Pandas is a data analysis library made for Python, and has some very practical data structures that make it easy to work with large amounts of data.

Seaborn is a Python library, based on Matplotlib, which can make data visualizations, such as heatmaps.

4.5.2 Measurements

Heatmaps The first thing we wanted to do was to get a good graphical representation of the data, to get a better understanding of how the typical customers in the bank were behaving. The data we had consisted of over 10 dimensions. In order to find some useful way of presenting it, we chose to select only a few dimensions at a time. The most useful dimension was the one representing login time. To show time in a useful way, we chose to split it into two dimensions; the day of the week, and time of each day. Using heatmaps became an obvious first choice for showing three-dimensional data in an intuitive way. A heatmap is a graph with two axes, the X-axis and the Y-axis. The third axis, the Z-axis, is represented as colors on a scale. For example, a heatmap that represents temperatures, can have a scale ranging from blue, for cold temperatures, to red, for warm temperatures. In our data, the color in an (x,y)-coordinate describes the number of logins that have happened in that time period.

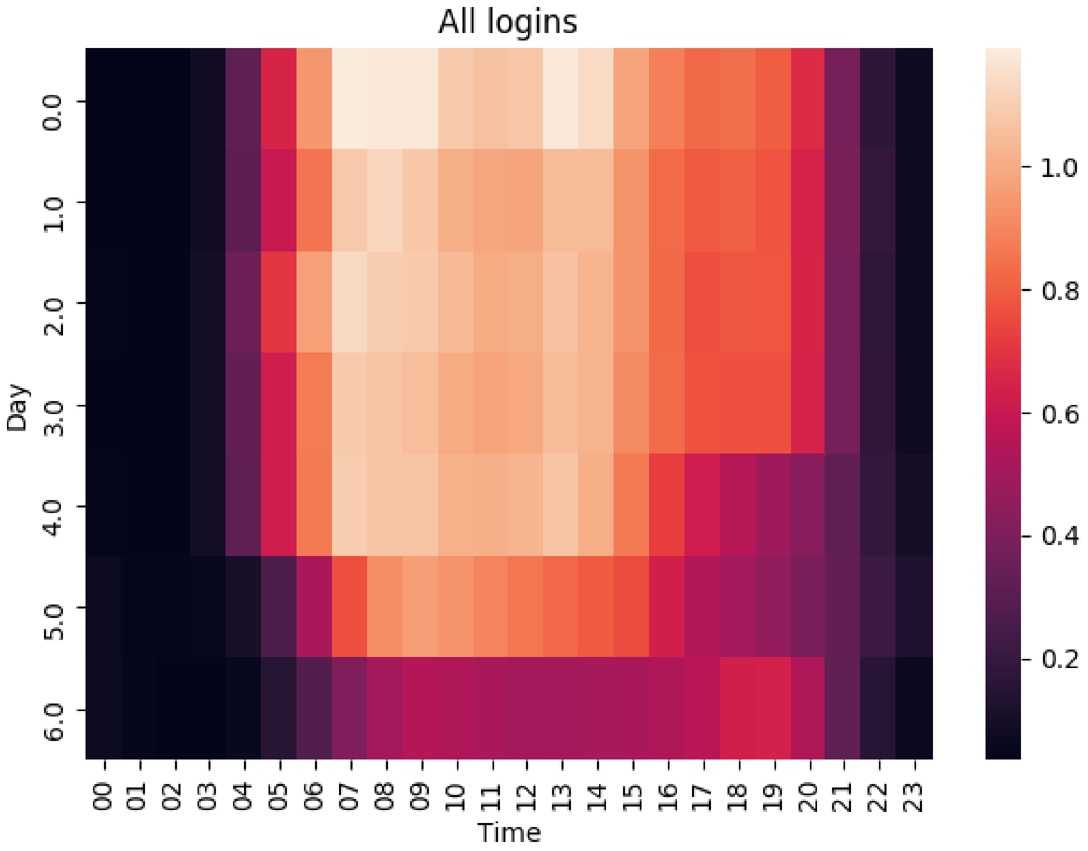


Figure 4.2: Heatmap showing the frequency of all customers logging into the bank. The value of each cell represents the percentage of logins that happened in that time-slot.

All heatmaps described in this chapter have the following pattern:

- X-axis: Time of day, beginning at 00:00, and ending at 23:59
- Y-axis: Day of week, with Monday=0, Tuesday=1 ... Sunday=6.
- Cells: All heatmaps, except for Figure 4.3, have cells that are color-coded according to the frequency of logins happening in that time-slot.

All the heatmaps described in this chapter have been normalized by percentage. This is done for two reasons. The first reason was to make it possible to compare the heatmaps with each other. The second reason was the need to hide the actual numbers of logins that happened in the bank, as this information did not need to become public knowledge.

The first heatmap, which can be seen in Figure 4.2, looks at all the customer activity. Since we had a massive amount of data, it was not very surprising to see that users logged in every hour of the day, every day of the week. The massive amounts of data that we had, made it ineffective to look at all the data as a collective set. Therefore, we looked at ways that we could split the data based on the different parameters that were available to us. We will now describe the results that were of interest for further work on this thesis.

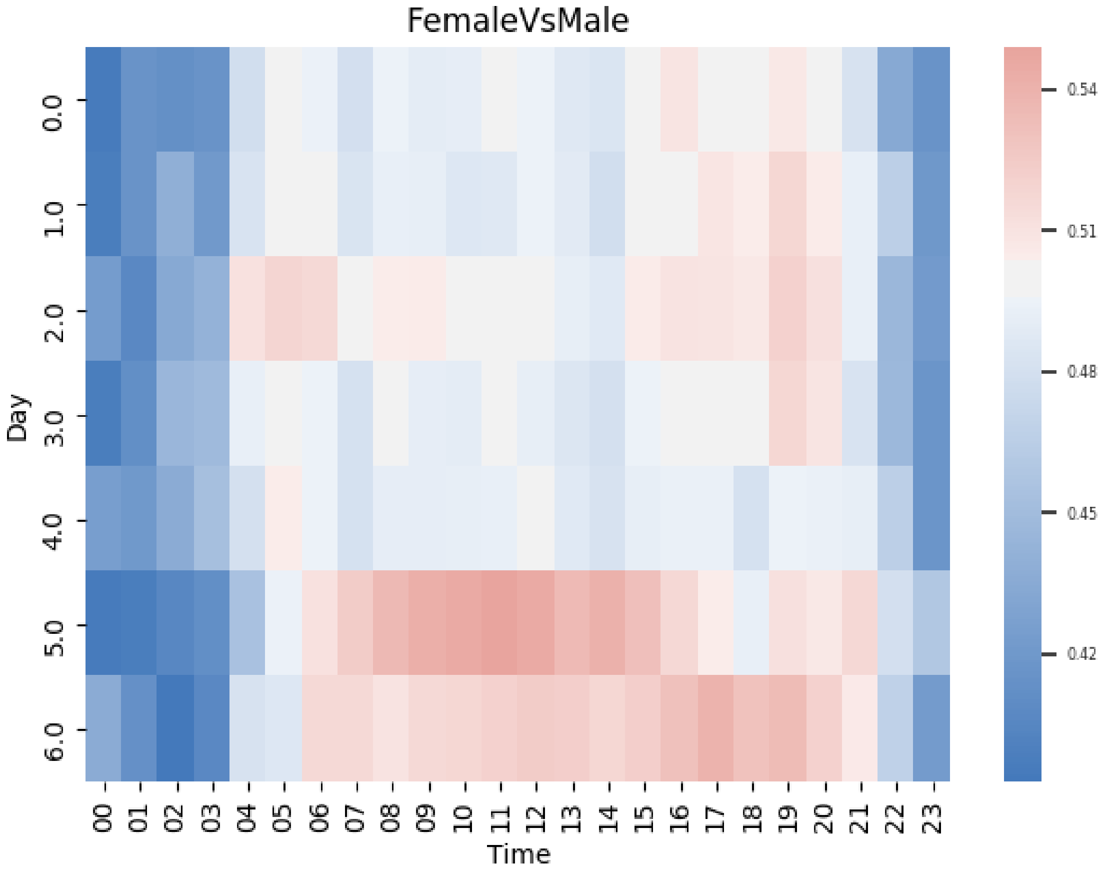


Figure 4.3: Heatmap showing the difference in frequency of female and male visitors of the bank. 0.5 is the center point, and is marked by the color white. White cells mean that there is an equal percentage of both genders visiting the bank. The more prominent colors, of either red or blue, show an increasing imbalance of the genders.

Gender Figure 4.3 shows the difference in the activity of male and female users of the bank. First, we extract the frequency of logins for each gender, so that each cell on the map represents the percentage of logins that happened in that specific time-frame. We then add these together, to see if there exists some time period where the genders have different use-patterns. Each cell is calculated by taking the percentage of female visitors to the bank, and dividing by the sum of the percentage of both female and male visitors.

- > 0.5: The percentage of female users logging into the bank is greater than that of the male users.
- $= 0.5$: There is an equilibrium of female and male users logging in at this time. 50% male, and 50% female customers logging in at this time.
- < 0.5: The percentage of male users logging into the bank is greater than that of the female users.

For example, if a cell has a frequency of 1.2% of the female users logging in, and 0.9% of the males, then the score shown in the cell will be $\frac{1.2}{1.2+0.9} = 0.57 =$ red.

There is a significant difference between the genders. During working hours, the changes are relatively small, but still significant. Logins by male customers are much more prominent during night time, and female customers have a greater amount of their logins happening during Saturday morning and early afternoon.

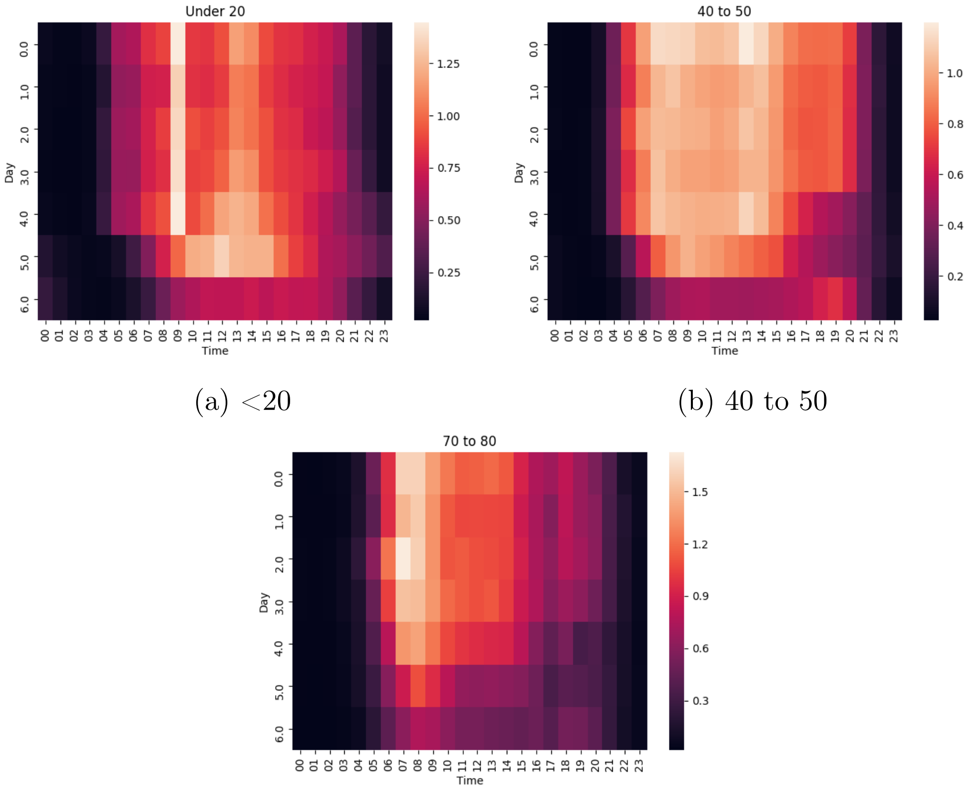
It is important to note that even though a cell shows that the distribution of e.g. female customers is more prominent, the actual number of male customers logging into the bank can be greater, since the total amount of male customers in the bank is higher.

Age Figures 4.4 show the login activity of different age groups. We made samples of many different age groups, by varying the length of the interval. The heatmaps shown in Figures 4.4 were chosen because of the way they clearly show the change in use-patterns.

The pattern for users under the age of 20 is somewhat special in the way that it differs from the other age groups. These users are often students in school, and tend to check their accounts often on weekdays at 09–10 am.

All other age groups follow mostly the same pattern of when they log in to the bank. Notable is that most logins are done inside working hours; between 07–15 on weekdays. This trend is the most extreme in the 40–50 years segment seen in Figure 4.4b, and becomes less extreme when one gets closer to either the 20–30 segment, or the 80-90 years segment.

Even when we split the users into age-segments, we still had logins every hour of every week—even for customers above the age of 80. The number of logins that happen late at night, or early morning, is not that high percentage-wise, but due to the size of the customer base, it is still significant.



(a) <20

(b) 40 to 50

(c) 70 to 80

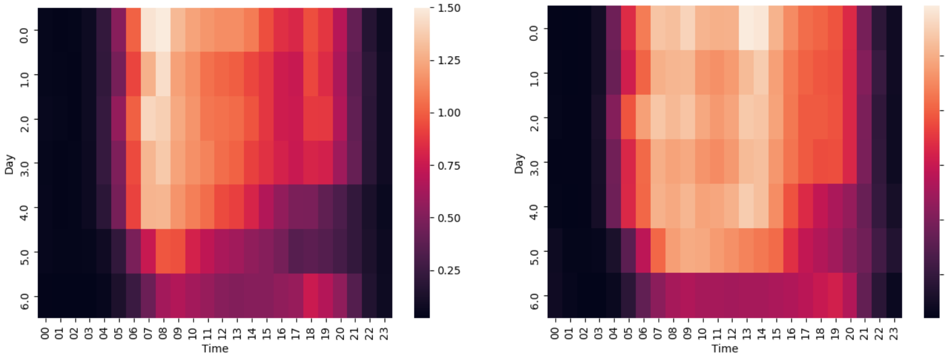
Figure 4.4: Heatmaps showing when customers of different ages use the online bank.

Login methods The bank has, as mentioned earlier, 6 ways of logging in:

Method	Percentage
Bank ID	6.69%
Code card	1.87%
SMS	6.71%
QR code	0.59%
Bank ID mobile	8.48%
App	75.66%

Table 4.4: Frequency distribution of what methods the customers use for logging into their accounts.

All methods for logging in show some of the same trends, with a large portion of all logins being during weekdays, in the time span 07–14. A clear distinction can be seen between the use of the app, and all the other methods that are usually used while on a computer. Use of a computer for logging in is causing the large spike in activity early in the day on weekdays, at around 07–09. The app is mostly responsible for the spike that can be seen later in the day, at 13–15, but use of the app is also more spread out over the entire day, as well as the weekend. The app is used a lot more than all other methods combined, which is probably due to the nature of the app being the easiest way of checking bank accounts while away from home.



(a) BankID, code card, SMS, QR code and BankID mobile

(b) App

Figure 4.5: Heatmap showing how the different devices divide into two main groups of daily activity.

4.5.3 Hypothesis 5: Individual Scoring of Login Activity

The heatmaps gave valuable insight into how the customer base behaved as a whole, but were not detailed enough to get a good estimation of how a specific customer would normally behave. We wanted to see if we could calculate a risk score for each individual customer.

Machine Learning and Anomaly Detection We began by looking at heatmaps with activity for individual users, to see if we could detect any patterns. We wanted to try making a distance-score using machine learning techniques for grouping the logins into groups or clusters. Our hope was that a new login that happened inside one such cluster would be safe, and outliers would be classified as anomalies, thus warning us of a possible identity theft.

Again, we used Python, Pandas, and one additional library called Scikit-learn. Scikit-learn is a Python library that offers simple and efficient tools that can be used for a wide range of machine learning techniques. We wanted to use some of the different clustering algorithms that they offered, in order to see if we could manage to cluster the existing data points for each user in any meaningful way.

Some of the techniques we tried were k-means, mean-shift clustering, and Gaussian mixtures. Our biggest dilemma was to figure out how we were going to score "closeness". What made one event closer to the other events, and what could make it a point of high risk?

Our Model We wanted more insight into how each individual customer acted over time. Machine learning might be a good way of finding anomalies in the customer base, but we saw that we first needed more insight about our dataset, in order to make good predictions for future models.

We also knew that whatever method we managed to make in Python, would have to be implemented anew inside the systems in Sbanken. This caused us to start looking into ways of making a more simple, mathematical model to score every single login by. The model could then give us increased insight into how the users acted over time, which could be used for further work in making a more sophisticated scoring model.

Simplifying the Model There existed six different methods for logging in, and each of these could be used in a total of $7 * 24 = 168$ different time-periods. This gave us a total of $168 * 6 = 1008$ possible locations to represent a login. We took a closer look at the results that we got from the measurements described in Chapter 4.5.2, and found some ways that we could simplify the model greatly. This could help us find a more intuitive way of measuring the difference between a new login, and previous logins for each customer.

From Figure 4.2, we see that there are only small variations between the pattern for weekdays. The pattern forming on the weekend is also more like each other. Some variations were observed, such as Friday acting like a weekday in the morning, and a weekend in the evening, which actually matches reality for most people pretty well. As a first step towards making the model, we concatenated all weekdays together, and Saturday and Sunday together. This caused us to have a 2x24 matrix, instead of a 7x24 matrix, to represent time.

We also found that at every login, we only cared for two different types of methods used; the method used in the newest login, and a collection of all logins regardless of method.

Our new graph now looked like a 2x2x24 grid, which gave us 96 possible points, instead of the original 1008.

Listing 4.1: Individual scoring of login activity

```
if (first time user logging in)
  return 1.0
newest_login = the current login session for the user
all_logins = list of all logins by that user
logins_same_method = list of logins made with the same
      login method as newest_login
if (highDensityArea(newest_login, logins_same_method))
  return 0.0
else
  if (highDensityArea(newest_login, all_logins))
    return 0.5
  else
    return 0.9



highDensityArea(newest_login, login_list):
  if(newest_login happened at a time of high activity, when
  considering the distribution of logins from login_list)
    return true
  else
    return false
```

Measurements We wanted to get a score that could say something about both the time that the user logged in, and what kind of method that they used to do so. Pseudocode of how we made this check can be found in Listing 4.1.

At the beginning, we check if this is a new customer or not. In the case of a new customer, then there is no further need for analysis. A new customer will get a high score by default. Our model is set to be run whenever a user wants to do a transaction. Because of this, we say that a new user that wants to transfer money out of the account, is sketchy. For now, such an event is marked as very suspect, but further analysis of the function's results is needed. It might be better to move it somewhere else, or maybe make it less severe. Preferably, this part of our function can become a part of a completely different check that can analyze the risk of new customers.

We want to see how similar a new login is to earlier sessions for that user. A login can be regarded as a good match, a somewhat close match, or a complete miss, and will be scored accordingly. The better the match, the lower the score. If there is a high chance of identity theft, then a high score will be returned.

The first thing we consider is if the user has used this method for logging in before, and if the newest login matches the pattern made by earlier logins. If this is the case, then we give this login a low score. If this is not the case, then we consider all of the earlier logins, regardless of the method used. We repeat the matching process here, and see if our login falls into a usual time-slot for the customer. If this is the case, we return a medium high score. The time is matching, but an unusual instrument for logging in is used, so we do not trust the action completely. If we do not match on either instrument or time, then we return a high score.

Density Function For each new login, we measure to see if this login happens in a "high-density area". Simply put, this means that we look at the number of logins that have happened in the same time-frame as the newest login, compare this to all the other times that this user has logged in, and see if this new login is likely or not.

To allow for some deviations in the user's pattern, we look at intervals of a specific size, instead of one single hour. Our chosen interval is of size 3. If a user logs in at e.g. 12:50, we round this number down to 12, and add and subtract one hour, so that we look at an interval from 11:00 to 13:59.

Regardless of the length of the interval, there is a total of 24 possible intervals to consider. We chose to only consider the intervals that had existing logins for that user. The function checks if our current interval is more or less likely than that of any other interval made from the user's old data.

There was no way of testing the method on the real customers before putting it into production. But, to get an idea of whether it would work as we intended, we ran it a few times on some fictive customers; some with only a few logins, some with prominent clusters of logins, and some with a more even distribution. We got the responses that we expected from the function, and it was included in Sbanken's systems.

4.5.4 Future Work

The method was put into production, but as of now, it is only logging its results so that we can later go in and observe them. We can then see if it had any useful information on detecting identity cases, or if it is too "trigger-happy", and reports too many false positives.

We did not know how much data we should feed into the method. For now, we feed it the entire login history of each user. We would like for the function to become more dynamic, to allow for changes in each user's pattern. One alternative is to consider the number of logins for each user, and choose the maximum of e.g. last 1000 logins and all logins from the last year. Another possibility is to have a weight for each login, which gets smaller and smaller the older the login is. That way, the newer logins will be regarded as more important, and affect the scoring function to a greater degree than the older logins.

4.6 Common Methods Used by the Criminals

We wanted to get as much information as possible from the known identity theft cases. This included looking at the distribution of the different login methods, and identifying what methods had been used to create each account.

Hypothesis 3 and 4 both state that we can find a common pattern between different identity theft cases, based on the way that the criminal has accessed the account. Hypothesis 3 states that the criminals have a preferred way of *registering a new account* in the victim's name, and Hypothesis 4 states that there is a preferred *login method* for the criminal.

The two hypotheses look at different stages of an identity theft, but they could both be researched by analyzing the same data; the log files about each victim's account.

As mentioned in Chapter 4.4, we already knew if each victim had reported that they were a victim of familial or non-familial identity theft. We wanted to use this knowledge to look for any correlation with the methods used by the criminal to access the accounts.

Hypothesis 3, which looks at registration methods, would need to be checked for cases of new account fraud. Hypothesis 4 could be checked for all cases of identity theft.

4.6.1 Hypothesis 3: Method for Registering a New Account

We wanted to see what methods had been used for creating a new account in all the cases where the victim reported that someone else had created a user account in their name. A customer could become a member of Sbanken by either showing a passport at a post office, or sign in with an already existing BankID issued by another Norwegian bank.

Table 4.5 shows the distribution of methods used for registering new customers in the bank. Only cases marked as new account fraud have been used for making these statistics.

BankID and BankIDMobile have been aggregated, as they both mean that the customer has used a BankID from another bank to register as a customer in Sbanken.

As we mentioned in Chapter 2, a letter is sent to notify the user that they can collect their OTP device at a post office. This happens automatically if the user registers by passport, and the letter is sent to their registered home address. Some of the victims reported that they became aware of the attempted fraud by receiving such a letter, which meant that the criminal was not able to retrieve the OTP device, nor successfully log into the bank. This is unfortunately not true for all the cases, and most of the cases that we looked at lacked any information about how the victim had discovered the identity theft. We have not been allowed to see the distribution for registering methods for all of the customers in the bank, so we are unable to say if the numbers we see in Table 4.5 show some trend unique to identity theft, or if it matches the pattern seen for all customers.

	Familial	Non-familial
BankID	32%	34%
Passport	11%	23%

Table 4.5: Distribution of identity theft victims, based on method for registering as a new user. We only used cases marked as new account fraud for finding these percentages. Together, the four cells summate to 100%.

4.6.2 Hypothesis 4: Method for Logging Into an Account

For all investigated cases, we wanted to identify what method had been used for logging into the bank in the sessions initiated by the criminal. This included the session in which they requested a loan, and when they managed to send money out of the account.

In order to do this, we needed to know which sessions had been initiated by the criminal. This could be very hard to figure out if the user did not detect the theft right away, and continued to use the account. The information would then be mixed up, and someone would need to go in and identify exactly what actions had been conducted by the criminal, and what had been done by the actual user.

Some customers did not acknowledge ownership of the account at all, meaning that they claimed that the criminal had created the account. This would make it possible for us to see directly what method that the criminal had used for creating the account. Most of the cases had no or little information about what had happened. This meant that we would have to manually go into each case and look through all actions done by that user. This was unfortunately not something we were allowed to do as outsiders. We got to look at some of the cases, but as the bank would have someone present while we looked at the data, it was not enough for us to get the analysis that we needed.

4.7 Reuse of Contact Information

Two of our hypotheses, 6 and 7, state that looking more closely at the contact information connected to a specific account, might uncover identity theft.

Hypothesis 6 states that changes made in the stored contact information of the user, especially if more than one such piece of information is changed during the same session, or in a short time-frame, is an indication of identity theft. An example of this could be a user that updates both their phone number, email address, and/or home address. We do not see such changes as a likely scenario for normal users, as there should be no reason for more than one of these factors to change at the same time.

Hypothesis 7 represents our belief that looking at reuse of information between accounts, can help detect identity theft. An example of this is if the same phone number is used as the primary number for two different accounts. This would mean that the same phone number could be used to get OTPs for two different user accounts, which should belong to two different people. This scenario is illustrated in Figure 4.6

Changing the phone number, or reusing a phone number for several accounts is seen as a clear warning sign, since it can be used to receive OTPs by SMS. Thus, phone number was the first thing we wanted to research for both of these hypotheses. Later we could add checks for changes in home address, email address, or any combinations of these.

When we started this project, we were under the impression that we could not do direct calls to the database to retrieve the information that we wanted, which made this task worse than what we expected. Someone would have to manually go into each case and look through their history. We were not allowed to do this task ourselves, so someone in the bank had to sit next to us and control the computer in order to show us the information that we needed. At the beginning of the project, there was no available time for them to do this, but luckily, we managed to get some of the needed information at the final stages of this project. In the last days of our work, it was discovered that it was actually possible to retrieve this information directly from the database. Someone would still need to write the queries, but this could save a massive amount of time.

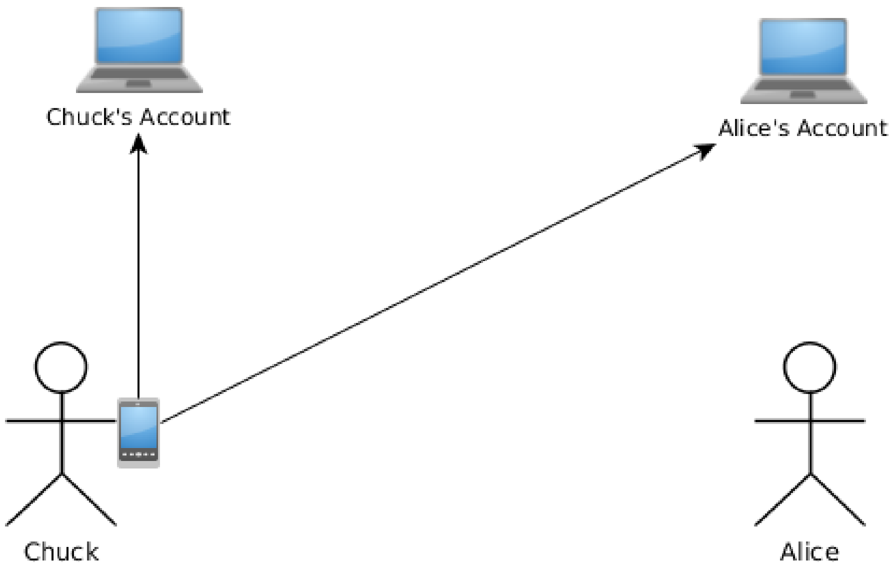


Figure 4.6: Representation of hypothesis 7; Chuck's phone is registered for both his own user account as well as the account belonging to Alice.

4.7.1 Hypothesis 6: Changes in Contact Information

We wanted to see how effective it would be to look at whether the user newly had made changes in their personal information. We believed that this check would be most effective for cases of account takeover, as opposed to new account fraud. If someone were to take over an account, then they might want to change some of the user information in order to make later logins easier for themselves. If the criminal had created the account, then they should not need to change any of the information, as they were the ones to add it in the first place.

Existing Cases We looked at logs of the reported identity theft cases and found that almost 20% of the cases had users who changed their phone numbers. We wanted to see when the change had taken place, as a phone number that had been changed years earlier should be of no consequence. What was really interesting was to see if someone changes their information, and then requests a loan. Unfortunately, we did not have access to look closely at the log files to find this information.

Only 5% of all the cases were what we saw as account takeovers where the phone number had been changed. This means that 15% of the identity theft cases were what we regarded as new account frauds. These were either cases where the victim denounced any ownership of the account, or we lacked information about the case.

We had expected the balance to be the other way around, with more reported account takeovers having changed their phone numbers. It might in fact be that these cases indicate a fraud attempt against the bank, instead of representing a real identity theft, but we have no way of knowing for sure. These identity thefts have to be researched more closely, by someone with more information about the cases and their verdicts, to understand why this is happening.

Implementation We started by implementing some new logging. Every time a user performs a transaction, we check if they have changed their contact information during the last 24 hours. If that is the case, then the information would be saved to the log files. By placing them at the time of transactions, they could be run simultaneously as the other fraud checks in the bank. This made the new check fairly easy to implement, as we already had the information we needed in place.

Our hope is that after some time, we will have collected enough information to see how often such a function would be triggered, and see if it could be of any help in detecting any new identity theft cases that may appear. We will need to see if the time-frame in which we look for changed contact information, should be expanded.

4.7.2 Hypothesis 7: Reuse of Contact Information

Our goal was to look at both phone numbers, email addresses, and home addresses, and any combined changes of these. We chose to start by checking phone numbers, as this would give us a decent measure of how easy the tests were to implement, and how we could find the information that we needed. It should be simple enough to add the other checks afterwards, and see if we can gain even more effective checks.

Existing Cases We managed to check this hypothesis for a few of the newer identity theft cases, but not all the cases, as we did with Hypothesis 6. The reason for this was how time-consuming it was to manually look through the cases, which someone in the bank had to do because of privacy regulations. For the cases we looked at, slightly over 15% of the cases had registered the same phone number as another customer of the bank.

There can be several reasons as to why more than one person has the same phone number registered. Here are some of the reasons we find the most likely:

1. A family only has one phone, which they have to use on all their accounts.

2. One of the user accounts is "real", and this user has used their phone number on their own account, as well as on a stolen account.

3. Both users are victims of identity theft.

One case was especially interesting, as it had both undergone a change in phone number, and this new number ended up matching the number of another customer of the bank. It would be very interesting to see if this other person was the culprit, or if they were both victims. Either way, such a result can be useful for the investigators that should look into the case.

Implementation We started to implement some logging, to see how many users actually had the same phone number as another customer in the bank. For the sake of convenience, we added this check at the same stage as the other checks; whenever a user wanted to make a transaction.

Our score would return 0 if there were no other customers with the same phone number as the user, or 1 if there existed someone else with that phone number. In the same manner as the other new logging systems, only the non-zero scores were added to the logs, to avoid unnecessarily messy files.

Primary Findings Interestingly enough, there existed a lot of accounts that used the same phone number as another account.

Our findings seem to match the 3 types which we described earlier in this subchapter. Most look like family members, who have used the same phone to register their accounts. Some might look like they have been helped by the same person to create accounts, and have all used this same number. This might not have been such a harmful thing a few years ago, when one could not use a phone to receive OTPs, but it should not be done today. We believe that these cases do not violate any policy of the bank, but it is strongly discouraged behavior.

Hopefully, banks can find a way to warn their customers of this, and encourage them to update their user information, as a step towards protecting the user account against identity theft.

4.7.3 Alternative Ways of Notifying the Customer

When looking closely at what contact information the bank has about each customer, it became a logical next step to see what methods the bank uses to notify their customers. Since we had just seen that the means of contacting the user could become compromised, we wanted to look into what other confirmation methods the bank can add for better security. One idea we had, was that the bank could send a mail to Digipost every time a user requested a loan. Digipost is a digital mail service that all Norwegian citizens can use to receive mail. Digipost has almost 2 million users in Norway, and 500 companies and organizations use it to send secure mail to their customers.

Adding this feature has been temporarily put on hold in the bank, possibly because it is not free to send messages to Digipost, but our hope is that all banks eventually will send out notifications to their customers in this manner. We also believe that this will give the customers more confidence in the banking system, as they now have a way of instantly knowing if someone were to take up loans in their name.

As the thief may very well also have access to the victim's Digipost-account, we recommend introducing a feature to make some messages impossible to delete. Unlike an SMS or email, we have the option to control what is seen and stored on Digipost, as it is owned by Norway Post, which again is owned by the Norwegian Ministry. This will mean that the victim can check their Digipost account to make sure that they do not have any loans they do not know about, even if someone else has access to their account.

4.8 Credits and Loans

Consequences for the Bank We wanted to get an overview of the actual loss in each identity theft case, as well as what the potential loss could be for both the bank, and the victim. We were not allowed to look at this information directly, but we were promised that another department would look into it, and report back to us with their findings. The information should include the actual economical consequence each of the identified identity theft cases had for the bank.

The department in question was contacted several times, but never reported back. In the end, this question still remains unanswered.

Consequences for Individuals The arguably most damaging part of an identity theft is if someone manages to take up loans in the victim's name. If a criminal steals someone's life savings, this is of course damaging, but the victim will at least only go back to zero. Taking up loans can push the victim into the negatives, and might make them indebted for the rest of their lives.

Thus, we wanted to add some more security mechanisms for detecting and stopping cases where a criminal manages to take up a loan in the victim's name. We had 3 guiding hypotheses in our work to do this, which we now will explain more in depth. Some of these hypotheses remain unanswered, and we will take a look at why we could not answer them.

4.8.1 Hypothesis 8: Visiting the Financing Web

Hypothesis 8 states that there is an increased chance of identity theft if someone takes up a loan without first visiting the financing web. This was a hypothesis that Sbanken strongly believed in. The financing web is Sbanken's own web-pages that have information about possible loans, and what kind of interest it is possible to get. The reason behind this hypothesis is that most people taking up a new loan should want to research this beforehand. A criminal might know exactly what they want to do in order to get the money that they want, but a regular user that wants to take up a loan in their own name, should want to research this closely before committing to a debt.

The first thing we needed to measure, was how many of the normal customers visited the financing web before requesting a loan. Maybe it was something everyone did, or maybe no-one actually did it.

A problem that appeared when we looked for a way to measure this, is that many users might not be logged in when they look at the financing pages. There could be many users that first look at the pages to figure out exactly what kind of loan they want, before then logging in to actually request the loan.

There is of course some logging of who visits the bank's web pages. The problem came in matching this information up against specific users that later logged in. We started some preliminary research into how such a check could be implemented, as it would mean that we had to add some more information into the log files. We tried to implement a very simple matching between the visitors and the customers that logged into the bank, based on IP addresses, but realized that doing what we wanted to do was a much bigger operation than what we first assumed.

The new scope of implementing this type of test, led us to set the entire check on hold. The structure of logs would have to be changed, which would lead to even bigger changes. Sbanken did not currently have the capacity to work further on the details of how this could be done. They still believe in this hypothesis, and will in all likelihood look further into finding a way of measuring this.

4.8.2 Hypothesis 9: Transactions

After someone manages to get approval of a request for credit in another person's name, then it is a logical conclusion that they would want to get this money out of the victim's account. This can be done by either transferring the money to another account, or to buy goods or services. Hypothesis 9 states that there is an increased chance of identity theft if the user immediately after receiving a loan, tries to transfer the money out of the account. This check would come in addition to existing checks that analyze each receiving account, which are already a part of Sbanken's fraud detection systems. An example of the already existing checks, is that every receiving account would be checked against a list of known mole accounts, which have been used for whitewashing money.

Implementation We implemented a check that would run each time a transaction was made by some user. The check would be triggered if the user had newly received a loan from the bank. In addition to this, Sbanken implemented a check that would look at the amount of money transferred out of the account. If the amount to be transferred was over 90% of the total amount in the account, then it would trigger a warning.

These two could be combined into a weighted score, so that in the event that both of them are triggered, then the resulting score will be much greater. Both of these two checks were fairly easy to include in the existing systems at Sbanken.

4.8.3 Hypothesis 10: Unsecured Credits

Hypothesis 10 states that identity theft cases have an abnormally high amount of unsecured credits. Unsecured credits are a type of loan that is not secured with property, or other collateral. Use of such credit can give the user almost instant access to money, but they also have a much higher interest, which makes them very expensive for the user.

To answer this hypothesis, we needed to know the amount of unsecured credits that had been given to the criminals. Then we could compare this to the amount of unsecured loans taken by the average customer in the bank.

We started the research by trying to figure out if the identified identity theft victims had a lot of this type of credit. This was not information that we could find ourselves, so we contacted another division of the bank and asked if they could help us. They responded that this should be possible to do, but over a year and several inquiries later, we have still not received any information from them. In all likelihood, someone decided that we should not get this information, as we were not employees of the bank.

We are still curious as to whether this hypothesis is correct or not, and hope that the bank will look further into the case after our engagement in the bank is over.

4.9 One Model

Sbanken already has a working model for detecting instances of fraud, which will calculate a score for each transaction in the bank. These existing checks focus on what information the bank has about the user sending the money, and the receiving bank account of the transaction. For the receiving account, the bank will check if it is black-listed, if this is the first transaction for this user to that bank account, and if it is the first time the bank sees this bank account. For the sender, the bank will check whether the user has logged in from a new IP address, device or browser.

Our goal was to expand upon this. To do so, we made several scoring models which can be used for warning of identity theft. These checks also run at the time of transaction, in addition to the existing fraud checks. Here is a list of our implemented checks:

- Is the transaction happening at an unusual time, or with an unusual login method for this user?
- Has the user newly received a personal loan?
- Has the user changed their phone number?
- Do other users in the bank have the same phone number as the user performing the transaction?
- Will this transaction empty, or almost empty, the user's bank account?

Most of the checks return either 0 or 1, with the exception of the check that looks for changes in the user's login pattern. That check returns either 0, 0.5, 0.9 or 1, depending on how suspicious we believe that login to be. All of these scores are aggregated into one score by summing them together, and then normalized by the number of positive scores. Eventually, we want to use this number as a threshold value, so that users with a score below this threshold should be trustworthy, and users above it may be victims of identity theft.

In Table 4.6, we have listed how often our checks triggered in the bank in April 2019. It shows what percentage of all transactions were triggered by each individual check.

Scoring Model	**Hit**
Unusual login time	16.862%
Unusual login method	7.675%
First time user logs in	0.014%
Received credit	0.116%
Change phone number	0.080%
Reuse of phone number	4.269%
Empty $\geq$ 90% of the account	6.506%
New recipient for the customer	30.058%
New IP for the customer	22.210%
Recipient is a blacklisted account	0.001%

Table 4.6: Percentages of how often each function is triggered, showing numbers from April 2019. The numbers in the table are the percentages of all transactions happening in that month, which triggered that specific check. The three last checks are listed as reference points for our new identity theft checks.

It is important to note that these numbers do not indicate any success of our functions, as we have no information about whether any of these transactions were the result of identity theft. However, what we can learn from these numbers, is something about how sensitive each of our functions are.

The low values for e.g. received credit and changes in phone number, might indicate that we have made some useful models; if the percentages had been too high, then it might not be a very useful indicator. This could be the case for e.g. the function that looks for changes in a user's login patterns, as over 20% of all transactions trigger this check. Then again, it could also turn out to be a good score, when seen in combination with the other scores.

It is our belief that we can achieve even better results by finding different combinations of the scoring models, so that the resulting single score will give a more accurate indication of when we should suspect identity theft. Of course, we want to find a model that can give the bank a good indication of when an identity theft has happened, without returning too many false positives. To do this, we need more analysis of what scores often trigger together, as well as testing different threshold values. We will discuss this more closely in the final chapter.

Chapter 5

Discussion and Conclusion

5.1 Summarizing our Results

We have looked closely at several different hypotheses in order to find a way to detect and stop identity theft in online banking. Some of these have led to the creation of new checks and security mechanisms, while others will require more time and research than what we had at our disposal.

Our largest hurdle in this work has been time. Firstly, it took more time than what we first assumed to collect the information that we needed. Then, after creating our new checks, it took time to integrate them into the existing systems. We are currently in the final waiting period; waiting for definite results of our checks being able to detect identity thefts happening in the bank.

We have started a process in the bank, where they will look more closely at identity theft cases, and analyze the results of our new logging systems. We have found some indications on whether a few of the hypotheses are correct or not, and several of them have led to changes in how Sbanken works with detecting identity thefts.

We will now review and discuss what we see as our most important findings. Afterwards, we will list our plans for future work on this subject, and give a brief conclusion of our work with mitigation of identity theft in online banking.

5.1.1 Characteristics of the victims

More Female Victims We found that there are more female than male identity theft victims. The data presented in Table 4.3 shows that there is a significantly higher percentage of female victims of familial identity theft fraud—2/3 of these victims were female.

We also found that the balance of male and female victims of non-familial identity theft, if one takes into consideration the gender distribution of the bank's customers, is approximately 50/50.

Our findings indicate that both genders are equally proficient in keeping their personal information or login credentials secure from a non-familial person. It is striking that there is such a large difference when comparing familial and non-familial identity theft.

In Chapter 2, we looked at how familial identity theft is often a consequence of misplaced trust in family members. One reason for why there is such a high percentage of female victims in this category, might be that they are quicker to trust others, and are thus more at risk of becoming victims of identity theft. Another explanation could be that many women may not control their own economy, either as a personal choice or as a result of cultural norms. It could also be that the balance is more even than what we have seen, but that a higher percentage of male victims of familial identity theft are more embarrassed by this fact, and thus avoid reporting it. This finding merits a deeper study on its own.

As stated in Chapter 4.4, we do not know the gender of the perpetrators. It could be very interesting to see more statistics about this issue as well.

Age In Chapter 4.4.2, we looked at the age distribution of identity theft victims. Our findings here, as depicted in Table 4.2, indicate that there are no specific age groups that are more susceptible than others; people of any age can become victims of identity theft.

5.1.2 Different Types of Identity Theft

We looked at the distribution of identity theft victims based on the type of identity theft that had been committed. As we have mentioned before, this information is extracted from the report made whenever a victim informs the bank of an identity theft, and could thus contain erroneous information, but we have no other way of measuring this than to trust the users on this point.

Table 4.3 shows the distribution of identity thefts based on the four types that we wanted to look at; familial, non-familial, new account fraud, and account takeover. In total, we see that 75% are new account cases, where the victim claims that someone else has created the account in question, and only 25% are accounts that have been taken over.

The distribution of familial and non-familial thefts is more balanced, but there is still a significant difference; 42% are familial cases, and 58% are non-familial. As we have mentioned before, "non-familial" can mean either that the victim does not know the perpetrator, or that they simply do not want to name the perpetrator to the bank.

It could be that far more identity thefts happen in a familial setting than what we have seen. Underreported cases of familial identity theft can have many causes; the victim might try to "protect" the other person, they can be embarrassed that they have been fooled by a family member, or they might even avoid reporting it in an attempt to hide evidence of their own sloppy security routines.

5.1.3 Contact Information

Hypothesis 6 and 7 state that changes in the contact information of each user is an indication of identity theft. We have looked at cases where the user has changed their phone number, and cases where they have the same phone number as some other customer in the bank.

There might not be anything criminal going on just because two users share the same phone number, but it would make identity theft easier to commit, and should thus be detected and hindered. Changing one's phone number could also be necessary in some cases—when trying to avoid an aggressive ex-boyfriend or girlfriend, or for celebrities who get their number leaked to the public—but it is not something people would normally ever do. Especially changing one's number before making a payment, or requesting a loan, should be seen as a clear warning sign of criminal activity.

We found that 20% of our known identity theft victims had at some point changed their registered phone number. We also found that 15% of the victims had the same number as another customer of the bank. Both changes in contact information, and reuse of said information, is now checked each time a user performs a transaction in the bank.

5.1.4 New Account Fraud

New account fraud is probably the hardest to detect, as we have no prior information about the correct user, and can thus not observe any changes in their use-patterns. This category makes up 75% of the reported identity theft cases, so finding a way of identifying these cases should help a great deal. Testing to see if a user has the same number as someone else can catch some of these cases, but to get more of them, we also need to look at how the criminal managed to create the new account.

Registration Methods We found that roughly 65% of the new accounts were created by using a BankID from another bank, and that only 35% had been created with passports. The distribution of familial versus non-familial thefts is almost the same for registrations using BankID. Use of passport is not as evenly distributed; 2/3 of the profiles created by this method were non-familial.

Several victims in the passport-cases reported that someone had tried to collect their OTP-device at a post office, but not succeeded. Others detected the attempted theft when they received a dispatch note from the postal service. We do not know exactly how many of these cases were "successful", meaning that they had managed to trick the post office worker with a fake passport, or had a real passport that they had pilfered from the victim. Nonetheless, we can assume that a significant portion of the attempted identity thefts done with the use of passports, were stopped.

5.1.5 Individual Pattern Detection

Analyzing the login behavior of regular users of the bank, we found that there are some very clear patterns discerning the genders, and different age segments. For now, we wanted to make an individual scoring for each user. That way, we can see if each user changes their use-pattern, as we postulated in Hypothesis 5 might be a good way of detecting identity theft. Including a comparison with the general patterns regarding age and gender segments might be useful in further improving an individual scoring of a customer—especially if they are a relatively new customer without a lot of prior data.

We implemented a function that looked at the login activity of the user, which gives each user a score in accordance with how well they match their own historical login activity in the bank. Specifically, we look at the time of the login, and what device they used to do so.

For now, we have the beginning of a model that analyzes the behavior of each individual user, but we have already seen many ways in which it can be further improved or expanded upon. For instance, instead of only looking at the time of login, we can also make a similar check that analyzes the pattern of when and how each user normally performs transactions in the bank.

Another way the function can be improved, is to find a good way of limiting the data that the function uses. One example of this, which we introduced in Chapter 4.5.4, is to have a weighted score in each login, so that newer logins will be weighted higher than older ones. This would allow for gradual changes in the user's pattern.

5.2 Further Work

As we have seen, detecting identity theft is no simple matter, and getting enough information about how and why it happens, takes time and effort. We still have many unanswered hypotheses. Some of them were made on assumptions that proved to be wrong, such as how time-consuming it would be to research that idea closer. For others, we were hindered by privacy concerns. It is our hope and belief that the bank will prioritize these hypotheses even after our time with them is over, so that they might get answers that can help them in gaining an even better understanding of identity thefts.

5.2.1 Introducing New Routines

We are still in a phase where we are collecting more information about how this type of crime happens, and what patterns the bank can look for. Our experiences lead us to believe that a new way of handling identity theft cases might make this work significantly easier. One example could be to have a register of all identity theft cases, with information about what has happened, and what type of case it is. Such a system can also be used for monitoring any changes in the type of identity theft that happens, as new cases will be easier to compare to the old ones. The system should also make it easier for anyone looking through the case to identify which actions the user claims are part of an identity theft.

5.2.2 Tuning Our Model

We ended Chapter 4 by summarizing the new scoring model Sbanken now uses for finding possible victims of identity theft. We explained how our functions will score each transaction happening in the bank, and that these scores will then be summed together. This final score can be used to see whether the user may be a victim of identity theft, and trigger further investigation.

We also explained how we need more time to analyze the results of our logging, and tune our models, in order to get a more accurate scoring. We believe that finding good ways of combining our different scoring models for identity theft, is crucial for scoring each transaction correctly. With combined checks, we can say that in the event that a specific combination triggers, then we see this behavior as especially suspect. This will give an even stronger indication of identity theft than if they had been triggered separately. One such example could be if the user both sends almost all of their money out of the account, as well as having only a partial match on their login pattern. Another would be if the user changes their contact information, e.g. their phone number, and this new number is already connected to another user of the bank.

We believe that using machine learning to find better weights and combinations for this type of scoring might give us much needed insight into how to detect identity theft. We already have a defined set of reported cases, and a control group. After our work on this thesis, we also have a set of parameters to test on, and have the knowledge of how to extract and look at these. It should then be possible to test different weighting strategies, to see which one can enable us to find the different types of identity theft that we have identified from the reported cases. This has the potential of giving us a much more reliable way of detecting future identity thefts, but to say anything about the success of the model, it is necessary to know the rates of TP, FP, TN, and FN that it gives us.

5.2.3 New Warning System

We have seen that many identity theft cases take time to detect, which can cause the sums in question to grow enormous. Some thieves also try to change the contact information connected to an account, making it even harder for the victim to detect that something is amiss. We propose to create a way of informing all Norwegian citizens whenever a loan has been taken up in their name. One example could be that all banks have to send a message to that user on Digipost, as we discussed briefly in Chapter 4.7.3. It might be a good idea to find a way to make these messages impossible to delete, to ensure that no other person with access to the account can hide such messages.

In Norway, there is currently an ongoing process of making a new register that financial institutions can use to see how much unsecured debt a person has. This is a fantastic system, which will most likely also have the side effect of mitigating the widespread consequences of identity thefts—the banks will have a clear way of seeing whom already have too much existing debt, and should not be allowed to take on even more, as is usually the case whenever someone is a victim of identity theft. We believe that this system may serve another purpose as well; hopefully, it can also become a way for people to keep tabs on their own debts, and to control that no one else has managed to take up any loans in their name.

In Chapter 3.3.3, we saw that almost 40% of a national survey responded that they did not trust the police to help them if they ever were the victim of a cybercrime. Knowing that they have a way of personally checking whether someone has taken up a loan in their name, might make them feel more at ease.

5.3 Conclusion

Our goal has been to find ways of better detecting identity theft in online banking. To do this, we needed to get a better understanding of how and why identity theft happens, to see what problems such cases cause, and why this has not been solved before.

We have seen that the consequences can become devastating for individuals affected by identity theft, and that the rules and regulations that should have protected them, have become outdated.

Our main hypothesis was that we could use the metadata about each user account to better detect identity thefts in the online bank. This has led us to create detection systems that now run in the bank, which look for changes in user patterns, or specific points of interest. We do not have conclusive proof that our hypothesis is correct, but our findings indicate that it has a strong possibility of being so.

Hopefully, other banks may also use this knowledge to improve their own systems. Together, the banks can take an important step towards eradicating identity theft from Norwegian society.

Bibliography

[1] Kjørven, M. E. "digitaliseringens pris" *NRK*.
https://www.nrk.no/ytring/digitaliseringens-pris-1.14505149, Apr 2019.
(accessed on 2019-04-8).

[2] Merriam-Webster. "Identity Theft".
https://www.merriam-webster.com/dictionary/identity%20theft.
(accessed on 2018-07-11).

[3] Nettvett.no. "Spørsmål og svar om ID-tyveri".
https://nettvett.no/sporsmal-svar-id-tyveri/, 2018.
(accessed on 2018-09-26).

[4] Sparebanken Hedmark. "Årsrapport 2009".
https://www.sparebank1.no/content/dam/SB1/bank/ostlandet/omoss/investor/rapporter2011-2014/aarsrapport_2009_norsk.pdf, 2010.
(accessed on 2018-11-04).

[5] Statistics Norway. "ICT usage in households".
https://www.ssb.no/en/teknologi-og-innovasjon/statistikker/ikthus.
(accessed on 2018-11-15).

[6] DIBS. "norsk e-handel 2018".
https://info.dibs.no/norsk-ehandel-2018-download?submissionGuid=07074af4-be2e-4c77-88b6-d8d89e25400b, 2018.
(accessed on 2018-10-12).

[7] Statistics Norway. "fakta om befolkningen".
https://www.ssb.no/befolkning/faktaside/befolkningen.
(accessed on 2018-11-15).

[8] NorSIS. "ID-tyveri og sikkerhet for egen identitet".
https://norsis.no/wp-content/uploads/2018/02/2017-12-20-Endelig-Rapport-Identitetssikring_des2017.pdf, 2017.
(accessed on 2018-09-25).

[9] Higgins, G. E., Hughes, T., Ricketts, M. L. & Wolfe, S. E. "Identity theft complaints: exploring the state-level correlates".
https://search.proquest.com/docview/235986102/E99FF60B71B84DEDPQ/1?accountid=8579, 2008.
(Accessed on 2019-04-05).

[10] Identity Theft Resource Center. "Identity Theft: The Aftermath 2017".
https://www.idtheftcenter.org/images/page-docs/Aftermath_2017.pdf, 2018.
(Accessed on 2019-04-05).

[11] Leiner, J. J. "A study of the experiences of fraud investigators in combating identity theft" *ProQuest*.
https://search.proquest.com/docview/1758252495, 2015.
(accessed on 2018-02-10).

[12] Navarro, J.C. & Higgins,G.E. "Familial Identity Theft".
https://search.proquest.com/docview/1867560656?accountid=8579, 2017.
American Journal of Criminal Justice : AJCJ, vol. 42, no. 1, pp. 218-230.

[13] National Research Council of the National Academies. *Who goes there?* National Academics Press, 2003. ISBN 0-309-08896-8.

[14] Statens vegvesen. "førerkortets sikkerhetselementer".
https://www.vegvesen.no/forerkort/har-forerkort/gyldig-forerkort-i-norge/eos-modell-2/sikkerhetselementer, 2018.
(accessed on 2018-10-15).

[15] Chaos Computer Club. "Hacking the Samsung Galaxy S8 Irisscanner".
https://media.ccc.de/v/biometrie-s8-iris-en, 2017.
(accessed on 2018-11-28).

[16] Syazana-Itqan, K., Syafeeza,A. R., Saad N. M., Hamid N. A. & Saad W. H. B. M. *"A Review of Finger-vein Biometrics Identification Approaches"*. Indian Journal of Science and Technology, volume 9 issue 32, 2016.

[17] Simmons, D. "BBC fools HSBC voice recognition security system". https://www.bbc.com/news/technology-39965545, 2017. (accessed on 2019-02-12).

[18] The Norwegian Tax Administration. "fødselsnummer". https://www.skatteetaten.no/person/folkeregister/fodsel-og-navnevalg/barn-fodt-i-norge/fodselsnummer/ . (accessed on 2018-11-14).

[19] The Norwegian Department of Finance. Høring - forslag til ny personidentifikator (fødselsnummer). https://www.regjeringen.no/no/dokumenter/horing---forslag-til-ny-personidentifikator-fodselsnummer/id2544699/ . (accessed on 2018-11-14).

[20] Difi. "ID-porten". https://eid.difi.no/nb/id-porten. (accessed on 2018-11-28).

[21] Hole, K. J., Moen V. & Tjøstheim T. *Case Study: Online Banking Security*. IEEE Privacy & Security, volume 4 issue 2, pp. 14-20, 2006.

[22] Julie Conroy. "Synthetic Identity Fraud: The Elephant in the Room". https://aitegroup.com/report/synthetic-identity-fraud-elephant-room , 2018. (accessed on 2019-05-20).

[23] Johansen, P. A. "Har slettet 70 falske barn i Folkeregisteret" *Aftenposten*. https://www.aftenposten.no/norge/i/g7EKB/Har-slettet-70-falske-barn-i-Folkeregisteret, Feb 2013. (accessed on 2019-02-26).

[24] Skatteetaten. "Modernisering av Folkeregisteret". https://www.skatteetaten.no/person/folkeregister/om/modernisering/prosjektet/. (accessed on 2019-02-26).

[25] Finansklagenemnda Bank. Uttalelse 2017-649.
https://publisering.finkn.no/statement/2017-649, 2017.
(accessed on 2019-02-13).

[26] IRS. "IRS Statement On Get Transcript”.
https://www.irs.gov/newsroom/irs-statement-on-get-transcript.
(accessed on 2018-12-18).

[27] Reuters. "Premera Blue Cross Says Data Breach Exposed Medical Data".
https://www.nytimes.com/2015/03/18/business/premera-blue-cross-says-data-breach-exposed-medical-data.html, 2015.
(accessed on 2019-02-19).

[28] Næringslivets sikkerhetsråd. "Mørketallsundersøkelsen 2018".
https://www.nsr-org.no/getfile.php/1311411-1539949973/Dokumenter/NSR%20publikasjoner/M%C3%B8rketallsunders%C3%B8kelsen/M%C3%B8rketallsunders%C3%B8kelsen%202018_ENG.pdf, 2018.
(accessed on 2019-02-18).

[29] Javelin Strategy & Research. "2017 Identity Fraud: Securing the Connected Life".
https://www.javelinstrategy.com/coverage-area/2017-identity-fraud, 2017.
(accessed on 2019-02-25).

[30] H. E. Malmedal, B. & Røislien. The Norwegian Cybersecurity Culture.
https://norsis.no/wp-content/uploads/2016/09/The-Norwegian-Cybersecurity-culture-web.pdf, 2016.
(accessed on 2018-09-25).

[31] Statistics Norway. "Lovbrudd anmeldt etter type lovbrudd. Absolutte tall".
https://www.ssb.no/lovbrudda/, 2018.
(accessed on 2019-02-03).

[32] Åsberg, A. R. "Morten deler identitet med en kriminell" *NRK*.
https://www.nrk.no/nordland/_-et-samfunnsproblem-at-fa-anmelder-id-tyveri-1.13402361, 2017.
(accessed on 2019-02-25).

[33] Spets, K. "Mathias (26) ble utsatt for ID-tyveri: Tok opp 1,2 mill. i forbrukslån på én uke" *Aftenposten*.

https://www.vg.no/nyheter/innenriks/i/216Jba/mathias-26-ble-utsatt-for-id-tyveri-tok-opp-12-mill-i-forbrukslaan-paa-en-uke, Feb 2019.
(accessed on 2019-02-19).

[34] Omland, P. "Når juss blir absurd" *Adresseavisen.*
https://www.adressa.no/meninger/kronikker/2018/03/09/N%C3%A5r-juss-blir-absurd-16236289.ece, Mar 2018.
(accessed on 2019-02-19).

[35] Clausen, V. "Eksen tok opp 2,1 millioner i forbrukslån i Jannes (29) navn: – Jeg må betale alt" *TV2.*
https://www.tv2.no/nyheter/10258162/, Dec 2018.
(accessed on 2019-02-19).

[36] Ervik, K. "Posttyven endret adressen til Anders (24) i Folkeregisteret" *TV2.*
https://www.tv2.no/a/5922170, Aug 2014.
(accessed on 2019-02-19).

[37] NorSIS. Nordmenn og digital sikkerhetskultur 2018.
https://norsis.no/wp-content/uploads/2018/11/Nordmenn-og-digital-sikkerhetskultur-2018-web.pdf, 2018.
(accessed on 2019-01-15).

[38] Javelin Strategy & Research. "2018 Child Identity Fraud Study".
https://www.javelinstrategy.com/coverage-area/2018-child-identity-fraud-study , 2018.
(accessed on 2019-02-27).

[39] Bank of America. "SafePass Online Banking Security Enhancements".
https://www.bankofamerica.com/privacy/online-mobile-banking-privacy/safepass.go.
(Accessed on 2019-05-13).

Appendix A

Norwegian Laws

A.1 Lov om finansavtaler og finansoppdrag (finansavtaleloven)

§ 34. Plikter ved bruk av betalingsinstrument

(1) En kunde som har rett til å bruke et betalingsinstrument, skal bruke det i samsvar med vilkårene for utstedelse og bruk, og skal herunder ta alle rimelige forholdsregler for å beskytte de personlige sikkerhetsanordningene knyttet til betalingsinstrumentet så snart instrumentet er mottatt. I tillegg skal kunden uten ugrunnet opphold underrette institusjonen, eller den institusjonen har oppgitt, dersom kunden blir oppmerksom på tap, tyveri eller uberettiget tilegnelse av betalingsinstrumentet, eller på uautorisert bruk.

(2) Institusjonen som utsteder et betalingsinstrument, skal, uten at det har betydning for kundens plikter etter første ledd, sørge for at de personlige sikkerhetsanordningene knyttet til et betalingsinstrument ikke er tilgjengelige for andre enn den kunden som har rett til å bruke betalingsinstrumentet. I tillegg skal institusjonen sørge for at kunden til enhver tid kan foreta underretning som nevnt i første ledd annet punktum eller be om at eventuell sperring av betalingsinstrumentet oppheves, jf. § 24a fjerde ledd. Institusjonen skal også sørge for at kunden i 18 måneder fra underretning som nevnt i forrige punktum kan dokumentere å ha foretatt slik underretning, og skal dessuten hindre enhver bruk av et betalingsinstrument etter at underretning etter første ledd annet punktum er foretatt.

(3) Det kan avtales at første ledd annet punktum og annet ledd tredje punktum ikke skal gjelde for småpengeinstrumenter hvis det ikke er mulig å sperre instrumentet for bruk.

(4) Institusjonen skal ikke sende betalingsinstrument uoppfordret, bortsett fra til utskiftning

av betalingsinstrument som tidligere er utlevert til kunden. Institusjonen har risikoen for sending av betalingsinstrument til kunden og personlige sikkerhetsanordninger knyttet til instrumentet.

§ 35.Misbruk av konto og betalingsinstrument

(1) Institusjonen er ansvarlig for tap som skyldes uautoriserte betalingstransaksjoner, med mindre noe annet følger av paragrafen her. En betalingstransaksjon er uautorisert dersom kunden ikke har gitt samtykke til transaksjonen, jf. § 24.

(2) Kunden svarer med inntil kr 1.200 for tap ved uautoriserte betalingstransaksjoner som skyldes bruk av et tapt eller stjålet betalingsinstrument dersom personlig sikkerhetsanordning er brukt, eller som skyldes uberettiget tilegnelse av et betalingsinstrument dersom kunden har mislyktes i å beskytte de personlige sikkerhetsanordningene og personlig sikkerhetsanordning er brukt.

(3) Kunden svarer for hele tapet ved uautoriserte betalingstransaksjoner dersom tapet skyldes at kunden ved grov uaktsomhet har unnlatt å oppfylle en eller flere av sine forpliktelser etter § 34 første ledd. Dersom betalingstransaksjonen har skjedd ved bruk av et elektronisk betalingsinstrument, svarer kunden likevel bare med inntil kr 12.000. Dersom tapet skyldes at kunden forsettlig har unnlatt å oppfylle forpliktelsene etter § 34 første ledd, skal kunden bære hele tapet. Det samme gjelder dersom tapet skyldes at kunden har opptrådt svikaktig.

(4) Kunden svarer ikke for tap som skyldes bruk av tapt, stjålet eller uberettiget tilegnet betalingsinstrument etter at kunden har underrettet institusjonen i samsvar med § 34 første ledd annet punktum, med mindre kunden har opptrådt svikaktig. Kunden svarer heller ikke for tap som nevnt i første punktum hvis institusjonen ikke har sørget for at kunden kan foreta slik underretning, jf. § 34 annet ledd annet punktum.

(5) Dersom kunden nekter for å ha autorisert en betalingstransaksjon, jf. § 24 annet ledd, skal bruken av et betalingsinstrument ikke i seg selv anses som tilstrekkelig bevis for at kunden har samtykket til transaksjonen, eller for at kunden har opptrådt svikaktig eller forsettlig eller grovt uaktsomt unnlatt å oppfylle en eller flere av sine forpliktelser etter § 34 første ledd. Det påhviler institusjonen å bevise at transaksjonen er autentisert, korrekt registrert og bokført og ikke rammet av teknisk svikt eller annen feil.

(6) Det kan avtales at bestemmelsene i første til tredje ledd samt femte ledd ikke skal gjelde for småpengeinstrumenter som anvendes anonymt, eller dersom institusjonen av andre grunner knyttet til betalingsinstrumentets karakter ikke vil kunne bevise at en betalingstransak-

sjon ble autorisert. Det kan avtales at fjerde ledd ikke skal gjelde for småpengeinstrumenter som ikke kan sperres for bruk. For elektroniske penger som definert i finansforetaksloven § 2-4 annet ledd gjelder likevel første til fjerde ledd med mindre institusjonen ikke kan sperre kontoen eller betalingsinstrumentet.

A.2 Lov om straff (straffeloven)

§ 202. Identitetskrenkelse

Med bot eller fengsel inntil 2 år straffes den som uberettiget setter seg i besittelse av en annens identitetsbevis, eller opptrer med en annens identitet eller med en identitet som er lett å forveksle med en annens identitet, med forsett om å
a) oppnå en uberettiget vinning for seg eller en annen, eller
b) påføre en annen tap eller ulempe.

Tilføyd ved lov 19 juni 2009 nr. 74.