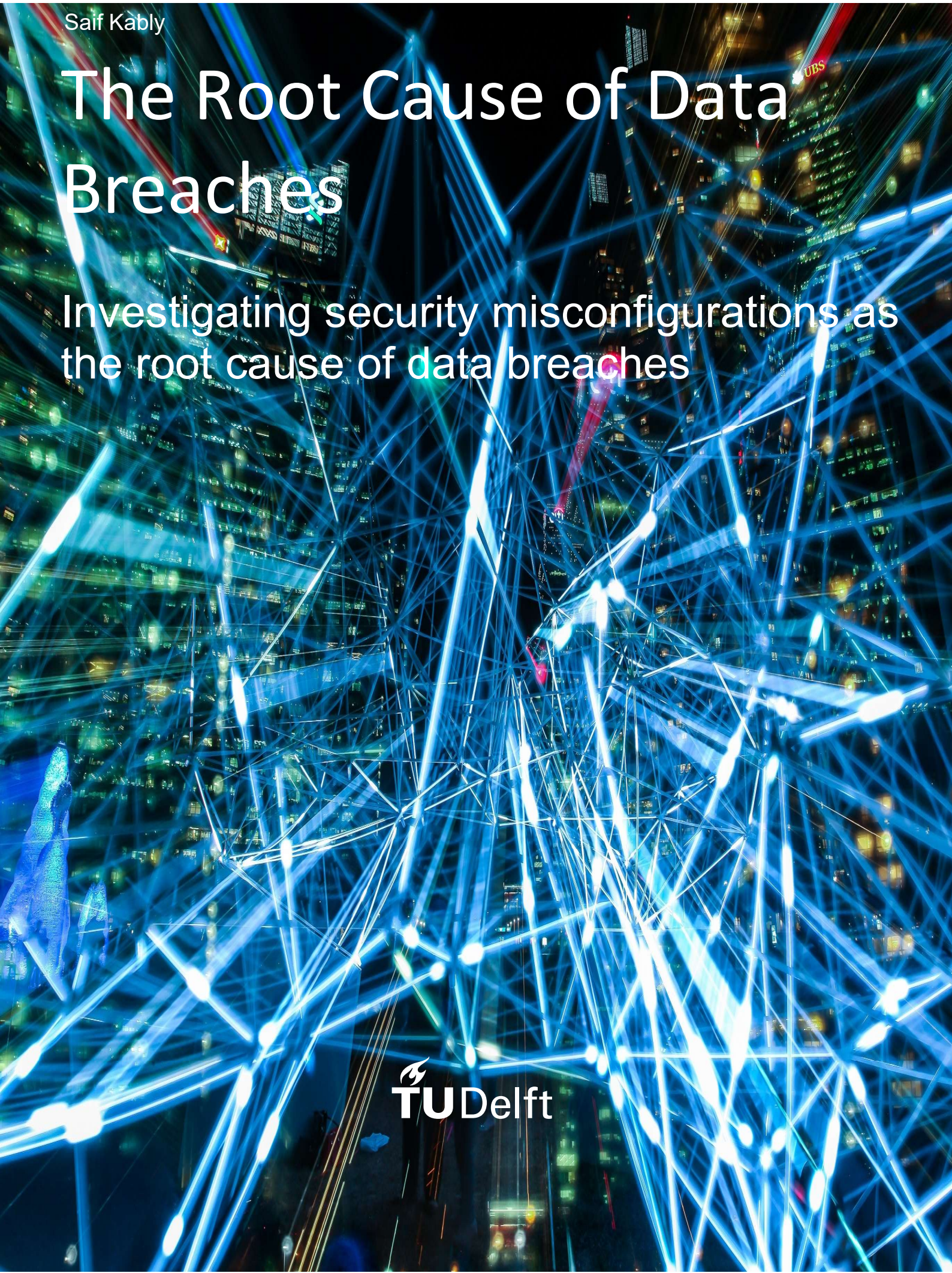
Saif Kably

The Root Cause of Data Breaches

Investigating security misconfigurations as the root cause of data breaches



TU Delft

The Root Cause of Data Breaches

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex Systems Engineering and Management**

Faculty of Technology, Policy and Management

by

Saif Kably

Student number: 1065408

To be defended in public on July 20, 2021

Graduation committee

Chairperson	: Dr.ir. G.A. de Reuver, Information and Communication Technology, TU Delft
First Supervisor	: Dr.-Ing. T. Fiebig, Information and Communication Technology, TU Delft
Second Supervisor	: Dr.ir. C. Hernandez Ganan, Organisation and Governance, TU Delft

TU Delft

Preface

I would first like to thank Dr.-Ing. T. Fiebig for his continuous support and patience. His guidance and feedback were very valuable for the research and writing of this thesis. I am very lucky and thankful that I had the opportunity to have had him as my first supervisor.

I would also like to thank Dr.ir. G.A. de Reuver (chairman) and Dr.ir. C. Hernandez Ganan (second supervisor) for their valuable comments and support.

I cannot express my gratitude enough to this Graduation Committee.

S. Kably
Delft, July 2021

Abstract

In the past decade, the world has experienced numerous severe and impactful data breaches, without indications of this development slowing down. Even worse, research has shown data breaches are still waiting to happen. The occurrence of a data breach has consequences for several involved parties and for society in general. It is therefore only natural that there exists a pursuit to prevent data breaches from happening. Research claims that data breaches happen because of simple and preventable errors made by human, also known as security misconfigurations. This study aims to investigate whether the root causes of severe data breaches are frequently related to security misconfigurations, which would make most data breaches preventable. No such structured research had been done before. We conducted a multiple case study, wherein a number of data breaches was analysed based on publicly available case literature. Assessing the data breaches with the help of our developed framework was part of that analysis, resulting in a systematic characterization of each data breach. The results indicate that in breaches the data are mostly subject to unauthorized access by outsiders, which frequently is made possible by poor security. The organizations directly responsible for that data are large organizations which get breached especially in their storage facilities. Next to the organization which got breached, these sizeable data breaches always affect individuals since at least part of the compromised data is about them or linkable to them. Usually this is not even discovered by the breached organization itself and sometimes only after a long period of time. Ultimately, it can be concluded that data are frequently caused by security misconfigurations and therefore are mostly preventable. On this basis, it is recommended that organizations responsible for sensitive data should be more incentivized to thoroughly combat security misconfigurations, instead of treating IT security as only a technical endeavor.

Contents

List of Tables

List of Figures

1 Introduction

1.1. Problem Introduction

Data breaches are security incidents that result in the compromise of private and sensitive data (Cheng, Liu, & Yao, 2017; Cichonski, Millar, Grance, & Scarfone, 2012; Grance, Kim, & Scarfone, 2004; Sen & Borle, 2015). In the past fifteen years, the world has experienced a large number of data breaches (Ayyagari, 2012; McCandless, Quick, Hollowood, Miles, & Hampson, 2019; Privacy Rights Clearinghouse, 2019). And it does not seem this kind of incidents is going to disappear anytime soon.

The occurrence of a data breach affects the organization which is responsible for the compromised data at the moment the data breach takes place, as well as the entities the compromised data relates to. These entities often concern human individuals but can also concern organizations such as the breached organization itself.

The affected entities become susceptible to threats. If data relating to individuals gets compromised, those individuals become threatened in their personal rights and freedoms. If data relating to an organization gets compromised, that may be damaging to that organization in various ways.

The breached organization, which holds the data and is responsible for them the moment the data breach takes place, also suffers from consequences. The issue needs to be investigated and fixed. Legislation may require the breached organization to disclose the breach and to notify government authorities as well as the victims. The legislation also requires taking measures to remedy the situation. Based on legislation, governments may impose fines and victims may claim compensations.

Nowadays, comprehensive data protection legislation exists in almost every country worldwide (Greenleaf, 2021a, 2021b). Legislation in itself should already incentivize organizations to improve their data protection. And in order to realize improved data protection, proper security should be in place that ideally prevents data from being compromised.

With a data breach becoming publicly known, the breached organization also suffers reputational damage. All the consequences can potentially result in an enormous financial blow to the breached organization, one which it might not recover from.

Because of these negative consequences, organizations, governments and societies in general naturally want to prevent data breaches from happening. However, only organizations that hold private and sensitive data are the ones that have direct influence on preventing data breaches.

In principle, prevention relies on anticipatory attitudes and actions. Therefore, it is useful to identify from an organizational perspective the causes underlying data breaches in order to understand how and why they happened.

More importantly, knowing what causes data breaches, provides insight in how to oppose them. Especially if they have a common denominator, which lies in preventable human-made error. We want to investigate the presence of human error as the root cause of data breaches.

1.2. Literature Review

The human factor in IT security has been widely discussed. We make a selection and proceed with an initial literature review on the human factor in IT security.

Naiakshina et al. (2017) consider a specific security mechanism and how it is handled. They conducted a qualitative research which provides useful insights with regard to the human factor in password storage by developers. They found that security knowledge does not guarantee secure systems. This obviously worsens when security knowledge lags behind the current security landscape. It appears that developers do not feel responsible for security if this is not explicitly mentioned or requested, even if the tasks are evidently related to security. Another finding is that developers focus on functionality first and consider security secondary. In combination with the factor developer misconceptions, this leads to security errors. However, the authors stress that these are preliminary results, based on a small sample, which need to be further studied.

Hauer (2015) created a criteria model to analyse data leakage incidents, mainly those which involve insiders, and found a significant portion to be caused unintentionally. The study shows that the human factor is a critical one in information security. To prevent data breaches, organizations have to rely on their involved members (mostly employees) and their acceptance of and cooperation with the organization' security policy.

Furnell and Clarke (2012) recognize that humans form a critical factor in achieving security, but also that they are often the point of failure. Especially considering the fast growth of sensitive data being utilized for all kinds of purposes, which means this data needs to be collected, stored, transmitted, or processed and the involvement of people therein. Thus, not surprisingly there is an increase in both security threats and technical security controls, which increased the burden on users. They acknowledge the human factor by emphasizing its significance as part of the solution for security issues.

There has been done research into why and how the human factor threatens IT security. Hadlington (2018) examined human factors, including personality factors, which influence security, based on the study of other works. In his research, these factors are related specifically to insiders who do not have malicious intent. The human factor in IT security turns out to be a very extensive and complex subject as the human element in security has multiple facets. Still, the author points out the significant risks the human factors pose to security and explains more understanding, and thus more research, is needed to address these risks. Because although some research already has been done on the relation between on the one hand specific human factors and on the other hand security behaviour and awareness, the understanding is still lagging behind the increasing risks.

The role of humans is emphasized in Tobias Fiebig (2017) and Dietrich, Krombholz, Borgolte, and Fiebig (2018), within the context of security incidents. Tobias Fiebig (2017) addresses that although IT security has been extensively researched and effective technical measures are available, impactful security incidents still occur often. He states that this is because of simple human errors in configuring. He refers to these errors as *"security misconfigurations"* and defines it as having *"… occurred, if the way an Internet service is deployed, i.e., configured, enables an attacker to taint either its confidentiality, integrity or availability, and the property could not have been tainted if the service would have been deployed and configured correctly, without restricting availability for legitimate clients"*.

Dietrich et al. (2018) state that security incidents are seldom the result of complex, sophisticated attacks, but are rooted in humans misconfiguring security. They define a security misconfiguration as an error in system operation which leaves that system exposed to threats to its confidentiality, integrity and availability. In the qualitative part of their research system operators were questioned about what they consider a security misconfiguration. This resulted in a number of security misconfiguration types. Furthermore, the quantitative part of the research brought forward to which extent the participants made or encountered the different security misconfiguration types. In general, they found that almost all participants encountered security misconfigurations, but that these do not always lead to incidents. However, when they do, they have significant impact. Based on the findings, mitigations are proposed.

Xu et al. (2013) claim that users are not completely and solely guilty to misconfigurations. In their work they attempt to make developers feel more responsible, so that they focus on developing software in such a way that operators do not quickly misconfigure security. Basically, this research claims that the secure configuration is not only the responsibility of the people who do the configuration in an organization, but also of the developers that develop the (application) systems which need to be configured.

Research has been done into strengthening the human factor. In Metalidou et al. (2014) the human factor in security is considered to pose a lot of weaknesses. Therefore, the authors seek to overcome this by addressing security awareness. To pursue this, they presented a framework that relates human factors to a lack of security awareness, and subsequently specific ways to raise awareness.

1.3. Research Gap

An security incident is defined as (M. Nieles, K. Dempsey, & V. Y. Pillitteri, 2017):

"An occurrence that actually or potentially jeopardizes the confidentiality, integrity, or availability of an information system or the information the system processes, stores, or transmits or that constitutes a violation or imminent threat of violation of security policies, security procedures, or acceptable use policies."

A data breach is a security incident which results in the compromise of private and sensitive data (Cheng et al., 2017; Cichonski et al., 2012; Grance et al., 2004; Sen & Borle, 2015). Hence, a data breach is a specification of a security incident. This means every data breach is a security incident, but not every security incident is a data breach.

Tobias Fiebig (2017) and Dietrich et al. (2018) claim in their work that security incidents and data breaches mostly happen because of simple and preventable errors and seldom are the result of complex and sophisticated attacks. They refer to these errors as *"security misconfigurations"*. More specifically, with a security misconfiguration they mean the simple human errors in the operation of systems which allow the system's security, in terms of confidentiality, integrity, and availability, to be harmed (Dietrich et al., 2018; Tobias Fiebig, 2017). The use of the term configuration inherently means this includes a human element in these errors, but this was still emphasized by Tobias Fiebig (2017) and Dietrich et al. (2018).

Hence, their claims implicate that most data breaches have security misconfigurations as the root cause. A root cause refers to the most basic reason for an event to occur. ISO standard 18238 defines a root cause as the *"original event, action, and/or condition resulting in an actual or potential undesirable condition, situation, nonconformity or failure"* (ISO, 2015b). Furthermore, their claims implicate that most data breaches therefore are preventable since they deem security misconfigurations preventable.

The human factor in IT security has already been discussed in literature. However, to our knowledge, no structured research has been conducted yet on the specific claim that most data breaches have security misconfigurations as the root cause. In this we see a research opportunity.

1.4. Research Focus and Scope

In this research, we investigate the claim that data breaches frequently have security misconfigurations as root cause, as ensued from the studies by Tobias Fiebig (2017) and Dietrich et al. (2018). A security misconfiguration consists of the technical misconfiguration, which results in a security flaw, and the contributing causes, which include the root cause (Dietrich et al., 2018). This distinction was made to better be able to identify misconfigurations and causes separately in their study.

The contributing causes basically are those which keep the operators from correctly and securely configuring the system they are responsible for (Dietrich et al., 2018). They facilitate the technical misconfiguration and therefore are called misconfiguration facilitators (Dietrich et al., 2018). Based on their own research, Dietrich et al. (2018) formulated three main categories of misconfiguration facilitators, namely systems, operators, and organizational environment.

Because of their approach, we cannot constrain our thinking to technical systems but also need to consider the operational processes and the people who use, operate, interact with, and are influenced by the technical systems and operational processes. As such, we need to carefully consider how the technical system interacts with(in) the broader socio-technical system. Hence, we address their claim in the broader socio-technical system domain, which means that we take a socio-technical perspective in our research. Taking this approach enables us to involve people, policies and processes as part of the system (G. Baxter & Sommerville, 2010; Sommerville, 1998). Furthermore, the socio-technical system perspective includes people making errors (Sommerville, 1998), which corresponds with the essence of the claim we conduct research on. Since security misconfigurations refer to human errors in the operation of systems, we consider the people responsible for the operation of systems, in short, operators.

The systems we focus on concern data breaches of a significant severity or impact. We consider data breaches that happened in the previous ten years. Naturally, there should be sufficient qualitative literature on the data breaches we want to include, otherwise we cannot assess them. Inherently, a data breach happens to an organization which was responsible for the compromised data, by holding or processing these data at the moment of breach. These organizations can be based everywhere in the world, and it does not matter in which sector they operate. Every organization that suffered a severe, impactful data breach in the last decade and on which there is sufficient relevant literature can potentially be included in our research. To identify or exclude security misconfigurations, we study the data breach within these organizations from an operational perspective, within a broader socio-technical perspective.

1.5. Research Objective

With this research we intend to study the claim that data breaches frequently have security misconfigurations as the root cause. Therefore, we need to investigate the extent to which security misconfigurations are the root cause of data breaches. Hence, our research objective becomes the following:

to identify to which extent root causes of data breaches are related to the configuration of IT infrastructure, by analysing multiple cases of significant severity in which sensitive data was exposed.

Since no earlier structured research has been done yet which investigates the claim that follows from Tobias Fiebig (2017) and Dietrich et al. (2018), we deem it sufficient to keep this research descriptive by analysing data breach cases and in each case focus on determining whether the root cause concerns a security misconfiguration or not. We do not intend to make any further classifications.

A security misconfiguration does not necessarily always lead to an actual incident, i.e., does not always lead to a data breach. On the other hand, data breaches are a materialization of threats that manifest themselves. These threats may have been present because of a security misconfiguration, which according to Tobias Fiebig (2017) and Dietrich et al. (2018) is often the case.

Hence, we can use a data breach to identify whether a security misconfiguration was present, based on an assessment of that data breach. After studying a significant number of data breach cases, we will be able to induce whether security misconfigurations as root causes for data breaches are more likely than other root causes.

The research is qualitative and is structured as a multiple case study. Only by studying multiple cases can we obtain more insight in whether data breaches actually do frequently have security misconfigurations as a root cause, based on appearance frequency.

1.6. Research Questions

Based on the research gap and the research objective, and within our scope, our main research question is formulated:

Main Question

Are there common root causes of severe data breaches, and are these preventable?

By answering this question, we determine if root causes of data breaches mostly are security misconfigurations. This enables us to establish whether data breaches are frequently preventable.

To be able to answer the main question more effectively, we first want to gain insight in known types of security misconfigurations which may cause data breaches. We can use these types when assessing data breaches. For each data breach case that we include in our study we strive to determine as closely as possible which of the known security misconfiguration types is applicable to the concerned security misconfiguration present in that specific case. Even if we cannot determine whether one of the known security misconfiguration types is applicable, knowing them helps us recognize whether the cause of a data breach can be attributed to a security misconfiguration or not. This supports us in determining the presence of a security misconfiguration in a data breach. Hence, our first sub question is:

Sub Question 1

What are known types of security misconfigurations that can lead to data breaches?

Identifying whether a security misconfiguration is present in a data breach case takes place as part of the assessment each case undergoes. This assessment of data breaches needs to take place in a structured and effective manner, and every included data breach case needs to be assessed in an equal manner. For these purposes, we formulate an assessment tool which we apply on each selected data breach case. Furthermore, this assessment tool includes factors for the purpose of descriptive analysis. Therefore, we formulated the following sub questions:

Sub Question 2

How can we assess data breaches in a structured manner?

Sub question 3

What are the key parameters and metrics to assess data breaches?

By answering the above questions, we obtain the elements for the assessment tool. All these elements work together to provide us with insight into the assessed data breach case and eventually help us determine the root cause. The results of all assessments then show whether most data breaches are caused by security misconfigurations. Our final sub question becomes:

Sub Question 4

Based on our assessment, are most of the severe data breaches preventable, and are there any other outstanding factors?

Indeed, when a significant majority of the data breaches that are to be studied turns out to be the result of a security misconfiguration, we consider most of the severe data breaches to be preventable.

1.7. Relevance

In the introduction we mentioned that a lot of data breaches have happened in the recent past, and still are occurring with no signs of disappearing soon. Even if the technical tools to prevent and to mitigate data breaches are available (Tobias Fiebig, 2017). Since the tools that enable organizations to prevent data breaches are available (Tobias Fiebig, 2017; Hauer, 2015), some other factors cause the data breaches still to happen. This research intends to establish the nature of those root causes.

This is relevant because of the significant impact that data breaches can have on society and especially the involved parties. Until recently, scientific literature on computer security focused mostly on advanced and complex attacks as the method with which private and sensitive data is compromised. Indeed, solutions for these attacks are also well researched and are mostly technical of nature (Tobias Fiebig, 2017).

The human factor in IT security has been widely discussed and recognized in literature (Furnell & Clarke, 2012; Hadlington, 2018; Kraemer, Carayon, & Clem, 2009; Metalidou et al., 2014; Pieters, 2013). Addressing the human factor proves to be very challenging because of its extensive, complex and multifaceted character (Hadlington, 2018).

This research intends to emphasize this human factor from an operator perspective and contribute to providing a point of attention for those operators who are responsible for IT security in organizations. Specifically, we address the claim ensuing from Tobias Fiebig (2017) and Dietrich et al. (2018) which states that data breaches mostly occur because of simple, preventable errors. To our knowledge, this has not happened before. As such, this research contributes to addressing the human factor in IT security from a unique angle and with a structured approach.

Since on the one hand research shows that nearly every system operator has encountered security misconfigurations and on the other hand security misconfigurations do not always lead to incidents – or more specifically, data breaches -, the assumption is that a large number of data breaches is lurking (Dietrich et al., 2018). Our research produces insights in whether data breaches generally can be considered preventable, and if so, where to put most focus on to improve preventability. This makes our study also practically relevant.

1.8. Document Structure

This master thesis document is structured in eight chapters. Chapter 1 and 2 respectively contain the introduction to the research and the research methodology. Chapter 3 contains the necessary background to identify whether a security misconfiguration is the root cause in a data breach. Chapter 4 treats the framework that we use for data breach assessments. Chapter 5 contains the assessed data breach cases, including a relevant description for each case. Chapter 6 shows the results of the assessments. Chapter 7 contains the discussion and limitations. In chapter 8, the conclusion is given. The last part contains the bibliography, which includes all the sources used during the research process.

2 Methodology

2.1. Approach

We investigated the question whether there are common, simply preventable root causes of severe data breaches. This came forward from the claim that data breaches frequently have security misconfigurations as root cause, which ensued from (Dietrich et al., 2018; Tobias Fiebig, 2017). We explored this under-researched topic and systematically described the characteristics of multiple data breaches, in order to assess whether the root cause of data breaches concerns a security misconfiguration. For this, we needed to study a number of data breaches, so we used the method of multiple case study to perform qualitative research.

Within our case study, we used document analysis. Each data breach case was studied and assessed based on qualitative data from primary and secondary sources. We kept our research descriptive and collected these data without intervening or influencing the studied phenomenon. Because we studied a number of data breaches, descriptive analysis was possible.

2.2. Case Study Justification

To answer our research question, we needed to find out whether data breaches frequently have security misconfigurations as the root cause. However, this meant that for a specific data breach we needed to investigate how and why that data breach happened. But doing this for only one data breach case, would not answer our research question. Hence, it was necessary to assess the how and why for multiple cases. Thus, we investigated the how and why in more than one data breach.

When conducting explanatory research and investigating contemporary events without having influence on behavioral events, the case study method is preferred (Yin, 2013). We focused on data breaches. Within our scope fall data breaches which happened in the last ten years. These are contemporary events which we had no influence on. It is not able to realistically simulate data breaches in a controlled environment for the purpose of what we want to achieve with our research.

Furthermore, we wanted to consider data breaches including their context, which are the technical, organizational, and operational settings and the involved people. Under these settings the data breach could happen and to obtain a complete accurate characterization of that data breach, the context within which it occurred need to be considered. According to Yin (2013), when relevant contextual conditions need to be considered, then case study design is suitable.

Multiple case study allowed us to study the data both within as well as across cases (Yin, 2013). With multiple case study, we were able to study multiple cases to understand the differences and similarities between them (Gustafsson, 2017). This understanding contributed to generating new insights and knowledge in relation to our topic, while evidence generated by multiple case study is considered strong and reliable (Gustafsson, 2017). We wanted to understand specific aspects of data breaches by systematically studying and describing them. The method of multiple case study allowed us to explore key characteristics and implications of multiple data breach cases.

2.3. Case Study Design

Case study research is defined by Yin (2013) as *"an in-depth empirical inquiry about a contemporary phenomenon set within its real-world context"*. Creswell and Creswell (2017) define case study research as *"a qualitative approach in which the investigator explores a real-life, contemporary bounded system (a case) or multiple bound systems (cases) over time, through detailed, in-depth data collection involving multiple sources of information, and reports a case description and case themes"*. We see that Creswell and Creswell (2017) define a case as real-life, contemporary bounded system. In the article by P. Baxter and Jack (2008), a case is defined as *"a phenomenon of some sort occurring in a bounded context"*.

We needed to establish what a case entails in our research, as the bounded system on which we focused, and which became our unit of analysis (Yin, 2013). This helped to further define the topic and its context (Yin, 2013). But it starts already with the research question, which should be clearly and well-defined (P. Baxter & Jack, 2008; Yin, 2013). Together with the research objective and scope, we were able to determine the case boundaries, as well as establish the depth of our research (P. Baxter & Jack, 2008). Our unit of analysis during the study was a severe data breach that happened in the past ten years.

Because we focused on how and why, our research is considered explanatory as well as descriptive. Not only did we use our case study to systematically describe data breaches in the real-life context in which they occurred (P. Baxter & Jack, 2008; Yin, 2013), but we also sought to confirm a presumed causality that was too complex for other research strategies (P. Baxter & Jack, 2008; Yin, 2013). We discussed this causality in chapter 1, as the claim ensuing from the works of (Dietrich et al., 2018; Tobias Fiebig, 2017).

The process used during our multiple case to study the workflow for the purpose of data collection was as follows:

1. Select a severe data breach case.
2. Find relevant data on that case. If not available, that case is discarded.
3. Study and analyze the found relevant primary and secondary data.
4. Systematically assess this case based on collected case data by means of an assessment framework. If this does not lead to satisfactory assessment, case is discarded. The framework is discussed in Chapter 4.
5. Write the case description.
6. Continue with the next case.
7. Repeat this process was repeated until saturation is reached.

The order of these process steps and the distinction between them were not carved in stone. Step 2 to 5 could take place with overlap. Also, this should be seen as an iterative process.

2.4. Case Selection

We conducted research on severe data breaches that happened in the last ten years, which was our population. The online database 'World's Biggest Data Breaches & Hacks' is the sampling frame where we drew our sample from (McCandless et al., 2019). This service keeps track of the biggest data breaches that become known, including one or two links to sources this service used for each data breach. At the time of writing this thesis report, the sampling frame listed 373 data breaches (McCandless et al., 2019).

To select cases to include in our research, we used non-probability sampling methods. This means data breaches were selected based on non-random criteria, which is often used in qualitative research. In our qualitative research it is not the aim to test hypotheses about a broad population of data breaches, but to develop a deeper understanding of data breaches within an under-researched perspective.

Our sampling method was based on both convenience sampling and purposive sampling. Convenience sampling because through the earlier mentioned online database, data breach cases are easily accessible. We could not find the criteria that service uses for including large data breaches, but their description suggest they include every large data breach that becomes known to them (McCandless et al., 2019). Hence, although they did not further specify what large means to them, we considered each data breach in the sampling frame to be large and therefore severe.

We used the knowledge we have of data breaches to select cases to be included in our sample in such a way that is useful for the objective of our research. The term data breach is used comprehensively in our research and includes various aspects. Therefore, each data breach case was judged on whether the collected case literature sufficiently considered our aspects of a data breach and whether that case literature made possible a useful assessment. Furthermore, we wanted to reach heterogeneous or maximum variation sampling, meaning that we strived to select data breaches with diverse characteristics. Additionally, we wanted to be able to select critical cases, enabling us to include severe data breaches with a significant societal or dramatic impact. All this explains why and how we applied purposive sampling during our research. In practice, our application of purposive sampling meant that it was also possible that we conveniently selected a data breach from our sampling frame, which we later would again discard because the case literature did not suffice.

Our sampling continued until we reached saturation. At 43 cases we established that studying another case (and more cases after that) was not likely to produce new insights. Especially not when we consider the characteristics of the typical population of data breaches.

2.5. Case Selection Limitations

When explicitly selecting cases for a research, inherently there also were data breaches which were not included. Even if this selection is based on non-probability sampling, leaving data breach cases with unequal opportunities to be selected. But besides the cases which were explicitly not included because they were not selected, there could also be cases of data breaches that were not included because they were not observable or for which there is no evidence. And that is because those data breaches have not been disclosed or even discovered. This also means that such data breaches are not in our sampling frame and have no possibility of being selected at all, compared to the data breaches which are listed in the sampling frame and have (more or less) chance of being selected.

An important factor determining whether data breaches are disclosed is regulation. Several data breaches were only disclosed because the breached organizations were obliged by law. Would those organizations not be obliged by law, part of them would likely have covered up the data breach. Furthermore, different countries or regions can have different disclosing requirements and criteria. This leads to different locations varying in the number of data breaches being disclosed, only because of regulation differences.

We did not further investigate the role of regulation as we consider this to fall outside of the scope of our research. Including this may quickly become too extensive and complicated. Because an organization does not only fall under regulation of the country in which it is based, but may also be subject to regulation of the countries where the people, whose personal data was breached, have their citizenship. And on the other hand, our sample also includes data breaches which were not disclosed because of regulation. Hence, we did not assess whether the disclosure of each data breach was because of which regulations.

As a result of the above, surely there are data breaches that have not even been considered to be included in our research. A concern may be how much the data breaches that were not included in our research as a case because of that, affected our research and the significance and reliability of its outcomes.

Furthermore, the sampling methods we used may not be statistically representative. Not every eligible data breach in the target population has the same opportunity to be included in our case study (Etikan, Musa, &

Alkassim, 2016). In their article, Etikan et al. (2016) present about this: *". . . study results are not necessarily generalizable to the population . . ."* This of course also then applies to our research.

Etikan et al. (2016) continue to present: *". . . [I]n purposive Sampling, subjects are selected based on study purpose with the expectation that each [case] will provide unique and rich information of value to the study."* When a selected case indeed offers such valuable information, it supports our analysis and understanding of data breaches within their natural context. Otherwise, we discarded the case.

With regard to subjectively selecting cases by using purposive sampling, Suen, Huang, and Lee (2014) claim: *"As a result, members of the accessible population are not interchangeable and sample size is determined by data saturation not by statistical power analysis."*

We realize that employing a non-probability sample may lead to a higher risk of sampling bias. Furthermore, we realize that because of that, our sample is not a good representative of the population. For the assessment we conducted in our study, we used the earlier mentioned sampling frame, which is a subset of the entire data breaches population. Even if every assessment is reliable, the evidence we have for interpreting the data may be limited to the extent that the selected data breaches are considered not to adequately reflect the whole population (Daniel & Onwuegbuzie, 2002). This is a concern regarding adequacy of evidence which reminds us that reliability is not the same as validity (Daniel & Onwuegbuzie, 2002).

A way to assess this influence is possible, although ex post. After our research, another researcher should conduct a research with the same research objective, research questions and methodology as in our research. Preferably there should be selected other data breach cases. If that yields results the meaning of which can be interpreted in the same way, this strengthens the reliability of the results. If the same research would be conducted by another researcher but that research would study the same cases as we did, and this gives the same results, than this would confirm the reliability of our methodology.

Furthermore, we must not forget the main objective of this research, which is to find root causes related to human error. Finding these and making inferences as to whether security misconfigurations as root causes are more common than other root causes, ultimately enables the prevention of data breaches. Not including non-observable cases does not affect this negatively, as having more insight in the root causes will at least contribute to preventing data breach cases which are similar to the cases that actually happened and were included in our research.

2.6. Case Data

The characterization of data breaches and the assessment of root causes was based on qualitative data. Therefore, for each data breach case we collected both primary and secondary data. Primary data consisted of information provided by breached organizations with regard to their internal data breach investigations. In some cases, a governmental authority was also involved in the investigations into a data breach, or was the sole investigator, and brought out its own report of the data breach. We consider these to be data from primary sources because these sources are closest to the phenomenon being studied. Secondary data consisted of information on data breaches as described, interpreted, analyzed, or evaluated by the likes of researchers, reporters and journalists based on the primary data.

As we studied data breaches that happened in the past, we were dependent on the primary and secondary sources for our research data. With regard to primary research data, we did not have access to breached organizations networks to investigate data breaches and gather this information directly ourselves. Furthermore, this would require specific expertise outside the scope of our research. Also, as we conducted a multiple case study, this would not have been possible within the time and research constraints of this research project. Hence our use of primary data as released by the breached organization is justified.

Enforced by legislation, government authorities sometimes did get access into breached organizations, enabling them to conduct their own investigations while requiring the breached organizations to cooperate. In those cases, and when made available, we were able to use the information released by those government authorities as primary data too. Because governmental authorities are supposed to act as neutral outsiders, protecting the rights of all entities involved in a data breach, they do not serve a particular party's interest and therefore their data can be considered more trustworthy.

The use of secondary data was useful. It provided insightful interpretations of the primary data by experts and experienced people in the field. Also, it could be used to judge the trustworthiness of primary data. Because when breached organizations released information about a data breach they suffered, they tended to do this in a strategic way, i.e., in a way that is least harmful to them in their perspective. This resulted in the released information often being manipulated to a certain extent. Furthermore, secondary data is useful to check the trustworthiness of other secondary data.

Considering we used both primary and secondary sources for our data, we could base our research on evidence as well as reasoning, since in principle primary and secondary sources respectively provide evidence and reasoning. Although generally primary sources provide data that functions as evidence and secondary sources provide data that interprets the evidence, in our research that distinction is not a hard one. Evidence and interpretations are sometimes intermingled. Because the occurrence of a data breach can be confirmed in various ways during and especially after it took place, secondary sources sometimes were also capable of showing evidence of a data breach. This meant the secondary data did not always needed to be based on primary data sources, and even that some cases only have secondary sources.

Because we used several data sources for each case, we were able to apply structural corroboration to support our interpretations of the data breaches, and to how and why they happened. Structural corroboration was proposed by Eisner (2017) as a synonym for triangulation and used *"to describe the confluence of multiple sources of evidence or the recurrence of instances to support a conclusion"*. This supported the trustworthiness of the data and the reliability and overall credibility of our qualitative research.

We mentioned earlier that secondary data consisted of information on data breaches as described, interpreted, analyzed, or evaluated by the likes of researchers, reporters and journalists. By using secondary data from such various sources, and where different sources on the same case showed agreement in the description, interpretation, analysis and evaluation of that case, we were able to establish consensual validation and contributed to the accuracy of assessing data breaches. Eisner (2017) defined consensual validation as *"agreement among competent others that the description, interpretation, evaluation, and thematics of a ... situation are right"*. These also include the analysis and assessment of a situation. To reach consensual validation, we needed to compare case literature from different sources. Achieving consensual validity contributes to trustworthiness and credibility.

Because of our approach, we believe we achieved reliability. Furthermore, after the design of our assessment framework, we tested it on a few cases. These cases also became part of our eventual data set, which means the framework was applied to them again during the actual research. This produced the same outcome for these cases.

The assessment framework has been formulated based upon what a data breach entails and is designed to describe all data breach specifics. This is further explained in chapter 4. As such, the framework measured all relevant aspects of the concept of data breaches. Because of this, the use of the framework respects research validity.

2.7. Data Sources

The main data source were electronic media in which publications were made with regard to data breach cases. These were found through searches on the Internet. We would start by trying to find data on a data breach through Google Scholar. If we were to find something, this would mostly concern scientific literature treating the data breach from some kind of perspective. Sometimes this was useful, if it included a relevant description of all or certain aspects of the concerned data breach.

From there, we continued to look through normal Google. Here we first searched for official statements made by the breached organizations themselves. Then we searched for governmental publications on the concerned data breach, which could be an investigation report or a public notification.

During the case study process, we learned that certain platforms regularly popped up as media coverage channels when searching for data on data breaches. They would often have useful contributions. Notably, Brian Krebs blog, ZDNet, Security Week, Infosecurity Magazine, CSO Online, SC Media and Data Breach Today. Of course, there were also other online newspapers and online magazine articles.

In some cases, the data breach would be treated by security researchers or experts which may or may not have done that on a personal title. Indeed, when they do this in name of a company they own or work for, this has commercial motivations.

2.8. Limitations

Being disclosed and publicized is what we used as evidence the data breaches occurred. However, we were aware that the use of public sources can have some limitations. When a data breach gets publicized based on the statements of the organization that was responsible for safeguarding the data, that organization may have released biased information. Some information they probably did not want to disclose or disclose it in the least harmful manner. There might be organizations that do not release any statement or even deny that a data breach occurred. In those cases, the information supply could come from people like journalists, researchers, the attackers, or the ones that detected the data breach. This information might also be biased, for instance when hackers release information with a certain agenda.

Furthermore, except the breached organization (and in some cases not even them) and the attacker (if there is an attacker), no one can be sure how the data breach came to be. These actors do not always like to share this information. The organization that was breached sometimes released limited and undetailed information, which is interpreted and analyzed by the people reporting on the data breach.

Despite the above, we have the opinion that public sources were sufficiently trustworthy and useful for our research. Especially in the case of severe and impactful data breaches. Data breaches have victims and that is why often they are made public. These victims mainly concern the subjects to whom the data pertains, although the organizations who were responsible for the data can also be considered victims to some extent. However, the public does not view them that way.

In the case of severe data breaches, there is an impact and a significant number of victims. And severe data breaches were where our attention went out to. In those cases, there will be a lot of pressure and attention from the relevant authorities, the public and the media on the organizations to release sufficient and trustworthy information concerning the data breach. Organizations will have a very hard time escaping that pressure and will have to comply to a certain extent.

Besides, we strived to use multiple sources for each data breach case, which offered a way to cross-check and evened out a part of the bias. Of course, we realize that various publications on a data breach could have the same original 'root' source, which might be a subjective and biased. That is why we searched for a many different sources as possible. Here we also applied the saturation principle: the moment we realized new sources did not provide new relevant case data, we stopped searching.

Another way to counter the 'root' source problem, was to remain critical during the assessment. When noticing the bias, we would lean into an assessment of that case that is the least positive for that breached organization. Especially when we noticed strategic behavior in the way they treated and disclosed the data breach. Because the 'root' source would mostly concern the breached organization, and they are inclined to take the least amount of responsibility for the data breach that happened under their roof.

3 Background Literature

In this chapter we treat scientific literature that serves as a source of relevant information for our research. We gain insight in the known types of security misconfigurations, which support us in recognizing their presence in data breach cases that we study.

These security misconfigurations types were intentionally left technical by Dietrich et al. (2018) to avoid blurry descriptions. We strive to recognize one or more of these security misconfiguration types in the data breach and from there identify the causes that contributed them. Hence these security misconfiguration types are to be seen as the technical mistakes. And if we identify the cause of the technical mistake to be a simple human error, we can actually speak of that data breach being caused by a security misconfiguration.

3.1. Technical Background

In the previous section we mentioned the security misconfiguration types brought forward by Dietrich et al. (2018). While the types of security misconfigurations they encountered in their research are rooted in human made errors, they are still technical of nature. We listed these security misconfiguration types in Table 1 below, and we will discuss them from a more technical perspective in this section.

Security Misconfiguration Type (SMT)	Example
Authentication	Faulty, weak or missing identity verification
Authorization	Incorrect assignment of privileges
Passwords	Weak or shared passwords
Encryption	Bad SSL/TLS settings
Firewalls	Missing isolation, disabled firewalls
Updates	Delayed or missed updates
Storage	Backups on the same drive as the productive system
Deployment	Publishing extended log files and version information
Integration	Insufficiently separated systems, not adapting configuration to new systems
Scripting	Faulty automation stalling system components

Table 1 Security Misconfiguration Types (Dietrich et al., 2018)

Authentication is defined by ISO (2018) as a *"provision of assurance that a claimed characteristic of an entity is correct"*. It is one of the pillars of Information Assurance, the discipline which next to confidentiality, integrity, and availability intends to achieve authentication and non-repudiation (Cherdantseva & Hilton, 2013). Shabtai, Elovici, and Rokach (2012c) define authentication as *"the task of verifying the identity of users who connect to a computerized system"*.

The IETF standard RFC2196 (Fraser, 1997) states that *"Authorization refers to the process of granting privileges to processes and, ultimately, users. This differs from authentication in that authentication is the*

process used to identify a user. Once identified (reliably), the privileges, rights, property, and permissible actions of the user are determined by authorization. Explicitly listing the authorized activities of each user (and user process) with respect to all resources (objects) is impossible in a reasonable system. In a real system certain techniques are used to simplify the process of granting and checking authorization(s).” Based on this statement, Jøsang (2017) defines authorization as the specification of policies, while access control is the enforcement of access policies. We assume that when system operators where asked about what they consider a security misconfiguration and responded with authorization (Dietrich et al., 2018), this includes access control as defined by Jøsang (2017).

We saw in the MongoDB case that this led to data breaches. The customer databases turned out to be openly accessible through the Internet because authentication and authorization were missing. Basically, it was a simple fault leading to significant impact.

The use of passwords serves as a security mechanism for access, used in authentication and authorization processes. It could be that this security mechanism is not installed according to standards, which would for example allow weak, simple passwords. Or password are not always installed securely (Naiakshina et al., 2017). On the part of end-users, in certain circumstances people also like to use a standard or shared password. All these present threats to information security.

Encryption is a method to protect confidential and sensitive data (Shabtai, Elovici, & Rokach, 2012a). M. Nieles, K. Dempsey, and V. Pillitteri (2017) state that *“Encryption transforms intelligible data, called plaintext, into an unintelligible form, called cipher text. This is reversed through the process of decryption”*. Thus, even if data gets exposed, lost, or stolen, encryption stops unauthorized access. Encryption does not prevent data breaches however, but is more a safeguard in order to mitigate the risk that compromised data poses. For example, when a network has weak or missing SSL/TLS settings, this negatively influences the ability to set up a secure connection between a client and server, leaving the data transmitted over that connection unprotected.

A firewall is defined by M Nieles et al. (2017) as *“A gateway that limits access between networks in accordance with local security policy”*. Basically, a firewall decides which traffic is allowed or not. Thus, when a firewall is not set up correctly, this poses a data breach threat, since the network traffic is not checked to be able to block it on the basis of the firewall’s policy.

When vulnerabilities of operating systems and other software become known, updates and patches are created and distributed, to repair the problem and prevent abuse of the vulnerability. Not updating and keeping the systems outdated, maintains known security weaknesses which can be exploited. The Equifax data breach is an infamous example of what a forgotten update could lead to (Brandom, 2017).

ISO (2015a) defines storage as a *“device (3.14), function, or service supporting data entry and retrieval”*, and a device as a *“mechanical, electrical, or electronic contrivance with a specific purpose”*. In the context of a data breach occurring in that storage, where the productive system houses as well as the backups are kept, now all that data is threatened, possibly making it very hard to retrieve the valuable original data.

Deployment contains *“the whole operational environment controlled and/or utilized by the operator”* (Tobias Fiebig, 2017). This term is comprehensive, and deployment being a security misconfiguration type might include any of the earlier discussed security misconfiguration types in this section, including the influence they can have on each other. However, we assume Dietrich et al. (2018) considered this already and do not include any of the other security misconfiguration types with the Deployment type.

Deployment is the *“phase of a project in which a system is put into operation”* (ISO, 2017b). It concerns the actual rolling out of a system into a live environment with real users. Fiebig (2017) states that deployment contains *“the whole operational environment controlled and/or utilized by the operator”*. The term deployment is comprehensive, and deployment being a security misconfiguration type may include any of the earlier discussed security misconfiguration types in this section, including the influence they can have on each other.

However, we assume Dietrich et al. (2018) considered this already and do not include any of the other security misconfiguration types with the Deployment type.

To determine whether in a data breach case the security misconfiguration falls under the type of deployment, we decide as such when the way a certain system is deployed is the applicable security misconfiguration leading up to the data breach, and this security misconfiguration does not fall under one of the other security misconfiguration types.

With regard to the security misconfiguration type Integration, this could refer to systems which are not sufficiently separated, for example Internet and intranet (Dietrich et al., 2018). Threats from the Internet possibly could more easily reach the internal network. Also, when updating the operational environment by adding new systems but not adjusting the configuration accordingly, this might result in security weaknesses (Dietrich et al., 2018).

ISO (2017a) defines a script as a *"set of graphic characters used for the written form of one or more languages"*. Schwalb (2003) states that *"a scripting language is a programming language that is used to manipulate, customize, and automate the facilities of an existing system"*. So, it is useful to automate certain tasks, especially repetitive, complicated and intensive tasks. The use of scripts can also introduce security vulnerabilities however, for example when the focus is too much on functionality and comfort and overall security is not considered sufficiently.

4 Assessment Framework

In this chapter we explain the descriptive characterization model that is used for our assessment of data breaches. This answers research questions 2 and 3. The actual assessment of data breaches by using this framework will eventually support the answering of research question 4.

The assessment of data breaches has been treated in literature before (Ayyagari, 2012; Tobias Fiebig, 2017). Furthermore, data breach assessments often focus on assessing the severity of data breaches in terms of their impact on the breached organization, data subjects, other involved parties, as well as impact on the environments they operate in (Galan Manso, Górniak, & European Network and Information Security Agency, 2013; Greve, Masuch, & Trang, 2020; Stiennon, 2013). In fact, that focus is precisely because of the significant impact that data breaches can have (Galan Manso et al., 2013; Greve et al., 2020; Stiennon, 2013).

Stiennon (2013) states that data breach severity depends on numerous factors. Hence, data breach assessments potentially involve a lot of factors. However, which factors are included in an assessment methodology as parameters depends on the objectives of the party which formulates that methodology or does the assessment, together with the demarcation and definitions that party employs.

We have looked at several different instances of data breaches assessment methods; this concerns literature in which a methodology is formulated, or sources in which such a methodology is applied to one or more data breach cases, or wherein both the formulation and application happens (Ayyagari, 2012; Galan Manso et al., 2013; Greve et al., 2020; McCandless et al., 2019; Stiennon, 2013). This provided us with a collection of factors which influence the severity of data breaches, or in another way are relevant for the assessment of data breaches. We made a selection from that collection and continued to formulate the specific parameters for our assessment framework. Particularly these formulated parameters are utilized in our assessment framework because they can be used for the assessment of every data breach case that meets our case selection criteria.

These parameters are used to establish which of their values are applicable in each data breach case that we study. The assessed values will provide us with a sufficiently in-depth understanding of what occurred in which way for each data breach. Hence, applying our assessment framework to a data breach supports the analysis and assessment of that data breach, eventually resulting in the final assessment of wherein the root cause of that data breach lies.

In the following sections we will describe and explain all the parameters of our assessment framework. These sections are divided like this to each represent a core element of a data breach and provides us with a practical distribution of the parameters. In the tool used for inputting the parameter values, this also gives an orderly overview.

The first section consists of the parameters that are used to specify the data breach, in terms of the breach type, the breach method, and the breach actor and its motive. Section two elaborates on the parameters related to the description of the breached organization, in particular the organization's sector, size, headquarters, breached component, and responsibility regarding the data. The third section explains the parameters that are applied to specify the compromised data, in terms of the data subject, the data amount, the data type, the data state, and the data necessity. Section number four treats the parameters that are used to describe the detection of data breaches, namely data breach detection and the time between breach and detection. Finally, the last section explains the parameter that is used to establish a root cause for the data breach.

4.1. Overview

In table 2 we present all the parameters from our assessment framework in an orderly overview, including the section in which the concerned parameter is explained.

Parameter Category	Parameters	Section
Data Breach Specification	Data Breach Type	4.2.1.
	Breach Method	4.2.2.
	Breach Actor	4.2.3.
Breached Organization	Organization Sector	4.3.1.
	Organization Size	4.3.2.
	Organization Headquarters	4.3.3.
	Organization Component	4.3.4.
	Data Responsibility	4.3.5.
Data Specification	Data Subject	4.4.1.
	Data Amount	4.4.2.
	Data Type	4.4.3.
	Data Sensitivity	4.4.4.
	Data State	4.4.5.
	Data Necessity	4.4.6.
Data Breach Detection	Data Breach Detection	4.5.1.
	Time between Breach and Detection	4.5.2.
Root Cause Assessment	Root Cause	4.6.

Table 2 Overview of the parameters from the assessment framework

4.2. Data Breach Specification

This section shows the parameters for describing what kind of data breach we are dealing with, in which manner the breach took place, by whom and with which intent. In order to do this, three parameters are going to be used, namely Data Breach Type, Breach Method, and Breach Actor, which each are explained in their own subsection. Assessing data breaches by utilizing these parameters improves our understanding of the mechanisms in those specific data breaches.

4.2.1. Data Breach Type

This parameter is used to get insight what happened in the breach, by distinguishing between types of data breaches we could be dealing with. In order to formulate a practical parameter and its possible values, we use a specific definition of a data breach.

In ISO/IEC 27040, an International Standard related to security technique and published by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), a data breach is defined as (ISO, 2015a):

"a compromise of security that leads to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to protected data transmitted, stored or otherwise processed."

From this definition we derived the formulation of our parameter *Data Breach Type*. We used the different identifying characteristics in that definition, which are used to determine whether a compromise of security indeed is a data breach, as values for our parameter.

Table 2 shows the values and their definition. On their part, these definitions include their own identifying characteristics. By means of each definition, we can decide for concerned value applies to the data breach being assessed. If we cannot establish the presence of one of these breach types, the value Inconclusive is used.

These values are not mutually exclusive, meaning multiple breach types can be applicable to a single data breach, since a data breach can lead to different ways in which data get compromised.

Parameter Values	Definition
Destruction	*"Result of actions taken to ensure that media cannot be reused as originally intended and that information is virtually impossible or prohibitively expensive to recover"* (ISO, 2015a).
Loss	*"Data loss is essentially the process that results in data being deleted, corrupted, or otherwise made unreadable by users, software, and other applications"* (Techopedia, 2019).
Alteration	Data has been changed and is no longer intact. This means data integrity has been affected.
Unauthorized Disclosure	*"An event involving the exposure of information to entities not authorized access to the information"* (Barker, Barker, Burr, Polk, & Smid, 2007).
Unauthorized Access	*"A person gains logical or physical access without permission to a network, system, application, data, or other resource"* (Stouffer, Pillitteri, Lightman, Abrams, & Hahn, 2015).
Inconclusive	None of the above values could be identified

Table 3 Values for the parameter Data Breach Type.

4.2.2. Breach Method

This parameter is used for specifying how the breach occurred, which could by way of a procedure, technique, action, or inaction. The values for this parameter are a slightly adapted version of the tactics that can be used in an attack or the errors that can be made, as mentioned in the report by (Widup, Spitler, Hylender, & Bassett, 2018). We will define and discuss each of the parameter values in its own subsection below. We also provide examples of manifestations of these values in data breaches. These parameter value explanations enable us to establish whether the respective parameter (i.e. breach method) is applicable to the data breach in question. Multiple of these values can be present in a single data breach and therefore they are not mutually exclusive.

4.2.2.1. Malicious Software

Malicious software, also known as malware, concerns any code added to, adjusted in, or removed from a system, and which is intentionally used to influence or misuse a system in such a way that it causes harm or subverts the intended function of the system (Namanya, Cullen, Awan, & Disso, 2018). Leading malware into a network can have various effects, depending om the malware's design intent as well as on the layout of the network (Namanya et al., 2018). Malware can be classified on the basis of its execution characteristics, payload, exploiting method, method of making the system vulnerable, and propagation method (Namanya et al., 2018). This results in the following subdivision of different types of malware (Namanya et al., 2018):

- Virus
- Worm
- Trojan Horse
- Spyware
- Adware
- Root Kit
- Bots
- Ransomware

When one (or more) of these types of malware is (are) identified in a specific case's literature, we can establish malware as a breach method present in that data breach. We do not deem it necessary to further define and explain these different types of malware as it is expected that the specific type, if present, will always be mentioned in the case literature. Hence, we do not need to recognize the specific malware type per se and there is no need to elaborate on detailed descriptions of these malware types. In any case, the malware description as given in the first sentence of this subsection is deemed sufficient.

4.2.2.2. Physical

The theft, loss or manipulation of physical devices or physical environments in order to bypass network security or physically access the device where sensitive data resides (European Network and Information Security Agency, 2020; Shabtai, Elovici, & Rokach, 2012b; Widup et al., 2018). This includes examples like (European Network and Information Security Agency, 2020; Shabtai et al., 2012b; Widup et al., 2018):

- the use of payment card skimmers
- physical tampering with Point-of-Sale devices
- the theft or loss of a (company) laptop holding sensitive data
- unauthorized access to a company building and subsequently plugging a malicious USB killer device into company servers or computers

- wire-tapping, which may be preceded by unauthorized access to a building to install the wire tap

4.2.2.3. Social Engineering

Social engineering refers to the skill of manipulating legitimate users of information systems into compromising those information systems in favor of the social engineer. These social engineers target users with legitimate access to systems and data and cunningly get them to divulge sensitive data or persuade them to perform specific actions. Social engineering attacks can be executed by humans and software through various channels, mainly e-mail but also by telephone, instant messaging, social networks, websites, or cloud. The attack methods that can be used are:

- **Phishing:** the attempt to acquire sensitive data or to influence an individual to perform desired actions by impersonating a trustworthy entity in an electronic communication medium (Krombholz, Hobel, Huber, & Weippl, 2015). This kind of attack is usually intended for large groups of people and often sent out in a random manner (Diogenes & Ozkaya, 2019; Krombholz et al., 2015).
- **Spear-phishing:** a phishing attack which is specifically targeted to particular individuals in an organization (Diogenes & Ozkaya, 2019; Krombholz et al., 2015). It requires the investigation of potential victims first in order to carefully tailor the attack to the interests of intended victims (Diogenes & Ozkaya, 2019; Krombholz et al., 2015).
- **Watering hole:** a targeted attack where advantage is taken of the trust certain individuals have in websites they regularly visit by compromising those websites (Krombholz et al., 2015). Such a compromise eventually results in the targeted individual's computer being infected and the attacker gaining access to that computer and the network it is connected to (Diogenes & Ozkaya, 2019).
- **Baiting**: an attack in which a malware-infected external storage device is deliberately left where intended victims will easily find it (Diogenes & Ozkaya, 2019; Krombholz et al., 2015). Relying on human curiosity, the expectation is that the victims will pick it up and plugs it into their computers (Diogenes & Ozkaya, 2019; Krombholz et al., 2015).

When one (or more) of these is (are) identified in a specific case's literature, we can establish social engineering as a breach method present in that data breach. However, this is not an exhaustive enumeration of possible social engineering attack methods. We do not deem it necessary to complement this list as it is expected that the specific way of social engineering, if present, will always be mentioned specifically in the case literature.

4.2.2.4. Misuse

Misuse was used by Widup et al. (2018) to label security incidents and refers to *"any unapproved or malicious use of organizational resources"* done by authorized organization members or with their cooperation. This concerns any form of privilege abuse or (unintentional) data mishandling (Widup et al., 2018). Privilege abuse concerns *"using logical access to assets ... without having a legitimate ... need to do so"* (Widup et al., 2018).

We use this interpretation of Misuse in our study as value for the parameter Breach Method, but specifically in case misuse has led to or was present during a data breach. As we have mentioned earlier in chapter 1, a data breach is a specific type of security incident. And we are investigating data breaches.

An example of unintentional data mishandling could be accidentally sending an email with sensitive data to multiple email addresses which should not have received it. Another example of unintentional data mishandling could be accidentally publishing sensitive data somewhere it is publicly accessible.

4.2.2.5. Poor or No Security

When a data breach occurred because the organization's network and endpoints are not properly secured, or even not secured at all, this is attributed to having poor or no security. When this indeed is the case, we expect this to come forward in that data breach's case literature in the statements made by security experts analyzing the data breach or reporters covering the data breach. This will then support us in establishing poor or no security as a breach method.

Examples for this are databases or cloud storage which are left open and unprotected, API without access control, or the implementation of systems not according to basic industry standards.

4.2.2.6. Hacking

We have seen various definitions of hacking in the literature (Carter, 2019; Gregorio, Mathanamohan, Mahmoud, & AlTaei, 2019; Imran, Faisal, & Islam, 2019; ISO, 2012; M Nieles et al., 2017). Based on those definitions, a comprehensive definition of hacking can be derived. It defines hacking as the activity performed by one or more individuals with the intent of tampering with the software, hardware, and data in a network, or taking over the infrastructure, for a purpose which is different from the purpose of the owner or operator of those systems, and without their consent.

For our research we further specify the definition of hacking as the use of a system to gain unauthorized access to data in the same or another system, which could result in the compromise of that data by way of the breach types we discussed in section 4.1.1.

This definition is still comprehensive enough to also overlap with the other Breach Method values. That is why we apply the value Hacking only if the criteria in the definition are met *and* certain methods present in the data breach are not clearly or completely covered by the other Breach Method values.

An example of hacking is the use of Denial-of Service attacks which can render data unavailable. Another example is the use of SQL injection to gain access to data in a database. The stealing of credentials may be the result of one of the earlier mentioned breach methods, for example the use of phishing. However, subsequently using these stolen credentials to gain access to data is considered hacking. A data breach in which this happened, will have at least both Social Engineering and Hacking as values for the parameter Breach Method.

4.2.2.7. Inconclusive

We use the value Inconclusive if none of the above breach methods can be identified in the concerned data breach.

4.2.3. Breach Actor

This parameter refers to the entity who is the breacher and its intent. In the literature, we have seen that there are different types of breach actors (Ayereby, 2018; Hadlington, 2018; Hauer, 2015; Shabtai, Elovici, & Rokach, 2012d; Stiennon, 2013). Based on the breach actor descriptions in these sources, we came up with our own typology of breach actors. Each breach actor type represents a value of our parameter. However, if we cannot determine the breach actor type, the parameter value Inconclusive is applied.

Breach actors can be insiders or outsiders in relation to the breached organization. The breach actor types and their description for insider actors are as follows:

- **Inadvertent Insider:** a definition for unintentional insider threat is presented by Hadlington (2018) as *"a current or former employee, contractor, or business partner who has or had authorized access to an organization's network, system, or data and who, through action or inaction without malicious intent, causes harm or substantially increases the probability of future serious harm to the confidentiality, integrity, or availability of the organization's information or information systems"*. We use this definition as the description that refers to an inadvertent insider in our research, i.e., the legitimate member of the breached organization who actually breached that organization's systems, but unintentionally. This could for example be an employee who during his legitimate task stumbles upon sensitive data left publicly accessible.
- **Malicious Insider:** Hadlington (2018) presents a definition which states that a malicious insider is *"a current or former employee, contractor, or business partner who has or had authorized access to an organization's network, system, or data and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization's information or information systems"*. We use this definition as the description that refers to a malicious insider in our research, i.e., the legitimate member of the breached organization who actually did the breach with malicious intent. Such insiders mainly act out of personal or financial gain (Hadlington, 2018). This could for example be a disgruntled employee acting out of resentment.
- **Ethical Insider:** basically, an insider who fits the description of the malicious insider, but who acts out of ethical motives. This could for example be a whistleblower.

The definitions presented by Hadlington (2018), which we use for our respective parameter values Inadvertent Insider and Malicious Insider, can be modified by us and used to respectively describe the parameter values for Inadvertent Outsider and Malicious Outsider. We do this for those values and additionally provide a description for Ethical Outsider.

- **Inadvertent Outsider:** we modify the definition for unintentional insider threat, as presented by Hadlington (2018), to describe an inadvertent outsider in our research as an individual without previous or current *"authorized access to an organization's network, system, or data and who, through action but without malicious intent, causes harm or substantially increases the probability of future serious harm to the confidentiality, integrity, or availability of the organization's information or information systems"*. This could for example be an individual surfing the internet and stumbling upon sensitive data left publicly accessible.
- **Malicious Outsider:** we modify the definition for unintentional insider threat, as presented by Hadlington (2018), to describe an inadvertent outsider in our research as an individual without previous or current *"authorized access to an organization's network, system, or data who intentionally causes harm or substantially increases the probability of future serious harm to the confidentiality, integrity, or availability of the organization's information or information systems"*. Examples are black and blue hat hackers, hacktivists, or state sponsored hackers.
- **Ethical Outsider:** basically, an outsider who fits the description of the malicious insider, but who acts out of ethical motives. This could for example be a white hat hacker.

When we cannot determine the breach actor type in a certain data breach case, we assign the value Inconclusive:

- **Inconclusive:** breach actor type cannot be determined

4.3. Breached Organization

This section treats the parameters for describing the organization where the data breach took place, meaning the victim organization that got breached. The parameters to be used for this are Organization Sector, Organization Size, Organization Headquarters, Organization Component, and Data Responsibility.

4.3.1. Organization Sector

This parameter is used for indicating the main sector that the breached organization operates in. It provides us with insight in the distribution of data breaches across different types of organizations and can eventually be used to compare between sectors to see organizations from which sectors are hit the most by a data breach. Furthermore, we can establish which type of organization from our sample suffer from security misconfigurations the most. Based on the organization types defined in Ayyagari (2012), Sen and Borle (2015), Widup et al. (2018) and McCandless et al. (2019) we formulated the following values as representatives for the sectors in which the breached organization operates:

- Financial and Insurance
- Retail (including Online Retail)
- Education
- Government and Military
- Medical and Healthcare
- Technology
- Other
- Inconclusive

These self-explanatory values are applied mutually exclusively. We understand that certain organizations operate in multiple sectors, but in such cases we choose the value that represents the breached organization's core sector at the time the breach took place.

4.3.2. Organization Size

This parameter is for describing the breached organization's size in terms of size categories. Applying this parameter provides us with insight in the distribution of data breaches across different organization sizes. The ranges for this parameter's ordinal values are based on the staff headcount thresholds the European Commission utilizes for medium and small size enterprises (European Commission, 2016). The mutually exclusive parameter values are shown in table 3 below.

Organization Size	Range Criteria
Small	< 50 employees
Medium	< 250 employees
Large	≥ 250 employees

Table 4 Values for the parameter Organization Size based on categories.

4.3.3. Organization Headquarters

This parameter is used to describe the affected organization's base location in terms of countries. This will give us an overview of where breached organizations are located and provide us with insight in the distribution of data breaches over the globe. Note that this distribution is also affected by our case selection, which on its turn depends on the available case literature. Another point is that knowing where the breached organization is located, does not necessarily mean potential attackers – in the case of an attack - are also located there. Furthermore, it does not necessarily have to mean the data subjects are from that country. Data subjects are treated in section 4.3.1.

Organization Headquarters is a nominal parameter, and its values are mutually exclusive. Theoretically, the values we can choose from to apply to an organization in a data breach consist of all the countries in the world. We are not going to explicitly list all these values, as this results in an unnecessary long list of values of which most will not be used eventually. Alternatively, we will identify the breached organization in each case and determine in which country its headquarters are based. To label the identified country, we will use the country codes as published by the International Organization for Standardization (ISO, 2020). If the base country of the breached organization cannot be determined, we will use the label Inconclusive.

4.3.4. Organization Component

This parameter is for describing more specifically the part of the organization's infrastructure where the data was breached. Hence, Organization Component is used to identify the organization assets where the data was present at the moment of compromise. The application of this parameter to a certain data breach provides us with more insight in and understanding of what took place where during that breach. This contributes to our assessment process in order to finally establish the root cause of a data breach. We will also see the number of times each component was involved in a data breach.

In various literature, we have seen that affected assets in breached organizations are being mentioned (Adebayo & Omotosho, 2013; Cheng et al., 2017; Saleem & Naveed, 2020; Widup et al., 2018). Based on those descriptions, we formulated the following parameter values for Organization Component:

- Database
- Web Server
- Mail Server
- Directory
- Website
- Mailbox
- Device
- Cloud Storage
- Other
- Inconclusive

In one and the same data breach, various organization components can be affected. This means these values are not mutually exclusive. This list is not exhaustive, and therefore we introduced the value Other, in case a component is affected which is not listed as one of other values. If we cannot identify a compromised asset, the value Inconclusive is used.

4.3.5. Data Responsibility

The responsible organization is the organization that was holding or using the compromised data when the breach took place and carried the responsibility for its safety and security. The parameter Data Responsibility is basically about whether the responsible organization owned the data, or only used it. The General Data Protection Regulation (GDPR) defines two roles in terms of data responsibility, namely the data controller and data processor:

- Data controller *"means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data"* (GDPR, 2016).
- Data processor *"means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller"* (GDPR, 2016).

We take these roles and use them as values for our parameter Data Responsibility, to indicate which of these two roles the breached organization fulfilled when the data breach took place. This will show us in how many data breaches the responsibility for the data was with both the controller and processor. The nominal parameter values are as follows:

- Data Controller
- Data Processor
- Inconclusive

Shown as the last in the list is the value we use if we are not able to determine the type data responsibility the breached organization was holding.

4.4. Data Specification

The parameters in this section are meant for describing the entity to which the breached data belongs, the amount of data implicated in the breach, the type of data involved, the state of the implicated data at the moment of breach, and the extent to which the responsible organization needed to possess or use the involved data. Respectively, these parameters are named Data Subject, Data Amount, Data Type, Data Sensitivity, Data State, and Data Necessity.

4.4.1. Data Subject

When a data breach occurs, there is always an affected party. Not only the breached organization under which responsibility the data becomes compromised, but also one or more entities to which the data originally belongs and for which the data has some kind of value. More importantly, these are the entities which are inconvenienced by the data breach and its aftermath.

Commonly, this concerns individuals, i.e., natural persons, whose personal information is compromised. In the General Data Protection Regulation (GDPR), which focuses on the protection of personal data, these natural persons are referred to as data subjects (GDPR, 2016):

"An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier."

We have seen references to data subjects, and what falls under them, in various literature (Article 29 Working Party, 2017; Cheng et al., 2017; European Data Protection Board, 2021; GDPR, 2016). Based on these references, we came to the following values:

- Citizens
- Customers
- Employees
- Patients
- Students
- Members
- Other

However, a data subject does not necessarily have to be a natural person. For instance, in Stiennon (2013) we saw that next to the total number of records (which are commonly related to natural persons), also the monetary value of breached proprietary information can be used as value for the concerned parameter in their Breach Level Index. And this proprietary information belongs to a certain party. Hence, to the list above we added:

- Breachee (the breached organization itself, for example when losing intellectual property)
- Trusted business partner (for example when losing intellectual property)

We also need a value for cases where we cannot establish a specific data subject type:

- Inconclusive

Except for the label Inconclusive, the values basically describe the relation of the data subjects to the breached organization which is responsible for their data. In a single data breach, multiple types of data subjects can be involved, meaning the values are not mutually exclusive.

Applying this parameter will provide us with an overview of the distribution of different data subjects across the data breaches.

4.4.2. Data Amount

Every data breach involves a certain amount of data which is compromised. Generally, the amount of compromised data is considered a relevant factor in determining the severity and impact of a data breach (European Network and Information Security Agency, 2020; Gemalto, 2019; Ponemon Institute, 2018, 2020; Posey Garrison & Ncube, 2011; Stiennon, 2013).

This amount is usually expressed in units of data records. Generally, a data record is defined as the collection of information that identifies the data subject whose personal information was compromised (Ponemon Institute, 2020). The content of such a record may vary in content and size, depending on which personal data is compromised.

In previous section, we have discussed in that in our research a data subject does not necessarily have to be a natural person. This means that a data record does not necessarily have to pertain to an individual. We will consider the loss of proprietary information belonging to an organization also in terms of data records. Because even if this may concern one or few data records, proprietary information can be of high value and therefore the severity and impact of this few compromised records can still be very high.

For this parameter Data Amount we created categorical values, which are ordinal. To come to certain ranges, we looked at the Breach Level Index (BLI) Methodology as proposed by Stiennon (2013). There, a formula is used to come to a BLI score. This formula is the following:

Log10 ($N * t * s * A$)

Where the variables can be one of certain predefined numerical values. The explanation of the variables is as follows:

- *N* stands for the total number of breached records
- *t* = the type of data in those records (possible values are 1 to 5)
- *s* = source of the breach (possible values are 1 to 5)
- *A* = the kind of action and intent, in other words whether the breached data has been used for harmful purposes, going from no action to the most harmful (possible values are 1 to 10)

The use of this formula gives a certain BLI score. The BLI also provides a scale which supports the interpretation and understanding of the BLI score. This scale has five categories of different BLI score ranges. And each category (i.e., range) has a characterization in terms of impact.

In this instance, we are only interested in the influence of the number of data records on that impact. Hence, apart from *N*, we decided to keep all other variables (*t*, *s*, and *A*) equal and use for them the value 1.

If we look at the interpretation scale, we see that from the third category upwards, the characterization includes significant impact. This third range starts a BLI score of 5. This means that if *t*, *s*, and *A* are 1, *N* should be at least 100.000. This is how we came to the first range for our own parameter Data Amount.

Because of the log10 function, the BLI score will increase by 1 if the number of records is ten times bigger. That is why our next ranges use 1 million and 10 million.

We end up with four values, which are mutually exclusive. If we cannot determine the number of records or the category, we use the label Inconclusive. The values can be found in table 4. As can be seen in table 4, each value represents a range.

After the assessment, we will create a bar chart for this parameter. This chart will indicate how many data breach cases each value has, i.e., how many data breaches fall in each category that the corresponding value represents. We introduce the letter R which we will use in our value labels. R simply represents the number of records (per million). The value eventually are used in the horizontal axis of the bar chart. We use the letter R to keep the horizontal axis of the bar chart readable.

The result of applying this parameter will provide us insight in the distribution of data breaches in relation to the number of compromised records, which eventually provides a general insight into whether most of the data breaches are less or more severe. That is why the use of ranges is more practical. Furthermore, in the literature regarding some cases there may not be available a specific number of breached records, nut only a very rough estimation. Potentially because of such cases, the use of ranges is also more practical.

R represents the number of records per million	**Range explanation**
R < 0.1	Less than 100 thousand records
0.1 ≤ R < 1	The number of records is equal or more than 100 thousand and less than 1 million records
1 ≤ R < 10	The number of records is equal or more than 1 million and less than 10 million records
R ≥ 10	Equal or more than 10 million records
Inconclusive	The number of records cannot be determined and therefore a range value cannot be selected

Table 5 Values for the parameter Data Amount

4.4.3. Data Type

The type of data involved in a data breach is relevant since different types have different impact. This should dictate how these data types are to be treated and how security should be configured to protect data from each data type.

We saw in McCallister, Grance, and Scarfone (2010), Death (2017) and Kostadinov (2019) that multiple types of information can be distinguished. In NIST this is considered from a security perspective which is necessary to achieve the security objectives of confidentiality, integrity and availability. At the highest level, information that needs to be protected can be personal or proprietary, since that information belongs to either an individual or an organization. From these two main types we work towards the eight subtypes that will constitute the values for the parameter Data Subject. These are the types we deemed relevant as data that could potentially be compromised in a data breach.

Personally identifiable information (PII) is defined as *"any information about an individual . . . , including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information"* (McCallister et al., 2010).

Although PII covers medical information that is linked or linkable to an individual, in the literature this kind of information is also separately defined as its own type and is named protected health information (Death, 2017; Kostadinov, 2019; McCallister et al., 2010). Based on these sources, we consider as protected health information (PHI) any health-related information which is linkable to a specific individual, including health status, health care and payment for health care.

Also corresponding with PII, but distinguished in the literature as its own type, is personally identifiable financial information (Bojanova, Yesha, Black, & Wu, 2019; Cooper, 2015; Gramm, Leach, & Bliley, 1999). Personally identifiable financial information (PIFI) concerns information provided by an individual to a financial organization, information resulting from a transaction involving a customer, or otherwise obtained by a financial organization (Gramm et al., 1999). As such, it concerns PII which can be linked to a specific individual's finances (TechTarget, 2008). According to this description there is overlap with PII. In our assessment, we will use PIFI to label compromised data which clearly can be considered financial information, like for instance payment card numbers and bank account balances. But for instance, not the names of bank clients, these will be labelled by PII.

Another type of personal information we established as a separate type is login information (LI). This concerns the information which is used by users to log into a system or application. Commonly, LI concerns usernames and passwords.

Strictly speaking biometric data (BD) also falls under PII, as we can see in the definition of PII above. However, for the purpose of our assessment we also single out BD, in order to use it as one of the values. This is because of BD's specific nature and sensitivity. This can be BD which is used for authentication purposes, and in that specific instance that information is labelled both as LI as well as BD. But there could also be different scenario's in which BD is gathered and stored by organizations for the purpose of specific products and services these organizations provide and which involve the collecting or processing of BD.

Proprietary information is information which is considered sensitive by an organization and relates to or is associated with that organization's input, processes, output, and other organizational matters (Death, 2017; Kissel, 2013; Kostadinov, 2019).

When this information is considered valuable, it will fall under the type Trade Secrets & other Intellectual Property (TSIP). When this information is not valuable, it will fall under Other (Business) Information (OBI). Whether proprietary information is valuable or not, is assessed on whether this information, if compromised,

would damage the organization and/or benefit other organizations, like competitors. This assessment is only based on the literature we use for that specific case.

When information is involved that is considered sensitive by a government organization, for instance strategic military information, this is labelled as classified information (CI). Additionally, when we cannot determine the type of the compromised data, we label it as inconclusive.

The values, their descriptions, their examples, and their labels are summarized and grouped in table 5 below.

Main Data Types	Parameter Value	Label	Description	Examples
Personal information	Personally Identifiable Information	PII	"*Any data that can be used to identify a specific individual"* (Park & Chai, 2018).	Name, Date of birth, Passport number, Social Security Number (McCallister et al., 2010).
	Protected Health Information	PHI	PII that can be linked to a specific individual's health.	Medical records, health care payments
	Personally Identifiable Financial Information	PIFI	*"Any type of PII that is linked to that person's finances"* (TechTarget, 2008).	Credit card number, account balance
	Login Information	LI	Data used by users in the process to gain access to a certain system and which is used by that system to authenticate that user, i.e. verify that user's identity.	Usernames, passwords
	Biometric Data	BD	*"The measurement and statistical analysis of [an individual's] unique physical and behavioural characteristics"* (Gillis, 2020). It is mainly used to establish that individual's identity based on its physical, chemical or behavioural attributes.	Fingerprints, palm prints, iris scans, retina scans, face recognition
Proprietary Information	Trade Secrets & Other Intellectual Property	TSIP	Refer to any confidential (business) information which has value for an enterprise or provides an enterprise with a competitive edge	Trade secrets, (patented) process, (patented) design, (patented) invention, copyright materials, pricing data, source code
	Other (Business) Information	OBI	Enterprise related data not falling in the TSIP category	Personnel matters, outdated and obsolete product or service specific information
	Classified Information	CI	Data which is considered sensitive by a government authority	Military information
Inconclusive	Inconclusive	INC	No data type could be determined	

Table 6 Values for parameter Data Type

4.4.4. Data State

From literature on IT security and data protection, we learned that data can have three states, namely data-at-rest, data-in-motion and data-in-use, which are relevant for the treatment and protection of these data (Andress & Leary, 2017; Hauer, 2015; Shabtai et al., 2012d; Shackleford, 2007). These form the values for our parameter Data State.

Data-at-rest (DAR) concerns data which are in storage, excluding data that are transferred across a network and data residing in volatile memory (Andress & Leary, 2017; Hoog & Strzempka, 2011; Shabtai et al., 2012d). These data may be residing in databases, on file servers, laptops, cloud-based systems, USB drives or any other non-volatile storage medium (Hoog & Strzempka, 2011; Shackleford, 2007). On these mediums, the data may be found in different types of files, for instance log files, configuration files or application files (Shackleford, 2007).

Data-in-motion (DIM) refers to data that are transferred between networks and/or within networks, and these may concern different types of networks (Andress & Leary, 2017; Hoog & Strzempka, 2011; Shabtai et al., 2012d; Shackleford, 2007). For instance, data which are being moved from local storage to cloud storage or an email which is being transmitted (Hauer, 2015; Lord, 2019; Shabtai et al., 2012d).

Data-in-use (DIU) refers to data which are stored in a nonpersistent state in volatile storage (Sachowski, 2016). For instance, this may be data located in Random Access Memory (Hoog & Strzempka, 2011). Furthermore, DIU also describes data which are being processed or data with which users are interacting (Andress & Leary, 2017; Shabtai et al., 2012d; Shackleford, 2007). Examples are screen-capture and copy-paste operations involving sensitive data, printing and faxing involving sensitive data or by inputting sensitive data on a web page or in another application (Hauer, 2015; Shabtai et al., 2012d). Data are considered to be in use only during the period in which they are processed, accessed, read, edited or otherwise utilized by or through systems (Andress & Leary, 2017).

Data which got compromised because of a data breach is labelled by the data state in which that specific data existed at the moment those data got compromised (Shabtai et al., 2012d). This is done by using the labels DAR, DIM, and DIU. If the data state cannot be determined, the label Inconclusive is used. During a single data breach, various data from different locations may be compromised in different ways. Hence, there may be involved data of more than one data state and a single data breach can have multiple Data State labels.

4.4.5. Data Necessity

The GDPR (2016) incorporates the principle of necessity with regard to the treatment and protection of personal data, which basically comes down to whether the personal data which are collected, processed and stored are necessary for the objectives and operations of the party handling those data. This inspired us to include Data Necessity as a parameter in our Assessment Framework, utilizing it in our own manner.

Data Necessity enables us to assess to whether the compromised data was necessary for either the data controller or data processor. In our assessment, we want to determine whether the data controller or data processor needed the compromised data for its operations and offerings. The keeping and processing of unnecessary data may already be the result of a security misconfiguration. And if that is not the case, the presence of unnecessary data makes the consequences of a data breach, which potentially is caused by a security misconfiguration, even worse anyway.

The keeping and processing of unnecessary data may already be the result of a security misconfiguration. And if that is not the case, the presence of unnecessary data makes the consequences of a data breach, which potentially is caused by a security misconfiguration, even worse anyway.

The values for this parameter are described in table 6 below. When part of a compromised data trove is found to be necessary data, and the other part's necessity cannot be determined, this is still labelled as Necessary. Similarly, when part of a compromised data trove is found to be unnecessary data, and the other part's necessity cannot be determined, this is labelled as Unnecessary.

Data Necessity values	**Label**
The compromised data consisted of necessary data	Necessary
The compromised data consisted of unnecessary data	Unnecessary
Part of the compromised data was necessary; another part of the compromised data was not necessary	Mixed
It could not be determined whether the compromised data was necessary or unnecessary	Inconclusive

Table 7 Values for parameter Data Necessity

4.5. Data Breach Detection

This section treats the parameters that are used to describe the detection of data breaches in terms of the detector, the way of detecting and the intent behind it, as well as the time that passed from the moment of breach until the detection. The first three constitute the parameter Data Breach Detection, the latter concerns the parameter Time Between Breach and Detection.

Applying these parameters to the selected data breach cases potentially provides us with insight in points of improvement for organizations with regard to having their own effective and timely detection capabilities.

4.5.1. Data Breach Detection

Data Breach Detection is used to describe how the breach was noticed or detected, in terms of the source of the detection, the method of detection, and the intent. These constitute the three parameters we use to assess Data Breach Detection. One of the criteria used to characterize data leakage incidents, namely Incident Detection as proposed by Hauer (2015), inspired us to formulate these three parameters and their values.

The first parameter Detection Source is used to indicate whether the data breach was detected internally or externally in relation to the breached organization. Its mutually exclusive values, their descriptions, and labels are shown below in table 7.

Detection Source Values	**Label**	**Description**
Internal	Internal	• Data breach is detected within the breached organization's perimeter • Detected by resources belonging to the breached organization
External	External	• Data breach is detected outside the breached organization's perimeter • Detected by resources not belonging to the breached organization
Inconclusive	Inconclusive_DS	Not possible to determine whether the data breach was detected internally or externally

Table 8 Values for parameter Detection Source

The second parameter Detection Method is used to indicate whether the data breach was detected in an automated way. More specifically, we consider the monitoring, searching and scanning of networks, endpoints and marketplaces for data or proof of incidents. And we assess whether these activities happened automated or manually. Unintentional encounter with data is labelled as having been detected manually. Detection Method's mutually exclusive values, their descriptions, and labels are shown below in table 8.

Detection Method Values	Label	Description
Automated	Automated	• Automated monitoring and analyzing of internal networks and endpoints for the purpose of detecting incidents • Automated searching and scanning of Internet and openly accessible networks and endpoints for incidents • Automated searching and scanning for data offerings on digital marketplaces, as well as scanning for information pointing in that direction
Manual	Non-Auto	• Manual monitoring and analyzing of internal networks and endpoints for the purpose of detecting incidents • Manual searching and scanning of Internet and openly accessible networks and endpoints for incidents • Manual searching and scanning for data offerings on digital marketplaces, as well as searching and scanning for information pointing in that direction • Unintentional encounter with data
Inconclusive	Inconclusive_DM	Not possible to determine whether the data breach was detected in an automated or manual way

Table 9 Values for parameter Detection Method

The third parameter Detection Intent is used to indicate whether the data breach was detected intentionally or inadvertently by the party which detected the breach. Its mutually exclusive values, their descriptions, and labels are shown below in table 9.

Detection Intent Values	Label	Description
Intended	Intended	Data breach is detected intentionally, the detector was looking for data breaches and the detection process was intentionally started to discover data or incidents in general
Inadvertent	Inadvertent	Data breach was discovered accidentally
Inconclusive	Inconclusive_DI	Not possible to determine whether the data breach was detected intendedly or inadvertently

Table 10 Values for parameter Detection Intent

Each assessed data breach case is labelled with one value of each parameter. The assessment is based on the information we find in the selected data breach case literature.

4.5.2. Time between Breach and Detection

The parameter Time Between Breach and Detection (TBBD) is used to describe how much time has passed between breach and detection. As a data breach may take a certain period itself, and even consist of multiple breach periods, this is assessed from the first known moment of the data breach. From that moment we measure the time until the first known moment of detection, independent of how the detection took place and by which party. This assessment is based on the information we find in the corresponding case literature.

This parameter TBBD has an ordinal scale which utilizes labels to classify assessed cases into ordered classes. These classes constitute the values for TBBD. We will explain the formation of these classes next.

In a study into the cost of a data breach, Ponemon Institute reported *"on the relationship between how quickly an organization can identify and contain data breach incidents and the financial consequences"* (Ponemon Institute, 2018). The institute found that *"companies that identified a breach in less than 100 days saved more than $1 million as compared to those that took more than 100 days"*. (Ponemon Institute, 2018). It also found that *"companies that contained a breach in less than 30 days saved over $1 million as compared to those that took more than 30 days to resolve"* (Ponemon Institute, 2018).

We used those findings to formulate the categorical, ordinal values of our parameter, as is displayed in table 10. The use of a categorical parameter eventually allows us to easily group data breaches on the different TBBD categories.

TBBD values	Description
Short	≤ 30 days
Medium	30 < days ≤ 100 days
Long	≥ 100 days
Inconclusive	The time which has passed between the first moment of data breach and detection

Table 11 Values for parameter Time between Breach and Detection

4.6. Root Cause Assessment

This section treats the parameter which we use to assess the root cause of the data breach. For clarity reasons, we describe the type of parameter we use in a separate section.

4.6.1. Is the Root Cause a Complex Attack or a Security Misconfiguration?

As earlier mentioned in section 3 of chapter 1, the root cause refers to the most basic reason for an event to occur. The International Organization for Standardization (ISO) standard 18238:2015 defines a root cause as the *"original event, action, and/or condition resulting in an actual or potential undesirable condition, situation, nonconformity or failure"* (ISO, 2015b). This standard also states that often there are multiple root causes to one problem (ISO, 2015b).

The IEEE Standard 1856-2017 (2017) presents the following definition of a root cause: *"The root cause is the most basic causal factor or factors that, if corrected or removed, will prevent the recurrence of an event."* Important is the condition that if the root cause is taken away, this prevents the event ensuing from it. We want to identify the basic causal factor in a data breach that meets this condition. Furthermore, when investigating the data breach for this purpose, we do this from a socio-technical perspective, as earlier mentioned in section 4 of chapter 1.

In this section, we discuss the parameter which enables us to assess whether a data breach has a complex attack or security misconfiguration as a root cause. These are two values for this parameter, together with the value 'Inconclusive' which is used when the root cause cannot be determined. To be able to categorize the root cause as either being a complex attack or security misconfiguration, we need to know what they mean. These definitions provide us with a classification and characteristics of both terms being defined. This supports us in our effort to recognize and identify them in case literature. We will formulate the definitions of both terms next.

A complex attack is a sequence of malicious actions directed at a network and its hosts, generally performed in multiple stages at different rates by multiple agents using multiple methods (Camtepe & Yener, 2007; Gregorio-de Souza et al., 2006; Mishra, 2019). Additionally, they usually are not easily detected and require the inspection of audit information and logs (Gregorio-de Souza et al., 2006).

Security misconfiguration stands for the failure of operators to configure the systems in an operational environment, which falls under their control and responsibility, in such a way that the confidentiality, integrity, and availability of those systems are not harmed (Dietrich et al., 2018; Tobias Fiebig, 2017). This failure usually consists of one or more human made errors, for instance:

- uninstalled updates because they are delayed, ignored or missed
- authentication and authorization not or improperly configured
- lacking or incorrect firewall rules
- etc.

For each data breach, this assessment is based on the case literature that we study. In that literature we search for direct or indirect evidence of the presence of security misconfigurations or of a complex, sophisticated attack that eventually resulted in that data breach. The evidence is mainly based on:

- statements made by the breached organization
- statements made by government authorities or other parties who received access to investigate the data breach within the breached organization or with the obliged cooperation of the breached organization

- statements made by experts and researchers in the field
- our own interpretation of case literature

A favorable situation would be the breached organization acknowledging there was indeed a security misconfiguration and explaining that situation, including how that misconfiguration could lead to the data breach. Since we expect that breached organizations are more inclined to deny their breaches were caused by security misconfigurations, we can assume that when they admit a security misconfiguration as the root cause, this is already trustworthy enough, and in principle we do not need to search further for evidence of a complex attack in such specific cases. But in all cases, we do want to look at the evidence for a complex attack also.

In case we do not find statements of the breached organization acknowledging the presence of security misconfiguration as cause for the data breach, we look at what experts and other knowledgeable persons declare in relation to that data breach. Of course, we look for multiple experts declaring the same, to give our assessment more ground. In some cases, we could have the possibility to consult investigative reports written by certain (government) parties.

If we are not able to find such statements, we can look at detailed descriptions of the data breach and form our own judgment on whether security misconfiguration was present as cause for the data breach. But this also happens as an addition to an admission by the breached organization or statements by experts.

Thus, security misconfiguration is clearly defined, and we can be assisted if the breached organization admits the presence of it as the root cause of the data breach. Additionally, we want to prevent jumping too soon to the assessment that the root cause was a complex attack. Potentially, this might be likely since a detailed description of a data breach may easily make the data breach appear complex and sophisticated which then again could more quickly lead us to unjustly assess this was the result of a complex attack.

On the other hand, complex attacks often are not easily detected. Therefore, in certain cases, it may be hard to find sufficiently informative literature describing and substantiating the existence of a complex attack as root cause for the data breach.

Furthermore, a data breach may involve sophisticated, complex actions of attackers, while also security misconfiguration being present. In such cases, we still consider the data breach to have security misconfiguration as root cause. Because with regard to such cases we assume that even if attackers use complex and sophisticated methods, the data breach could not have happened if the security misconfiguration was not present, unless the case literature clearly and explicitly states otherwise. We have seen the same condition in Tobias Fiebig (2017), where the definition of security misconfiguration includes that *“the property could not have been tainted if the service would have been deployed and configured correctly”*. As we presented earlier in this section, the definition of a root cause in the ISO standard 18238:2015 (2017) includes basically the same condition, but more in general. This condition is an important part of our root cause assessment process.

We want to clarify that absence of evidence does not always imply evidence of absence. This means that if we are not able to find evidence for security misconfiguration, this does not necessarily mean that there was no security misconfiguration which eventually caused the data breach to happen. However, if there is sufficient evidence for a complex attack, it is valued as such. Otherwise, it is labelled as Inconclusive. Because we have to determine the root cause based on evidence in the case literature.

Conversely, the same applies to complex attacks. If we are not able to find evidence for a complex attack, this does not necessarily mean that there were no elements of a complex attack which eventually caused the data breach to happen. However, if there is sufficient evidence for a security misconfiguration, it is valued as such. Otherwise, it is labelled as Inconclusive. Here also we have to base the determining of the root cause based on evidence in the case literature.

What we described in this section, forms our root cause assessment process. For clarification, we have formulated a flowchart which visually and orderly shows the essential steps we take during the root cause assessment process, i.e., to show the steps we take to ultimately determine whether the root cause is a security misconfiguration, a complex attack or undeterminable. The flowchart is shown in figure 1.

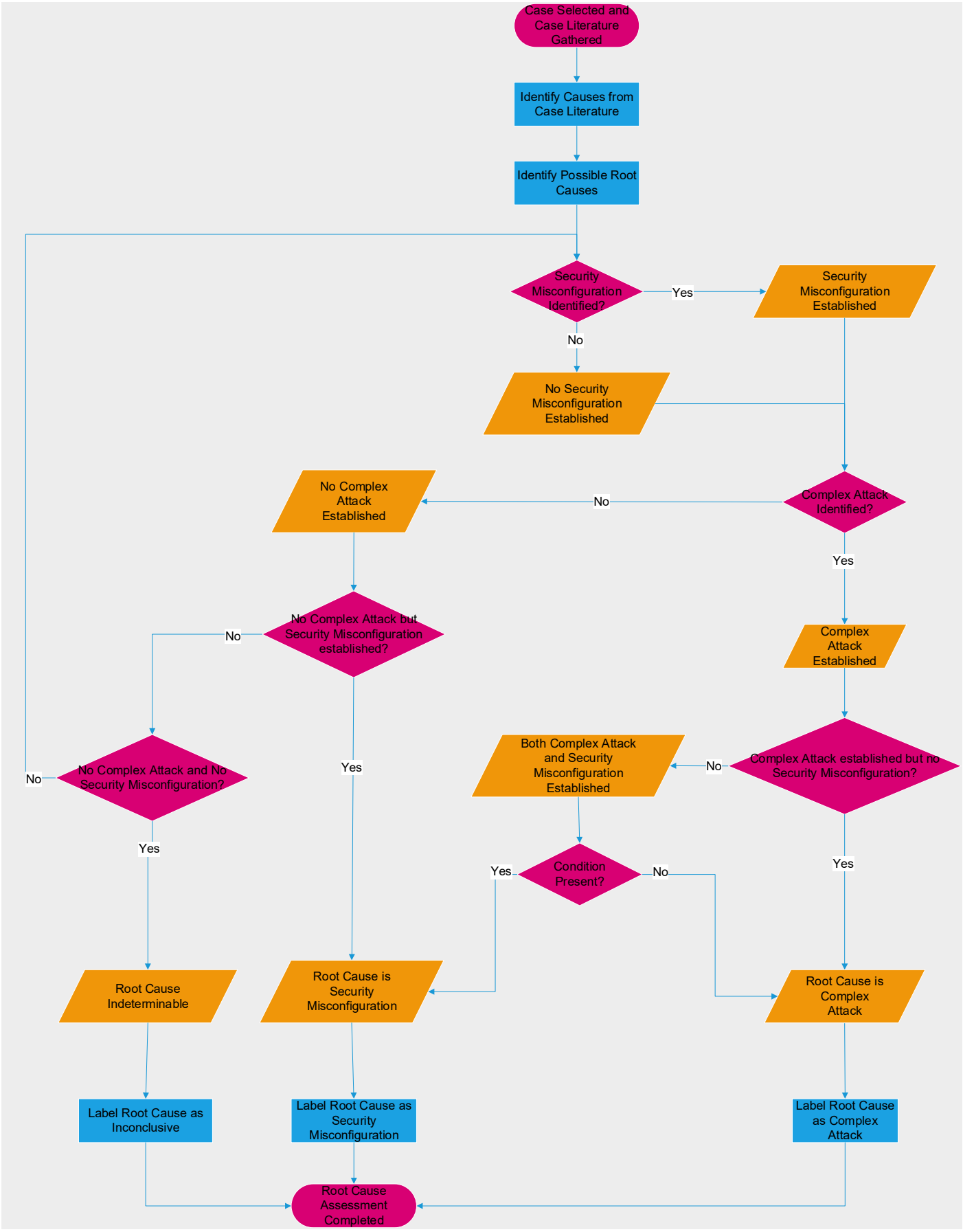


Figure 1 Root Cause Assessment Flowchart

4.6.2. Parameter Type

Our parameter is nominal because it has unordered categorical values. Since the outcome of our root cause assessment can either be Complex Attack, Security Misconfiguration or Inconclusive, the parameter values are mutually exclusive.

5 Case Studies

5.1. Introduction

This chapter contains all the data breach cases that we analysed and assessed for our research. Working through them case by case, we studied the selected sources which covered, described and/or analysed all those cases. Subsequently, for each case we filtered out the information relevant to our research and formulated an applicable case description. This was supported by applying our assessment framework and its parameters to the case, which eventually resulted in extracting the key characteristics for each case. The assessment framework functioned as a structure for the case descriptions that we have formulated based on the corresponding case literature. Because of the framework parameters it remained clear what to include in the case description and what to leave out, in such a way that a case description was formulated that it only contains the relevant information with regard to this research. Hence, by doing so, this chapter consists of the actual reporting of the results of applying the assessment framework per case, in a textual manner.

This all took place according to the approach as discussed in the methodology. For a better overview, each case description is summarized in a table which only contains the values for our defined parameters.

5.2. Case Descriptions

This section contains the descriptions and assessments of all individual cases included in this research. The assessments were done systematically and follow the structure of the framework.

5.2.1. Sony Pictures Entertainment (2014)

After receiving a threat a few days prior, delivered to some higher level Sony employees by email, Sony Pictures Entertainment (SPE) got the bad news of being breached when noticing employee computers were taken over. Also SPE's social media channels were overtaken, making obvious the attack. As significant part of SPE's computer network was already wiped or destroyed. This was the beginning of what ended up being an extensive data breach, including intrusion, data destruction and disclosure of sensitive company data.

The attack vector could not be determined exactly, but similar recent attacks used phishing methods, while other analysts state there might have been help from insiders. Attribution could not be determined with certainty: some analysts were convinced this was done by a state-sponsored entity (the state being DPRK), others believed it is more likely the work of another motivated entity like a hacktivist group.

Once access was gained to the SPE network, the intruders took their time to map the network and identify critical systems and files, in order to plan further theft and destruction, for which malicious software needed to be installed all over the network. This malicious software used Microsoft's Windows own network and management to propagate and to eventually attain technical disruption and destruction. The data exfiltration likely has been done during the period between intrusion and the moment the actual technical destruction started.

Eventually, next to damaged hardware and software, a vast amount of internal and confidential corporate data, including sensitive employee data, login information, intellectual property (like movies) and other sensitive business information was taken, which according to the perpetrators consisted of 100 TB. Subsequently 40 GB thereof was gradually leaked.

Considering the nature of the attack, which requires knowledge of SPE's network, there must have unauthorized access for a longer period of time enabling the attackers to map the network, or there must have been some insider knowledge (or both). A long period of unauthorized access in which no unusual activity was detected at all, in combination with the available information about how to be able to detect the malicious software before and after it has been triggered, as well as disclosed weaknesses of SPE security practices, brings us to the conclusion that while being present some sophistication on both strategical and technical level of the attack, SPE could have done enough to prevent or mitigate this data breach. Before the data breach the Cybersecurity and Infrastructure Security Agency (CISA) had already released tips to handle destructive malware. After the data breach the CISA released an alert concerning the specific type of destructive malware used, which included preventive measures to be used by operators to protect their networks. These measures not new and should have been implemented already. Therefore, we consider this to be a case of security misconfiguration.

Parameters	Values for Sony Pictures Entertainment
Data Breach Type	Unauthorized Access Unauthorized Disclosure Destruction Loss Alteration
Breach Method	Hacking Malicious Software Social
Breach Actor	Malicious Outsider

Organization Sector	Other
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Database Mailbox Cloud Storage
Data Responsibility	Data Controller

Data Subject	Employees Breachee Trusted Business Partners
Data Amount (Records)	10.000.000
Data Type	PII PHI PIFI LI TSIP OBI
Data Sensitivity	Internal Confidential
Data State	DAR
Data Necessity	Mixed

Data Breach Detection	External Non-Auto Intended
Breach Period (Days)	60
TBBD	Medium
TBDA	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 12 Summarized overview of identified values for Sony Pictures Entertainment

5.2.2. Saks Fifth Avenue and Lord&Taylor (2018)

This data breach came to light when the attacker, an organization known as JokerStash and Fin7, announced the sales offer of payment card data of more than five million credit and debit cards which was discovered by cybersecurity company Gemini Advisory (2018). These data were released in batches, such that financial institutions would not easily detect that stolen payment card data was stolen and illegally used (Schwartz, 2018d). While not sure if eventually everything has been released, at least 125,000 were actually offered for sale on the dark web (Gemini Advisory, 2018; Schwartz, 2018d).

Gemini Advisory (2018) analysed this released sample of the compromised data in cooperation with various financial institutions, and was able to confirm the analysed data came from cards used to shop by customers at Saks and Lord & Taylor in North America (Schwartz, 2018d), their common point of purchase. Saks and Lord & Taylor operate in the retail sector and are subsidiary companies of the Hudson Bay Company, which eventually acknowledged that a breach had occurred (Schwartz, 2018d).

Gemini Advisory (2018) estimated the window of compromise to run from May 2017 to March 2018. During that period, Saks' and Lord & Taylor's networks have been compromised. The data was very probably pilfered and exfiltrated by using malicious software. In breaches concerning payment card data, typically malicious software that scrapes payment cards is installed onto point-of-sale terminals, although in some cases malicious software is installed on the servers to which payment card data gets transferred during the payment process (Schwartz, 2018d). It is uncertain how the malicious software got in place, but it is assumed the attackers employed phishing e-mails sent to Hudson Bay Company employees (Associated Press, 2018; Gemini Advisory, 2018; Goel & Abrams, 2018; Schwartz, 2018d).

Even when subject to phishing, the malicious software should have been prevented from reaching the point of sale POS devices by implementing network segmentation in the first place (Gamer, 2015; Schwartz, 2015), which apparently has not been done properly. Because of the manner that almost any malicious software reaches the POS devices, a device linked to POS system is used, and these links should be decreased in numbered as well as be secured properly (Gamer, 2015; Schwartz, 2015). Furthermore, it appears that a very significant amount of retailers does not change the default passwords on their POS devices, making it easier for these devices to be infected with malicious software (Schwartz, 2015). Also, in the case that malicious software still gets installed in the POS system, there have to be tools in place to prevent the pilfering of the payment card data (Gamer, 2015; Schwartz, 2015).

All Saks and Lord & Taylor shops had already installed the EMV standard (Schwartz, 2018d), which should imply that payment card is safe and the payment card data is secure (Rosenblum, 2018). However, applying EMV does not stop the malicious software if customers are allowed to swipe their payment cards, instead of card dipping which uses the card chip (Schwartz, 2018d). Even when imposing customers to card dipping, the payment card data might not be safe against theft if the EMV standard is not implemented correctly (Schwartz, 2018d), i.e. if the operators do not set up their point of sale (POS) systems correctly and completely with all required components. The EMV standard misses an important component, namely the encryption of data transferred from the POS to the payment card processor which would leave the payment card data readable and available to be stolen (Constantin, 2014; Rosenblum, 2018). Further encryption methods that complement EMV are available, for instance Point-to-Point encryption and tokenization, but it appears that these were not implemented by Saks and Lord & Taylor (Rosenblum, 2018).

Our assessment is that the methods and practices mentioned above were not implemented, or at least not sufficiently and correctly, and therefore the networks of the affected organizations were not configured correctly such that a data breach could be prevented and mitigated.

Parameters	Values for Hudson Bay Company (daughter companies Saks Fifth Avenue and Lord&Taylor) Data Breach
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Social Malicious Software
Breach Actor	Malicious Outsider
Organization Sector	Retail
Organization Size	Large
Organization Headquarters	CA-CAN-124
Organization Component	Device
Data Responsibility	Data Controller
Data Subject	Customers
Data Amount (Records)	5.000.000
Data Type	PIFI
Data Sensitivity	Internal
Data State	DIU
Data Necessity	Necessary
Data Breach Detection	External Non-Auto Intended
Breach Period (Days)	335
Time between Breach and Detection	Long
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 13 Summarized overview of identified values for Saks Fifth Avenue and Lord&Taylor

5.2.3. FriendFinder Networks (2016)

In November 2016 it came to light that the online dating and adult entertainment company FriendFinder Networks (FFN) suffered a data breach about a month earlier, the result of which left over 412 million of its users' accounts compromised (McDaniel, 2019; Whittaker, 2016). These accounts were from several of FFN's services, with the greatest amount (more than 300 million) consisting of accounts from AdultFriendFinder.com, but also accounts that should not have been there at all such as 15 million accounts that supposed to have been deleted and 7 million Penthouse accounts while Penthouse was already sold (Dickey, 2016; Whittaker, 2016).

The breach was discovered by breach notification website LeakedSource, which also acquired the data after these were made available on the dark web (Daitch, 2016; Kirk, 2016; Ragan, 2016b). The breached data mainly concerned customer data amassed during twenty years and was contained in six databases, five corresponding to FFN's services and one from a unknown domain, and consisted of usernames, email addresses, passwords, membership status, browser information, date and IP address of last visit, and user purchase information (Daitch, 2016; Storm, 2016; Whittaker, 2016). Next to customer data, some other business information was also included in the data breach (Ragan, 2016b).

A portion of the passwords were simply plaintext, while the other portion was hashed (Kirk, 2016; McDaniel, 2019; Ragan, 2016b, 2016c; Whittaker, 2016). This hashing was done by using SHA-1 (Kirk, 2016; Whittaker, 2016), and a portion of those combined this with pepper (Ragan, 2016b, 2016c). However, the SHA-1 hash function was not considered secure enough at that time already (Kirk, 2016; McDaniel, 2019; Ragan, 2016c; Whittaker, 2016).

The hack presumably happened by way of Local File Inclusion (LFI) which enabled access to FFN's complete network including those of its services (Dickey, 2016). LFI allows input to be supplied to a web application by a potential attacker; in this way a web server can be redirected to execute a locally stored file which was included by the attacker (Kirk, 2016; McDaniel, 2019).

Remarkably, around the time of the data breach a security researcher had alerted Adult FriendFinder (one of FFN's services) that their website was vulnerable to LFI (Kirk, 2016). FFN confirmed to have received notifications of potential security issues and to investigated these (Kirk, 2016). While some turned out the be part of extortion attempts, FFN admitted to having found a code injection vulnerability and fixed this, without stating whether this concerned the LFI vulnerability mentioned earlier (Kirk, 2016; McDaniel, 2019).

There are methods available to test whether web applications and websites have LFI vulnerabilities, and vulnerability testing should be part of an organization's IT security program. Furthermore, a responsible operator should in general just never trust user input and configure the web application or website accordingly. On top of this, password were not encrypted or weakly encrypted. Based on these findings we consider the root cause for this data breach to be a security misconfiguration.

Parameters	Values for FriendFinder Networks
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Hacking
Breach Actor	Malicious Outsider

Organization Sector	Other
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Web Server Database
Data Responsibility	Data Controller

Data Subject	Customers Breachee
Data Amount (Records)	412.214.295
Data Type	PII LI OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Mixed

Data Breach Detection	External Non-Auto Inadvertent
Breach Period (Days)	Unidentifiable
Time between Breach and Detection	Medium
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 14 Summarized overview of identified values for FriendFinder Networks

5.2.4. MyFitnessPal (2018)

Acquired by Under Armour in 2015, MyFitnessPal is an application, mainly used as mobile app but also accessible through the website, which enables users to keep track of their calorie intake and their calories burned as a means for a healthy lifestyle (Bradley, 2018; Gonzalez & Jung, 2019).

On March 25, 2018, Under Armour announced that MyFitnessPal discovered that in February of the same year an unauthorized party gained access and was able to retrieve data related to 150 million MyFitnessPal user accounts (Agencies, 2018; Bradley, 2018; Gonzalez & Jung, 2019; McGee, 2018; MyFitnessPal, 2018; Under Armour, 2018). The breached data consists of usernames, email addresses, and passwords (Gonzalez & Jung, 2019; MyFitnessPal, 2018; Under Armour, 2018).

The largest part of the passwords were encrypted by using bcrypt, which is considered to be a secure password hashing method and protected that data despite being stolen (Bradley, 2018; McGee, 2018; Under Armour, 2018). The rest of the passwords, but also the usernames and email addresses, were only hashed by using SHA-1 hashing function, which is considered to be insecure and results in this part of the stolen data to be much more vulnerable (McGee, 2018; MyFitnessPal, 2018; Sprecher, 2018).

MyFitnessPal or its parent organization Under Armour did not provide further explanations of how access was gained and by whom, and how the data was pilfered, even after investigations in collaboration with law enforcement and security firms (Agencies, 2018; MyFitnessPal, 2018; Sprecher, 2018). One year later more than 150 million MyFitnessPal accounts records were offered for sale on the dark web as part of a large trove of data stolen from multiple organizations, with each record holding a user ID, username, email address, IP address and weakly encrypted passwords (Williams, 2019). The entity offering the large trove of data claim to have exploited typical security vulnerabilities within web applications to deploy remote code execution and from there went on to take user account data (Williams, 2019).

Although there not provided a lot of clues to the public with regard to where the root cause of this data breach lies, MyFitnessPal's and Under Armour's focus on the protection of the stolen data and on data that was not stolen (MyFitnessPal, 2018; Under Armour, 2018), seems to be there to divert the attention of possible faulty security. Together with some reports indicating the attack vector was in MyFitnessPal's web application and its possible vulnerabilities (Gonzalez & Jung, 2019; Williams, 2019), leads us to believe the root cause concerns a security misconfiguration.

Parameters	Values for MyFitnessPal
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Hacking
Breach Actor	Malicious Outsider

Organization Sector	Other
Organization Size	Medium
Organization Headquarters	US-USA-840
Organization Component	Database
Data Responsibility	Data Controller

Data Subject	Customers
Data Amount (Records)	150,633,038
Data Type	PII LI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	Internal Unidentifiable Unidentifiable
Breach Period (Days)	Unidentifiable
Time between Breach and Detection	Short
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration
References	[Numbers still to be inserted]

Table 15 Summarized overview of identified values for MyFitnessPal

5.2.5. *Coffee Meets Bagel (2017/2018)*

The large data trove offered for sale in a digital marketplace on the dark web in February 2019, included more than six million accounts of Coffee Meets Bagel's users next to hundreds of millions of data records from other organizations (O'Donnell, 2019a; Perez, 2019; Williams, 2019). Coffee Meets Bagel (CMB) is a mobile dating application offered by the company with the same name (Crunchbase, 2020; Perez, 2019).

After noticing the data pertaining to their users being advertised on February 11 of 2018, CMB three days later emailed a notification to its customers "that an unauthorized party had gained access to a partial list of user details" and that therefore the email's recipient may have his or her data stolen (Coffee Meets Bagel, 2019). In total it consists of 673 MB of data stolen late 2017 and mid-2018 (Dickey, 2016; Williams, 2019). In their email notification and in other communications CMB stated that the data only concerned names and email addresses of customers who became a member before May 2018 (Coffee Meets Bagel, 2019; Doe, 2019a; Perez, 2019). However, outside reporting says that records consist of full names, email addresses, age, registration dates, and gender (O'Donnell, 2019a; Perez, 2019; Williams, 2019).

CMB did not make statements about how the unauthorized party had accessed their system and data, and how the data got exfiltrated, only that they were still investigating this in collaboration with forensic security (de Looper, 2019; Perez, 2019). We did not find any other reports explaining the data breach details. Additionally, taken into account the reported statement from the data seller on the dark web marketplace telling us that typical web application security vulnerabilities were exploited to execute remote-code and eventually exfiltrate data (Williams, 2019), we assess this case to have a security misconfiguration as root cause.

Parameters	Values for Coffee Meets Bagel
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Hacking
Breach Actor	Malicious Outsider
Organization Sector	Other
Organization Size	Medium
Organization Headquarters	US-USA-840
Organization Component	Database
Data Responsibility	Data Controller
Data Subject	Customers
Data Amount (Records)	6.174.513
Data Type	PII
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary
Data Breach Detection	External Non-Auto Inadvertent
Breach Period (Days)	Unidentifiable
Time between Breach and Detection	Long
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration
References	[Numbers still to be inserted]

Table 16 Summarized overview of identified values for Coffee Meets Bagel

5.2.6. Government Payment Service Inc. (2018)

GovPayNet is an online service offered at GovPayNow.com by Government Payment Service Inc., and this service is used by state and local governments to receive payments with regard to fines, licensing fees, bail, taxes, etc (Krebs, 2018c; Osborne, 2018a). IT Security blogger Krebs (2018c) reports that Government Payment Service Inc. has exposed *"more than 14 million customer records dating back [to 2012], including names, addresses, phone numbers and the last four digits of the payer's credit card"*, until notified by him. Krebs (2018c) further reports that his blog received a statement from Government Payment Service Inc. in which the issue is confirmed: *"GovPayNet has addressed a potential issue with our online system that allows users to access copies of their receipts, but did not adequately restrict access only to authorized recipients"* (Krebs, 2018c).

Part of GovPayNet's service, is that when a payment was done at GovPayNow.com, an online receipt would be displayed to the payer. However, by changing the digits in the URL of an online receipt, receipts from other customers could be viewed (Krebs, 2018c). This enabled unauthorized parties to access customer receipts, which was confirmed by Government Payment Service Inc. (Krebs, 2018c). The receipts were stored and numbered sequentially, represented by the digits in the URL, which made it easy to count the number of potentially exposed records (Krebs, 2018c; Muncaster, 2018a). Because of this, we also assume this issue existed since the first customer record, which would mean the exposure period was very long.

This data exposure was preventable by taking into account potential data disclosure vulnerabilities in advance, since these were known and common vulnerabilities and GovPayNow.com (and the whole GovPayNet service) could easily have been set up more securely (Daitch, 2018; Krebs, 2018c). In this case, Government Payment Service Inc. should not have used sequential record numbers and additionally could have applied encryption on unique portions of the URL belonging to the displayed receipt by using HTTPS when displaying the receipts (Daitch, 2018; Krebs, 2018c; Robinson, 2018b). Furthermore, it should have been set up in such a way that a user needs to be logged in and needs to have permission to view the receipt for the payment that user has done, and that user does not have permission to view other receipts (Krebs, 2018c; Robinson, 2018b; Taylor, 2018).

Parameters	Values for Government Payment Service Inc.
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider

Organization Sector	Government and Military
Organization Size	Medium
Organization Headquarters	US-USA-840
Organization Component	Web Server
Data Responsibility	Data Controller

Data Subject	Customers
Data Amount (Records)	14.000.000
Data Type	PII PIFI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	External Non-Auto Inadvertent
Breach Period (Days)	Unidentifiable
Time between Breach and Detection	Long
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 17 Summarized overview of identified values for Government Payment Service Inc.

5.2.7. South-East Regional Health Authority Norway (2018)

The South-East Regional Health Authority (SERHA) is an organization that manages healthcare institutions in 18 counties in Norway's southeast (Ashford, 2018a; Cimpanu, 2018a). When Norway's Computer Emergency Response Team (CERT) for the healthcare sector (HelseCERT) discovered suspicious traffic from SERHA's network (Helse Sør-Øst, 2018), an investigation followed at SERHA's ICT provider which revealed there has been a data breach (Ashford, 2018a; Cimpanu, 2018a; Sæther, 2018). After seven days, SERHA confirmed that 2.9 million citizens' healthcare records may have been exposed (Ashford, 2018a), because what was sure is that unauthorized access into SERHA's network had taken place (I. Ashok, 2018a; Helse Sør-Øst, 2018).

From the coverage of this case it appears that SERHA and its partners displayed strategic behaviour after this incident. Firstly, SERHA stated it did have not an indication that patient information had actually been taken, but at the moment of that statement it was too early to be sure about that as the investigation was still going on (I. Ashok, 2018a; Cimpanu, 2018a). Several sources include the statement that SERHA and its partners were actively working on resolving the issue and investigating the cause, details, size and scope of the data breach (I. Ashok, 2018a; Cimpanu, 2018a; Helse Sør-Øst, 2018). Our assessment is that if it is too early to be sure whether personal information had been taken because it is not known what happened and how and who is involved, then it is too early to make the statement telling patients that there is no indication that personal information was actually taken. Besides, according to our definition of a data breach (chapter 1, section 3), it does not matter whether the data was actually taken or not.

Also, SERHA appears to downplay their own role and responsibility in this incident, by stating that the intrusion was carried out in an advanced manner by a skilled party (Helse Sør-Øst, 2018; Sæther, 2018; Sæther, Bugge, & Sarmadawy, 2018), thus try to sell it as an sophisticated attack which they could not prevent. Because it seems contradictory to be able to conclude that it was a sophisticated attack while also stating that the investigation is still ongoing and that the nature of the attack far from completely known. This leads us to assess that SERHA either did not know the specifics yet but already tried to minimize their responsibility and liability, or SERHA did know sufficient about the incident but chose not to disclose (yet) because of certain reasons. Involved parties did hint to the latter, by stating that for the sake of incident handling there could not be present further details concerning the incident (Mikalsen, 2018; Sæther, 2018).

We were not able to find the results of the mentioned investigation, which leads us to believe this was not disclosed eventually. The disclosure of and coverage on this incident appears to include bias and was not specific enough to determine whether this was a sophisticated attack or a security misconfiguration.

Parameters	Values for South-East Regional Health Authority Norway
Data Breach Type	Unauthorized Access
Breach Method	Hacking
Breach Actor	Malicious Outsider

Organization Sector	Medical and Healthcare
Organization Size	Large
Organization Headquarters	NO-NOR-578
Organization Component	Inconclusive
Data Responsibility	Data Controller

Data Subject	Patients Citizens
Data Amount (Records)	2.900.000
Data Type	PII PHI
Data Sensitivity	Internal
Data State	Inconclusive
Data Necessity	Necessary

Data Breach Detection	External Automated Intended
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Inconclusive

Table 18 Summarized overview of identified values for South-East Regional Health Authority Norway

5.2.8. MyHeritage (2017)

On June 4, 2018, MyHeritage announced that it had suffered from a data breach (Deutsch, 2018). MyHeritage is a business in the genealogy sector which offers direct-to-consumer genetic testing and ancestry services (Krebs, 2018b; T. Wong, 2019). Nowadays MyHeritage also offers tests that result in comprehensive health reports, but this was not the case at the time of the data breach (Reuters, 2019).

According to the announcement, on June 4, 2018 a security researcher notified MyHeritage's Chief Information Security Officer that he had found a file named 'myheritage' on a private server outside of MyHeritage (Deutsch, 2018). This file was reviewed by MyHeritage's Security Team which confirmed that the email addresses and hashed (plus salted) passwords belonged to 92,283,889 users that signed up to MyHeritage up to and including October 26, 2017 (Deutsch, 2018). MyHeritage states that October 26, 2017 is the date the data breach took place (Deutsch, 2018). But MyHeritage did not explain how it came to this assessment however, whether it was only based on the fact that the file contained accounts created up to that date, or that additional investigation reinforced this. In the announcement MyHeritage stated that family ancestry data and genetic data are stored on segregated systems with added layers of security, while financial information is processed by and stored at trusted third parties (Deutsch, 2018).

The MyHeritage data resurfaced eight months later as part of the large data trove of 620 million accounts that was put up for sale on the dark web on February 11, 2019 (Williams, 2019). These 92,284,478 account records contained email addresses, SHA1-hashed passwords and salt, and the date of account creation (Williams, 2019). Samples from the offered data trove were confirmed to be real by MyHeritage (Williams, 2019). According to MyHeritage this data is almost exactly the same as the data from the October 2017 data breach, and therefore assumes that it came from the October 2017 data breach (Williams, 2019).

The sales offer on the dark web also made clear that the passwords were encrypted with a weak hashing mechanism (Williams, 2019). This was already suspected because in its announcement after discovering the October 2017 data breach, MyHeritage did not disclose specifically which hashing mechanism it had used in what manner for encrypting user passwords, while one would expect a data breach victim would be eager to tell if a secure hashing mechanism was used as an eased circumstance (Krebs, 2018b).

MyHeritage did not elaborate on how the data was taken from their network and by whom. Regarding the sales offer on the dark web, this was likely the work of malicious outsiders rather than an inadvertent insider. These malicious outsiders would probably try to decrypt the user passwords, or sell the encrypted passwords to other parties that would try to decipher them. Because of the weak hashing mechanism and the available tools they would likely have been successful to some extent (CipherCloud, 2019; Krebs, 2018b). With those passwords they would not only gain access to MyHeritage users' accounts, but they would also possibly gain access to MyHeritage users' personal email boxes since a lot of people reuse password over different platforms. This would potentially enable the malefactors with access to genetic information, as email confirmation was a step in the download process of genetic information (Forman, 2018). Furthermore, they could also circumvent the password reset procedure which also uses email confirmation. MyHeritage did not expire user accounts and force this procedure on its customers immediately after the data breach notification (Krebs, 2018b). Because of the access that could easily have been gained to MyHeritage users' personal email accounts, the data breach could have been much worse, since more sensitive data could possibly have been taken. Genetic information is considered to be the one of the most sensitive types of information there is (CipherCloud, 2019; T. Wong, 2019).

With such sensitive information involved, MyHeritage should have implemented a more secure authentication method, for instance multifactor authentication, which should be obligatory to use by each user. If password authentication would still be used as a part of that, these passwords should have used a more secure hashing mechanism.

Hence, security misconfigurations were present anyway, even if we do not know exactly how the data was taken during the data breach. However, we cannot be sure that if these security misconfigurations were not present and that if this was configured securely, that the data breach would not have happened. Additionally,

MyHeritage has not been transparent about what happened either. Because of the above, we cannot assess whether the data breach was the result of a complex attack or a security misconfiguration.

Parameters	Values for MyHeritage
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Inconclusive
Breach Actor	Malicious Outsider
Organization Sector	Other
Organization Size	Large
Organization Headquarters	IL-ISR-376
Organization Component	Database
Data Responsibility	Data Controller
Data Subject	Customers
Data Amount (Records)	92,284,478
Data Type	PII LI OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary
Data Breach Detection	External Non-Auto Intended
Breach Period (Days)	Short
Time between Breach and Detection	Long
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Inconclusive

Table 19 Summarized overview of identified values for MyHeritage

5.2.9. ClixSense (2016)

On September 10, 2016 hackers published links in a post on PasteBin leading to files with more than 2.4 million of ClixSense user accounts (Hunt, 2016; Khandelwal, 2016; Wagner, 2016). ClixSense was a company that claimed to pay users for clicking on and viewing ads, as well as for completing online surveys (Goodin, 2016; Wilson & Hingnikar, 2019). The account records consisted of first and last names, dates of birth, gender, usernames, plaintext passwords, email addresses, home addresses, IP addresses, accounts balances and payment histories (Goodin, 2016; Khandelwal, 2016). Security expert Troy Hunt, who also is the creator and operator of the website 'Have I Been Pwned', checked the files and verified the authenticity of the data (Goodin, 2016; Khandelwal, 2016).

The PasteBin post advertised stolen ClixSense data and said the published data only concerned a sample of personal information of more than 6.6 million ClixSense user accounts which were taken from a compromised database (Goodin, 2016). The complete data set was said to be consisting of 6,606,008 user account records and also included users' Social Security Numbers and answers to security questions, and next to the user accounts even 70,000 internal emails and the website's complete source code (Goodin, 2016). We consider this case to have 6676009 records breached (6,606,008 plus 70,000 plus 1). The post also mentioned that the sample data was publicly released because ClixSense initially denied having suffered a data breach where (user) data was taken (Goodin, 2016; Kovacs, 2016). Another interesting statement in the post from the attackers mentions that servers were only protected with weak credentials with no additional security (Goodin, 2016; Kovacs, 2016).

Furthermore, the post mentioned that the most of the personal information is from very recent and that the sample list includes only the early users, i.e. the oldest personal information (Goodin, 2016). The more recent the personal information is, the more valuable. Interested parties had the opportunity to bid and the highest bidder would then obtain the rest of the data (Khandelwal, 2016).

On September 4, 2016 ClixSense already noticed it was the victim of an incident, which began with its website being redirected to a porn site by hackers that took over the Domain Name System (Goodin, 2016). The day after the hackers gained access to ClixSense's hosting provider and turned off all its servers, gained access to its Microsoft Exchange server and changed the password of all ClixSense's email accounts (Goodin, 2016). On September 6 the perpetrators gained access to an old unused and unsecured server which was still directly connected to their main database server, from which ClixSense's user table was copied (Goodin, 2016; Khandelwal, 2016). ClixSense confirmed that this database contained more than 6.6 million user accounts (Goodin, 2016). Furthermore, code was ran to change account names, forum posts were deleted, and account balances were set to zero (Khandelwal, 2016). All this was disclosed by ClixSense itself and it confirmed the claims done in the PasteBin post, but not before September 11 (Goodin, 2016; Khandelwal, 2016; Wagner, 2016).

After discovering the breach, ClixSense terminated the outdated vulnerable server (Goodin, 2016; Osborne, 2016). But of course, by then it was already too late and this should have been done before or at least should network segmentation and segregation could have been implemented to securely seclude that server. Furthermore, ClixSense made a mistake by storing passwords in plaintext which also is not according to standard industry practices. In addition, the servers were poorly secured. The ClixSense system designers and system operators involved with the website and back end either did not know of and understand security practices, or chose to ignore them for the sake of other factors (Cobb, 2017). We consider these to be security misconfigurations.

Parameters	Values for ClixSense
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Poor or No Security
Breach Actor	Malicious Outsider

Organization Sector	Other
Organization Size	Small
Organization Headquarters	US-USA-840
Organization Component	Database Mail Server Mailbox Web Server
Data Responsibility	Data Controller

Data Subject	Members
Data Amount (Records)	6.676.009
Data Type	PII PIFI LI OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	Internal Non-Auto Inadvertent
Breach Period (Days)	3
Time between Breach and Detection	Short
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 20 Summarized overview of identified values for ClixSense

5.2.10. SVR Tracking (2017)

SVR tracking is a company which offers GPS tracking solutions in the USA (SVR Tracking, n.d.-a). These solutions enable tracking of vehicles my means of detailed location data and monitoring (SVR Tracking, n.d.-a). In practice, it was used by SVR Tracking's customers mainly to track their vehicles, in real time as well as a limited past time, by using a hidden physical GPS tracking device to monitor and recover these vehicles in case they are stolen (Khandelwal, 2017a). For this they used the SVR app, in which the customer has to log in (Waqas, 2017). In the meantime, SVR has expanded their assortment of solutions (SVR Tracking, n.d.-a). The solutions are built into a number of applications (SVR Tracking, n.d.-a).

On September 21, 2017, SVR Tracking was notified by the Kromtech Security Center (KSC) that it had discovered a publicly accessible online storage facility containing a database with data belonging to SVR Tracking (Khandelwal, 2017a; Sushko, 2017; Wells, 2017). KSC reportedly discovered the database already on September 18 (Cameron, 2017; Waqas, 2017). KSC needed a day to determine who the data belonged to and when it discovered it belonged to SVR Tracking, SVR Tracking was notified on September 20 (Cameron, 2017).

This online database was used to store backups of customer related data (Wells, 2017). It concerned a Amazon Simple Storage Services (S3) Bucket, the cloud storage service offered in Amazon's Web Services, and this cloud storage bucket was misconfigured which left it unsecured (Khandelwal, 2017a; Paganini, 2017; Sushko, 2017; Waqas, 2017; Wells, 2017).

Hence, KSC had access to the database and found it to contain 540,642 SVR Tracking account records with their ID numbers, customer names, email addresses, passwords, license plate numbers, vehicle identification numbers and the GPS devices' IMEI numbers (Khandelwal, 2017a; Sushko, 2017; Wells, 2017). This database included information indicating the exact location of the GPS device within the car (Khandelwal, 2017a). Furthermore, the database also contained 339 logs with photographic data, vehicle status data and maintenance data, as well as a document with customer data pertaining to dealerships (Khandelwal, 2017a). We consider this to be a total of 540,982 (540,642 plus 339 plus 1). By the way, many account holders had multiple devices, which means it concerns multiple vehicles (Khandelwal, 2017a; Waqas, 2017).

Some sources reported that the passwords were hashed with SHA-1 which is a hashing mechanism offering weak security (Khandelwal, 2017a; Paganini, 2017; Sushko, 2017; Waqas, 2017). However, when SVR Tracking came out with its Incident Response it stated that the passwords were hashed with SHA-256 Cryptographic Hash Algorithm (Wells, 2017). SHA-256 is considered to be much more secure than SHA-1 (Gilbert & Handschuh, 2004).

After being notified, SVR Tracking immediately reviewed the situation and configured the storage bucket correctly in order to prevent further unauthorized access (Khandelwal, 2017a; Sushko, 2017; Waqas, 2017; Wells, 2017). On September 18, SVR Tracking imposed a forced password reset (Wells, 2017).

SVR Tracking investigated the data breach and received cooperation from KSC, where KSC provided the IP addresses it used in order for SVR Tracking to verify that KSC's IP address was the only unauthorized IP address that had accessed the data before the misconfiguration was resolved (Wells, 2017). SVR Tracking did not disclose how long the database was publicly accessible, while only SVR Tracking and Amazon can know this for sure, but of course are not eager to share this information (Cameron, 2017). And the exposure period could not be determined from the other resources, but of course this does not mean there was no data breach. Even if SVR Tracking concluded that no other IP addresses than KSC's had unauthorizedly accessed the data. Our definition of a data breach does no require access to have taken place.

Obviously, since SVR Tracking admitted this data breach exposing customer and business information was due to a configuration issue, this case is labelled to be caused by a security misconfiguration.

Parameters	Values for SVR Tracking
Data Breach Type	Unauthorized Access
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider

Organization Sector	Technology
Organization Size	Small
Organization Headquarters	US-USA-840
Organization Component	Cloud Storage
Data Responsibility	Data Controller

Data Subject	Customer
Data Amount (Records)	540.982
Data Type	PII LI OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	External Non-Auto Intended
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 21 Summarized overview of identified values for SVR Tracking

5.2.11. SKY Brasil (2018)

SKY Brasil (SKY) is a telecommunications company which offers subscription based Internet and TV services and one of Brazil's largest providers (Crunchbase, n.d.-g; Ilascu, 2018; SKY Brasil, 2020). SKY used ElasticSearch, including an ElasticSearch server, for storing business related data. ElasticSearch is a search engine known for its powerful and fast storage, search and analysis capabilities which therefore became a favourite (Pascu, 2018) (Cimpanu, 2018b; Elastic, 2020; Pagano Dritto, 2019).

On November 22, 2018, security researcher Fabio Castro found an ElasticSearch server of SKY publicly accessible on the Internet by using the advanced features of the Shodan search engine (Ilascu, 2018; Pascu, 2018). Shodan is used as a tool to find and index devices and systems connected to the Internet, and is often used by organizations to check and improve their cybersecurity (Bodenheim, Butts, Dunlap, & Mullins, 2014; Cimpanu, 2018b). The encountered server was accessible through two IP addresses and was indexed on Shodan since at latest mid-October (Cimpanu, 2018b).

According to the security researcher, the server stored 28.7 GB of log files and 429.1 GB of API data, of which the latter contained information of 32 million SKY residential and business customers (Cimpanu, 2018b). The data included full names, birth dates, home addresses, phone numbers, email addresses, encrypted passwords, customer IP addresses, and billing details (Cimpanu, 2018b; Ilascu, 2018). Furthermore, the researcher encountered customer specific product and service data, including the models and serial numbers of the devices used by the customers (Ilascu, 2018).

After verifying the data and who the data belonged to, Castro notified SKY the same day (Cimpanu, 2018b). SKY never responded directly to Castro, but the researcher noticed the was secured on the next Monday morning, November 26, with a password restricting unauthorized access (Cimpanu, 2018b). It is not clear how long the data had been exposed exactly, and if anyone had accessed the data illegitimately before that time (Abel, 2018).

The cause of this data breach was that the system operators did not set up password protection for this server, or in general did not setup authentication and authorization for this server and the data it held (Cimpanu, 2018b; Pascu, 2018). Elastic stated that their servers had been developed primarily for internal network use and were not meant to be exposed to the Internet, which was why setting up authentication or authorization were not default or required steps (Bressers, 2019; Cimpanu, 2018b). Even so, we consider this to be the responsibility of the system operators. Especially since at that time misconfigured ElasticSearch servers were emerging more and more, and security experts even were already predicting that these misconfigurations would be causing more data breaches (Cimpanu, 2018b; Ilascu, 2018; Mares, 2018). We conclude that this data breach was caused by a security misconfiguration.

Parameters	Values for SKY Brasil
Data Breach Type	Unauthorized Access
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider

Organization Sector	Technology
Organization Size	Large
Organization Headquarters	BR-BRA-076
Organization Component	Cloud Storage
Data Responsibility	Data Controller

Data Subject	Customers
Data Amount (Records)	32.000.000
Data Type	PII LI OBI PIFI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	External Automated Intended
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 22 Summarized overview of identified values for SKY Brasil

5.2.12. LocalBlox (2018)

LocalBlox is a company which collects data from the Internet to build all kinds of profiles that it can sell to interested parties (Johnson, 2018). Among these, LocalBlox produces psychographic profiles, consisting of personal data scraped from public sources (Johnson, 2018; Whittaker, 2018a). Whittaker (2018a) reports: *"a little-known data firm was able to build 48 million personal profiles, combining data from sites and social networks like Facebook, LinkedIn, Twitter, and Zillow, among others -- without the users' knowledge or consent"*. The scraping of personal data from sites and social networks is against their terms (Johnson, 2018; Whittaker, 2018a). However, LocalBlox places the responsibility for preventing scraping at the individual sources (Johnson, 2018). Additionally, information is sometimes collected by LocalBlox from non-public sources, for instance by purchasing data (Johnson, 2018).

On February 18, 2018, the UpGuard security researcher and ethical data breach hunter Chris Vickery and his Cyber Risk Team (CRT) discovered a publicly accessible and downloadable file containing 48 million psychographic profiles on a cloud storage repository (UpGuard, 2018). This storage concerned an Amazon Web Services (AWS) Simple Storage Service (S3) bucket, which was left unlisted and unprotected (Johnson, 2018). UpGuard (2018) reports that *"[t]he bucket contained one 151.3 GB compressed file, which, when decompressed, revealed a 1.2 TB ndjson (newline-delineated json) file"*. UpGuard (2018) further reports: *"the massive file contains 48 million records, each in json format and separated by new lines."* Such files are readable by humans (Whittaker, 2018a).

The file was download and analysed (UpGuard, 2018). The 48 million records varied in the type and amount of data they have. In general, the data consisted of personally identifiable information, including names, dates of birth, physical addresses, past and current employment information, email addresses, IP addresses, phone numbers, postal addresses, and other information (UpGuard, 2018; Whittaker, 2018a). Sometimes marital status, net worth and contact preferences are also included (Whittaker, 2018a). The compromised data does not concern customers of LocalBlox, but rather is the product that LocalBlox offers (UpGuard, 2018). Hence, we will label the data subjects as citizens, of which a maximum of 48 million are affected since a single individual can have multiple records (UpGuard, 2018).

Paganini (2018b) reports that *"[t] he analysis of metadata in a header file allowed the researchers to attribute it to LocalBlox"*. On February 28, UpGuard's CRT notified LocalBlox of the data breach and later that day the cloud storage repository was secured (UpGuard, 2018). Whittaker (2018a) reports that they had contact with LocalBlox's Chief Technology Officer (CTO) who stated that apart from UpGuard, *"no other individual is believed to have accessed this file from the S3 bucket"*. Whittaker (2018a) further reports that the same CTO stated that *"most of the data was fabricated and for internal tests"*. However, the CTO did not want to specify what most meant (Whittaker, 2018a). Because of the size of the found file and the fact that LocalBlox secured that cloud storage, we consider this to be strategic behaviour to downplay or even deny a data breach. And by the way, internal testing should definitely include testing the security of sensitive data which apparently was not the case.

This data breach was caused by LocalBlox not having set up password protection, or any other form of authentication and authorization for the concerned repository. Especially a business like LocalBlox which handles and stores people's personal data and already does so in a questionable manner, should handle the involved data more responsibly and securely (Johnson, 2018; Paganini, 2018b). Would this have been configured correctly, this data breach would have been prevented. This case clearly concerns a security misconfiguration (Claburn, 2018; UpGuard, 2018).

Parameters	Values for LocalBlox
Data Breach Type	Unauthorized Access
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider
Organization Sector	Technology
Organization Size	Small
Organization Headquarters	US-USA-840
Organization Component	Cloud Storage
Data Responsibility	Data Controller
Data Subject	Citizens Breachee
Data Amount (Records)	48.000.000
Data Type	PII
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Mixed
Data Breach Detection	External Inconclusive Intended
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 23 Summarized overview of identified values for LocalBlox

5.2.13. MBM Company (2018)

MBM Company Inc. is a firm which mainly operates as its jewellery brand Limogés Jewelry and partners with businesses like Walmart (I. A. Ashok, 2018b; NJCCIC, 2018; Wilson & Hingnikar, 2019). On February 6, 2018, the German security firm Kromtech Security Centre (KSC) encountered an unprotected Amazon Web Services (AWS) Simple Storage Services (S3) bucket (Hughes, 2018; O'Donnell, 2018). Robinson (2018a) reports: *"The open S3 bucket, named "walmartsql," housed an MSSQL database backup."* Initially, the bucket name did make the data appear to belong to Walmart (O'Donnell, 2018; Paganini, 2018a).

After further analysing this file, KSC discovered it contained personal information consisting of 1,314,193 million customer records (I. A. Ashok, 2018b; Hughes, 2018; Paganini, 2018a). Also, KSC was able to determine that the file and its content actually belonged to MBM Company (O'Donnell, 2018; Paganini, 2018a). NJCCIC (2018) reports the following: *"The open S3 bucket, named "walmartsql," contained customers' names, addresses, ZIP codes, phone numbers, email addresses, IP addresses, plaintext passwords, encrypted credit card numbers, and payment details for purchases made between 2000 and early 2018."* This concerns mainly Limogés Jewelry customers (NJCCIC, 2018; O'Donnell, 2018).

Cluley (2018) mentions in his blog post: *"Limogés Jewelry is sold through Walmart, but it's also sold via other major retailers including Amazon, Sears, Kmart, Target, and countless online third-party stores. And the exposed database was found to contain database records connected with jewelry purchases from many other retailers besides just Walmart."*

O'Donnell (2018) presents the following from KSC's report: *"It also contained internal MBM mailing lists, encrypted credit card details, payment details, promo codes, and item orders, which gives the appearance that this is the main customer database for MBM Company Inc. Records were seen with dates ranging from 2000 to early 2018."*

Hughes (2018) reports the following about the database backup file: *"The backup file was named 'MBMWEB_backup_2018_01_13_003008_2864410.bak,' which suggests the file was created on January 13, 2018."* This also suggests the cloud storage bucket and the database file were publicly accessible at least from that date (NJCCIC, 2018; O'Donnell, 2018).

KSC stated that it had notified Walmart of the unsecured AWS S3 bucket immediately after discovering it (O'Donnell, 2018). Right away Walmart quietly made sure the bucket was secured (NJCCIC, 2018; O'Donnell, 2018; Paganini, 2018a). Around the time of discovery, there were no indications that the data had been accessed or taken (Hughes, 2018; O'Donnell, 2018). But even so, this is still considered a data breach by our definition since the data was left exposed, and for quite some time.

The unsecured AWS S3 bucket causing this data breach concerns a misconfiguration (NJCCIC, 2018; O'Donnell, 2018). Especially since before this data breach occurred, several incidents had taken place as the result of AWS S3 buckets that had not been configured properly (I. A. Ashok, 2018b; NJCCIC, 2018; O'Donnell, 2018). Furthermore, to use an unprotected AWS S3 bucket to store an unprotected database file made it even worse (I. A. Ashok, 2018b; O'Donnell, 2018). But having stored the passwords in plaintext in that database was the icing on the cake (I. A. Ashok, 2018b; O'Donnell, 2018).

Parameters	Values for MBM Company
Data Breach Type	Unauthorized Access
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider
Organization Sector	Retail
Organization Size	Medium
Organization Headquarters	US-USA-840
Organization Component	Cloud Storage
Data Responsibility	Data Controller
Data Subject	Customers
Data Amount (Records)	1.314.193
Data Type	PII PIFI LI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary
Data Breach Detection	External Non-Auto Intended
Breach Period (Days)	25
Time between Breach and Detection	Short
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 24 Summarized overview of identified values for MBM Company

5.2.14. Capital One (2019)

The financial corporation Capital One is one of the largest banks in the U.S. and the world, having 28 billion American dollars in revenue in 2018 and more than 50 thousand employees (Novaes Neto, Madnick, de Paula, & Malara Borges, 2020). Capital One was notified on July 17, 2019, by an anonymous person mailing to Capital One's responsible disclosure email address, and who stated that their leaked data surfaced on someone's Github page (Krebs, 2019d; Novaes Neto et al., 2020).On July 29, 2019, Capital One made the announcement that ten days earlier it had discovered there had taken place unauthorized access to its network (Lu, 2019; Novaes Neto et al., 2020). The unauthorized access already took place about four months earlier on March 22 or 23 of 2019 (Novaes Neto et al., 2020). Later, the indictment against the intruder speaks of a period between March 12, 2019, and July 17, 2019, during which the intruder gained access to Capital One's network (US District Court at Seattle, 2019b). This does not necessarily mean that it happened continuously between those dates without pause, in this instance it is far more likely the access and exfiltration happened on certain moments in that period. However, as it is not possible for us to exactly determine those moments, we assume the breach period to be that whole period.

The intruder was able to obtain personal information of 106 million Capital One's customers and individual credit card applicants (Capital One, 2019a). Capital One (2019a) publicized the following:

> *Based on our analysis to date, this event affected approximately 100 million individuals in the United States and approximately 6 million in Canada. . . . The largest category of information accessed was information on consumers and small businesses as of the time they applied for one of our credit card products from 2005 through early 2019. This information included personal information Capital One routinely collects at the time it receives credit card applications, including names, addresses, zip codes/postal codes, phone numbers, email addresses, dates of birth, and self-reported income. (Capital One, 2019a)*

Capital One (2019a) further reports:

> *Beyond the credit card application data, the individual obtained portions of credit card customer data, including:*
>
> - *Customer status data, e.g., credit scores, credit limits, balances, payment history, contact information.*
> - *Fragments of transaction data from a total of 23 days during 2016, 2017 and 2018.*
>
> *. . .*
>
> *The individual also obtained the following data:*
>
> - *About 140,000 Social Security numbers of our credit card customers.*
> - *About 80,000 linked bank account numbers of our secured credit card customers.*
>
> *. . .*
>
> *For our Canadian credit card customers, approximately 1 million Social Insurance Numbers were compromised in this incident. (Capital One, 2019a)*

The data was stored in Amazon's cloud storage facility, Amazon Web Services (AWS) Simple Storage Service (S3) buckets, which were used by Capital One for data storage operations (Lu, 2019; Novaes Neto et al., 2020).

The Github profile led to the hacker, who was a former employee of Amazon's cloud storage department, and was knowledgeable enough to be able to exploit a configuration vulnerability in Capital One's infrastructure (Lu, 2019). Krebs (2019d) reports on his blog: "*According to a source with direct knowledge of the breach investigation, the problem stemmed in part from a misconfigured open-source Web Application Firewall (WAF) that Capital One was using as part of its operations hosted in the cloud with Amazon Web Services (AWS). . . . the misconfigured WAF for whatever reason was assigned too many permissions . . .*"

Because of the misconfigured firewall, the hacker could run a Server Side Request Forgery (SSRF) attack (Krebs, 2019d; Novaes Neto et al., 2020). This SSRF method tricked Capital One's WAF into running commands it was not intended to run, and actually should not have been permitted to run (Krebs, 2019d; Novaes Neto et al., 2020). Krebs (2019d) reports: *"it was allowed to list all of the files in any buckets of data, and to read the contents of each of those files."* Eventually, this enabled the hacker to copy those contents (Krebs, 2019d). The vulnerability was immediately fixed by Capital One upon discovery (Lu, 2019).

SSRF was already a well- known attack method at the time (Krebs, 2019d). Furthermore, Novaes Neto et al. (2020) showed several things that could have been done by Capital One to prevent unauthorized access and data exfiltration. On Amazon's part, also some things could have been added to the AWS which would probably have prevented the data breach, but Amazon was quick to deny any responsibility in this incident (Krebs, 2019d; Novaes Neto et al., 2020). Whoever responsible for the misconfiguration is not relevant for our research, we are interested in its presence as the root cause for a data breach. In this case, a security misconfiguration was definitely the cause (Capital One, 2019b; CloudSploit, 2019; Lu, 2019; Novaes Neto et al., 2020; US District Court at Seattle, 2019a).

Parameters	Values for Capital One
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Hacking Poor or No Security
Breach Actor	Malicious Outsider

Organization Sector	Financial and Insurance
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Cloud Storage
Data Responsibility	Data Controller

Data Subject	Customers
Data Amount (Records)	106.000.000
Data Type	PII PIFI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	External Non-Auto Inconclusive
Breach Period (Days)	127
Time between Breach and Detection	Long
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 25 Summarized overview of identified values for Capital One

5.2.15. European Central Bank (2014)

The European Central Bank (ECB) is “the central bank of the 19 European Union countries which have adopted the euro”, and its main task is “to maintain price stability in the euro area and so preserve the purchasing power of the single currency” (European Central Bank, 2020). In 2014, the ECB had an average number of 2155 staff employed (European Central Bank, 2015).

On July 24, 2014, the ECB announced that here had been an security incident involving its website, and eventually resulting in a security breach of the database serving the public ECB website (European Central Bank, 2014; Honan, 2014). The database serves part of the website that processes registrations for events like conferences and visits hosted by the ECB (European Central Bank, 2014; Honan, 2014). The ECB refused to comment on when the data breach occurred and for how long.

The ECB was unaware of the data breach, until on July 21, 2014, they were contacted anonymously by a party claiming to have accessed their infrastructure and taken the data, and which sought financial compensation for the data to be returned (European Central Bank, 2014; Honan, 2014; Schwartz, 2014). Further, that party tried to extort the ECB by threatening to publish the stolen data if the financial demands were not met by the ECB (Honan, 2014). The ECB refused to comply with any demands (Honan, 2014). Upon discovering the data breach, the ECB security team went on to find and fix the exploited vulnerability (Schwartz, 2014). The ECB did not want to elaborate on where and how the intruders gained access (Schwartz, 2014).

The ECB did not specifically state which data was in the database, only that it concerned contact information for event registrants and that most of the data was encrypted, while “parts of the database included email addresses, some street addresses and phone numbers that were not encrypted” (European Central Bank, 2014). The database also held data on downloads (which may have been done by individuals) from the ECB website, and these were encrypted (European Central Bank, 2014; Schwartz, 2014). The ECB refused to provide further technical details on which encryption methods were used (Schwartz, 2014).

According to an ECB spokeswoman the database stored approximately 20 thousand email addresses, and a smaller number of other data (Schwartz, 2014). The ECB explicitly wanted to make clear that the database was physically separate from any internal ECB systems, and that no other sensitive data were compromised (European Central Bank, 2014). While not mentioned, we assume other personal information like names, dates of births, etc. were also involved.

ECB’s security team quickly identified and fixed the exploited vulnerability (Honan, 2014), because of which we tend to think a simple misconfiguration was the cause. Furthermore, the ECB is very secretive and provides minimal information about the data breach, which organizations tend to do faster in the case of a security misconfiguration than if a sophisticated attack would have been the cause. Additionally, the fact that part of the data was unencrypted indicates a security error but does not per se mean there was a misconfiguration that is the root cause for the data breach (i.e., if this specific misconfiguration would not be present, it does not mean the data breach would not have occurred).

However, assuming that this data breach was caused by a security misconfiguration based on these superficial findings alone would be speculative. Hence, the reports and coverage could not provide us with sufficient information and thus it remains inconclusive whether in this case the root cause was a security misconfiguration or a complex attack.

Parameters	Values for the European Central Bank
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Hacking
Breach Actor	Malicious Outsider

Organization Sector	Financial and Insurance
Organization Size	Large
Organization Headquarters	DE-DEU-276
Organization Component	Database
Data Responsibility	Data Controller

Data Subject	Other
Data Amount (Records)	20.000
Data Type	PII OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	External Non-Auto Intended Notified by the Breach Actor
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Inconclusive

Table 26 Summarized overview of identified values for the European Central Bank

5.2.16. JP Morgan Chase (2014)

Ranked in the top ten of the world's 100 largest banks (Ali, 2020), JP Morgan Chase (JP) is a financial services company offering its services worldwide (J.P. Morgan, 2020). In June 2014 a JP employee's personal computer got infected with malware, which probably happened because this employee clicked on a phishing email attachment or link which then would have redirected the employee to a website with malware (Jeng, 2015). The infected computer resulted in the employee's login credentials being stolen (Jeng, 2015). The hackers that stole these login credentials were able to gain access to JP's internal network when the compromised employee used VPN to connect to JP's network (Jeng, 2015).

Once inside, the hackers ran malicious programs which were specifically designed to be able to move through JP's network while avoiding being stopped by security (Jeng, 2015). By exploiting several zero-day vulnerabilities, the hackers managed to successfully obtain the highest administrator privileges to and take over control of more than ninety servers (Jeng, 2015). Data exfiltration was spread out over a longer period of time to avoid being detected (Jeng, 2015). The hackers had access over a period of two months (Perlroth & Goldstein, 2014; Roman, 2014b).

Late July 2014, the computer security firm Hold Security discovered a repository of a billions stolen usernames and passwords, a part of which of people who had registered for charity races on JP's Corporate Challenge website which is separated from JP's corporate network (Perlroth & Goldstein, 2014). Hold Security notified JP of its findings (Perlroth & Goldstein, 2014), which triggered JP's security department to look into its own corporate network, which then resulted in discovering JP's corporate network also suffered a breach (Jeng, 2015). Shortly after discovering the breach, the security stopped the intrusion (Perlroth & Goldstein, 2014).

The hackers partly succeeded in concealing their activities by deleting a significant number of log files (Jeng, 2015). Likely, if Hold Security had not discovered the stolen data, the hackers would have had access for a longer period of time (Jeng, 2015). Hence, with the risk of stealing more data in terms of amount and sensitivity, or in general more risk that this would have resulted in a more severe incident.

The data stolen by the hackers contained names, addresses, email addresses and phone number for 76 million households and 7 million small businesses (Jeng, 2015). Partly included was personally identifiable financial information with regard to mortgages, credit cards and private banking (Jeng, 2015). Furthermore, the hackers acquired lists of programs and applications running on JP's devices (Silver-Greenberg, Goldstein, & Perlroth, 2014).

Although JP spends a lot of resources on IT security and the hackers showed quite some level of sophistication (Jeng, 2015), this data breach could have been prevented if JP did not forget or neglect to install a relatively simple security measure on one of its servers (Goldstein, Perlroth, & Corkery, 2014; Kurane, 2014). If the overlooked server was upgraded with multifactor authentication the stolen login credentials would have been useless (Jeng, 2015). Furthermore, Jeng (2015) showed that had certain security measures been in place, that there was a high probability of malware being detected and stopped, unnecessary access of the compromised employee to data would not have been available, log files would not have been deleted, and more.

Regarding the above, we assess that the root cause for this data breach was a security misconfiguration.

Parameters	Values for JP Morgan Chase
Data Breach Type	Unauthorized Access
Breach Method	Hacking Malicious Software Social Poor or No Security
Breach Actor	Malicious Outsider

Organization Sector	Financial and Insurance
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Database
Data Responsibility	Data Controller

Data Subject	Customers Breachee
Data Amount (Records)	83.000.000
Data Type	PII PIFI OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	External Non-Auto Intended
Breach Period (Days)	60
Time between Breach and Detection	Medium
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 27 Summarized overview of identified values for JP Morgan Chase

5.2.17. The Home Depot (2014)

"The Home Depot is the world's largest home improvement specialty retailer" (The Home Depot, 2014). In 2014, The Home Depot (THD) had 2,266 retail stores mainly across Northern America and employed more than 300 thousand people (The Home Depot, 2014).

On September 2, 2014, THD received reports from banks and law enforcement indicating that THD's systems may had been breached by criminals, after which THD immediately started its investigation (The Home Depot, 2014). This followed after multiple financial institutions could trace back a batch of stolen payment cards, which were offered for sale in an online underground web store, to have been used in THD stores (Krebs, 2014b). On September 8, 2014, THD confirmed that it had suffered a data breach incident (Hawkins, 2015).

Data related to approximately 56 million payment cards had been stolen, as well as 53 million email addresses (Hawkins, 2015; Seals, 2014). Based on the payment card data offered for sale, the compromise included customers' names, card numbers and expiration date (Ragan, 2014a). These data could probably be categorized based on location, since the data also holds exactly which stores are involved, so some location information can be derived (Ragan, 2014a). The data is related to customers of THD which made payments in THD's stores between April and September 2014 (Krebs, 2014c).

The attackers managed to obtain login credentials from one of THD's third party vendors, which they used to gain access that vendor's environment within THD's perimeter (Hawkins, 2015; Seals, 2014). At that point they did not yet having direct access to THD's point-of-sale (POS) devices (Seals, 2014) Hawkins (2015) reports about what followed: *"Then they exploited a zero-day vulnerability in Windows, which allowed them to pivot from the vendor-specific environment to the Home Depot corporate environment."*

Seals (2014) reports: *". . . [T]he hackers then acquired elevated rights that allowed them to navigate portions of Home Depot's network and to deploy unique, custom-built malware on its self-checkout systems in the US and Canada."* Hawkins (2015) reports the following on the same: *"Once they were in the Home Depot network, they were able install memory scraping malware on over 7,500 self-checkout POS terminals."* These self-checkout lanes were in approximately 1800 of THD's retail stores located in the US and Canada (Krebs, 2014d).

Although having an endpoint antivirus solution installed on the POS devices, an important feature which would act as a host intrusion prevention system, was not turned on (Hawkins, 2015). Furthermore, Point-to-Point (P2P) encryption was missing, which would have encrypted the payment card data from the moment of payment contact (mostly by swiping) and would have allowed the data to reach the memory already being encrypted (Hawkins, 2015). In order to implement P2P technology, the POS devices' operating system needed to be upgraded first (Hawkins, 2015). Next to this, the POS devices functioned on an outdated, vulnerable version from Windows anyway, which should have been upgraded to a newer version (Hawkins, 2015).

Also, THD should have done a better job in segregating the POS network from the rest of its corporate network, by restricting access between these network parts and by implementing a restricted Virtualized Local Area Network (Hawkins, 2015). Furthermore, since the attackers gained first access by stealing third-party vendor's credentials and from there were able to access other parts of THD's network, this indicates the access management was not properly done and least privilege access should have been implemented (Hawkins, 2015).

These and more host-based and network-based measures to prevent and detect the data breach by addressing misconfigurations have been treated by Hawkins (2015), and he has shown that the measures are readily available. Furthermore, Seals (2017) reports: "*Security professionals agree that most certainly [this data breach was] a result of poor IT practices within the company."*

Because of these findings, we consider the root cause for this data breach to be a security misconfiguration. Especially since more similar retail data breaches occurred relatively frequently in a relatively short period of time before THD became a victim itself, which should have made THD more alert (Hawkins, 2015).

Parameters	Values for The Home Depot
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Hacking Malicious Software
Breach Actor	Malicious Outsider
Organization Sector	Retail
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Other
Data Responsibility	Data Controller
Data Subject	Customers
Data Amount (Records)	56.000.000
Data Type	PII PIFI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary
Data Breach Detection	External Non-Auto Intended
Breach Period (Days)	150
Time between Breach and Detection	Long
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 28 Summarized overview of identified values for The Home Depot

5.2.18. Indiana University (2013)

On February 25, 2014, Indiana University (IU) notified the Indiana attorney general's office of an incident which had left exposed the personal data for some students and ex-students (Indiana University USSS, 2014). On the same day Indiana University issued a press release, which stated the exposed data included names, birth dates, addresses, Social Security Numbers, and enrolment and degree information for approximately 145,966 students who had attended the university in the period from 2011 to 2014, and a small number of students that graduated before 2011 (Indiana University USSS, 2014).

A university staff member who wanted to access the files with the data for internal use discovered the issue (Indiana University USSS, 2014). On February 21, 2014, the university was able to determine that adjustments in the security settings of the website that contained the data, inadvertently allowed unauthenticated access (Indiana University USSS, 2014). These adjustments were made in March 2013, which means the data were insecurely stored for eleven months (Indiana University USSS, 2014). Upon discovery, IU locked down the website and made sure to properly protect the data by moving it to a secure location (Indiana University USSS, 2014).

IU mentions that the files were safeguarded to hide the nature of the contained data (Indiana University USSS, 2014). This was done by using file names and extensions that are meaningless to external parties, which is part of standard security practices (Fater, 2014). But the data itself was not encrypted (Roman, 2014a).

Based on the review of access logs, the involved data had only been downloaded by three web crawlers, which are automated data mining applications used to enhance Internet search capabilities by indexing and copying the websites they visit (Fater, 2014; Indiana University USSS, 2014). Hence, the data probably had been cached by the search engine (Fater, 2014). Apart from this, the data was not accessed or downloaded according to IU, and found no other evidence that the data had been viewed or used (Indiana University USSS, 2014). IU cannot guarantee that the data has not been viewed or taken (Fater, 2014), and according to our definition this is still a data breach.

An IU representative admitted this was the result of human error (Fater, 2014). Furthermore, Jimenez-Gomez (2017) assessed the compromise was the result of poor security. Based on our own analysis, we agree with these assessments and therefore consider this case to have a security misconfiguration as root cause.

Parameters	Values for Indiana University
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Poor or No Security
Breach Actor	Inadvertent Insider

Organization Sector	Education
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Web Server
Data Responsibility	Data Controller

Data Subject	Members
Data Amount (Records)	145.966
Data Type	PII
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	Internal Non-Auto Inadvertent
Breach Period (Days)	330
Time between Breach and Detection	Long
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 29 Summarized overview of identified values for Indiana University

5.2.19. British Airways (2018)

The British airline company British Airways (BA) is part of the International Airlines Group, and in 2017 BA operated a fleet of almost 300 passenger aircrafts flying worldwide (Bosnell, 2018). BA had more than 45 million passengers in 2017 and more than 46 million passengers in 2018 (Mazareanu, 2020). In the beginning of 2020 BA still had 42 thousand staff members, although it was announced in April of the same year that it would be reduced by cutting up to 12 thousands jobs (BBC News, 2020).

In September and October of 2018 it became clear that BA had been breached, after being notified on September 5 by a partner which monitored their websites, and subsequently having investigated the incident (Bosnell, 2018). The breach eventually affected 429,000 payment cards (Bosnell, 2018). The names, email addresses, billing addresses, and payment card information including card number, expiry date and Card Verification Value (CVV) of 321,000 payment card holders were affected. The other 108,000 card holders did not have their CVV compromised, but for the rest were affected in the same way (Schwartz, 2018a). The stolen data was found by security researchers being offered in an online underground shop (Klijnsma, Kremez, & Herman, 2018; Muncaster, 2018c).

BA announced that 244,000 of the compromised card holders did a transaction between August 21 and September 5 of 2018, while the other 185,000 made their transactions between April 21 and July 28 of the same year (Schwartz, 2018a). These transactions were made by customers who purchased or changed their tickets, or made another kind of payment, using BA's website or BA's mobile app (Bosnell, 2018; Schwartz, 2018a).

The attack was attributed by information security researchers to an organization named Magecart, which was already known for stealing from numerous organizations the personally identifiable (financial) information of their customers (Klijnsma, 2018; Kolesnikov & Parashar, 2018). The group targets websites with an insecure payment infrastructure (Perhar, 2018). One cybersecurity firm compared and analysed how the scripts running on BA's website changed over a period of time before and after the breach notification, which it was able to do because of the crawls it performed daily involving a huge amount of web pages (Bosnell, 2018; Klijnsma, 2018).

Magecart apparently was able to compromise BA's websites and webservers, and then modified one of the JavaScript files, which supported the payment form functionality, to include a logging script which would then grab the desired data from customers making a payment on the payment web page (Bosnell, 2018; Klijnsma, 2018; Kolesnikov & Parashar, 2018). The affected JavaScript file concerned a Modernizr JavaScript library (version 2.6.2.) which was hosted on BA's own servers (Kolesnikov & Parashar, 2018; Schwartz, 2018a).

The grabbed data subsequently got sent to a server owned by Magecart in Romania, for which the domain name 'baways.com' and a paid-for SSL certificate were arranged in order to appear legitimate (Bosnell, 2018; Klijnsma, 2018; Kolesnikov & Parashar, 2018). Notably, the certificate was issued some time before the date which BA had reported as the start date of the breach, which makes it plausible that Magecart already had access to BA's website quite some time before (Klijnsma, 2018).

All this was done while making sure the intended original functionality was maintained in order to avoid detection (Bosnell, 2018). Basically, this method is the digital variant of card skimming, and by modifying the JavaScript a virtual card skimming device was implemented (Bosnell, 2018). Once the customers hit the button to submit their payment, the entered personal and payment data were extracted (i.e. copied) from specific form fields and sent to Magecart's server (Klijnsma, 2018).

When payments were made through the BA mobile app, it loaded a web page with the same components as the BA website, which included the malicious script and thus also affected payments done through the mobile app (Bosnell, 2018; Klijnsma, 2018; Kolesnikov & Parashar, 2018).

Certainly Magecart executed an attack with a certain level of sophistication by carefully constructing a targeted attack which was customized especially to operate seamlessly with BA's infrastructure while remaining unnoticed (Klijnsma, 2018). However, some configuration vulnerabilities were also present. First, although BA did not state to which extent Magecart gained access to their network, but this access had to be significant if Magecart was able to modify a site resource (Klijnsma, 2018). The malicious script was injected into a baggage claim information page that was poorly secured and had not been modified for a significant amount of time (Perhar, 2018), and loaded from there during the transaction (Klijnsma, 2018).

Second, after the data breach was announced, BA's payment web page was checked to find out which JavaScript files were loaded (Greenwood, 2018; Reeve, 2018). It turned out that files from seven external domains apart from www.britishairways.com were being loaded, which should not be on web pages that process payment card data (Greenwood, 2018; Reeve, 2018). According to the Payment Card Industry Data Security Standards (PCI DSS) only files necessary for the payment processing should be loaded in web pages that receive payment card data (Reeve, 2018).

It was Greenwood (2018) who investigated this and he reported: *"We can see files loaded from 7 external domains (excluding www.britishairways.com). These include files from analytics, customer service and A/B testing tools. These should not be present on web pages processing customer card data."*

Based on this finding by Greenwood (2018), Reeve (2018) stated: *"It would be a violation of Payment Card Industry Data Security Standards (PCI DSS) which says only files necessary to the processing of payments should be loaded into pages that take credit card data."*

Furthermore, payment card data should have been isolated and captured in a PCI DSS compliant manner, which could have been done with a secured inline frame (iframe) (Greenwood, 2018; Perhar, 2018; Reeve, 2018). If a secured iframe had been used in which payment card data would have been collected, this data could not have been stolen by the malicious script, since JavaScripts loaded in the main window cannot access the contents of an iframe (Greenwood, 2018; Perhar, 2018).

Without these measures an unnecessary JavaScript file could be loaded, including the malicious script which could then read and copy the data from the payment form fields on the payment web page, could process them, and could send it to Magecart's servers hosted on baways.com (Klijnsma, 2018; Reeve, 2018; Schwartz, 2018c).

The Office (ICO) Information Commissioner's Office (2019) has stated the following in a news post: *"Following an extensive investigation the ICO has issued a notice of its intention to fine British Airways £183.39M for infringements of the General Data Protection Regulation (GDPR)."* This happened after an extensive investigation by the ICO, which found that data was compromised because of poor security arrangements (Information Commissioner's Office, 2019). This corresponds with the previous findings with regard to the incorrect or missing configuration. Based on our assessment, we consider this case to have security misconfigurations as the root cause.

Parameters	Values for British Airways
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Hacking Malicious Software Poor or No Security
Breach Actor	Malicious Outsider
Organization Sector	Retail
Organization Size	Large
Organization Headquarters	GB-GBR-826
Organization Component	Website Web Server
Data Responsibility	Data Controller
Data Subject	Customers
Data Amount (Records)	429.000
Data Type	PII PIFI
Data Sensitivity	Internal
Data State	DIU
Data Necessity	Necessary
Data Breach Detection	Internal Automated Intended
Breach Period (Days)	115
Time between Breach and Detection	Long
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 30 Summarized overview of identified values for British Airways

5.2.20. Cathay Pacific (2018)

At the end of 2018, the Hong Kongese airline company Cathay Pacific (CP) operated a fleet of 212 aircrafts, together with its subsidiaries Cathay Dragon (CD) and Air Hong Kong, flying to 109 destinations in 35 countries (Cathay Pacific, 2019). CP and CD carried 35.5 million passengers in 2018 (Cathay Pacific, 2019). Mid 2018, CP and CD employed respectively more than 32.4 thousand and 3.4 thousand people (Cathay Pacific, 2018b).

On October 24, 2018, the Hong Kongese airline company CP announced that it had discovered unauthorized access to some information systems in its network (Cathay Pacific, 2018a; Fu, 2018). This discovery was made after CP detected suspicious activity in the form of a brute force attack on March 13, 2018, which resulted in about five hundred staff members being locked out of their user accounts (S. K.-y. Wong, 2019). CP stated it was still being attacked in the period from March and May of 2018 (Cathay Pacific, 2018c).The data breach eventually covered a period from October 2014 to May 2018 (Venkat, 2020). After becoming aware of the data breach, CP immediately started blocking the attackers and started remediation activities (S. K.-y. Wong, 2019).

Immediate investigation followed and eventually showed that four of CP's systems in their vast network were affected (S. K.-y. Wong, 2019). Parsons, Lin, and Phillips (2019) summarized the findings of S. K.-y. Wong (2019) as follows: *"(i) the customer loyalty system, (ii) a shared back-end database used to support web-based applications, (iii) a reporting tool that extracted and complied data from other databases; and (iv) a database used to allow customers to redeem non-air rewards . . ."* From CP's customer loyalty system, one database was attacked (S. K.-y. Wong, 2019).

This affected approximately 9.4 million CP passengers, as well as passengers of its subsidiary CD (Fu, 2018). The amount and type of compromised personal data varied for each affected passenger, depending on what data was taken from which system (Cathay Pacific, 2018a). For each affected passenger the name was compromised (S. K.-y. Wong, 2019). For more than half of the affected passengers, but in varying amounts, the compromised data consisted of *"flight numbers and dates, titles and email addresses"* (Parsons et al., 2019). Around a third of the affected passengers had their membership number compromised, a quarter their home address, and one fifth their phone number (S. K.-y. Wong, 2019).

The nationality of twelve percent of the affected passengers was compromised, and for nine percent their passport number, and for eight percent their date of birth (S. K.-y. Wong, 2019). A very small amount (and percentage) had their (mostly expired) credit card number compromised (S. K.-y. Wong, 2019). Notably, six percent of the affected passengers had their identity card number compromised, of which around 240 thousand should already had been removed from CP's possession anyway (S. K.-y. Wong, 2019). Historical travel information was also involved (Cathay Pacific, 2018a).

Three main incidents could be identified (S. K.-y. Wong, 2019):

- a keylogger attack in October 2014, of which it could not be found out how the intruders accessed and placed the malicious software. This enabled the attackers to obtain valid user account login credentials and from there to navigate further through the network and place other hacking tools to further obtain domain credentials. The attackers were able to access personal data contained in this part of the network they could move through.
- suspicious activity in an Internet facing server which could be exploited because of a vulnerability allowing unauthenticated and unauthorized administrative access. This vulnerability was already publicly known for a long time.
- the earlier mentioned brute force attack. CP stated it was unable to determine of this attack resulted in data being compromised.

In his investigation report, Hong Kong's Privacy Commissioner found that CP is responsible for certain missing, incomplete and incorrect security measures (S. K.-y. Wong, 2019). As S. K.-y. Wong (2019) states: *"Cathay had not taken reasonably practicable steps not to expose the administrator console port of the Internet Facing Server to the Internet, as a result of which a gateway for attackers was opened."*

As such, *"Cathay Pacific's administrator console had been configured to be accessible externally rather than limited to internal network access, and this was found to be deficient"* (Parsons et al., 2019). This old publicly known vulnerability should have been found and patched by that time (Duckett, 2019; S. K.-y. Wong, 2019).

Furthermore, except CP's IT Support Teams no one was required to use multifactor authentication, which should have been implemented at least for remote access users trying to access internal systems (S. K.-y. Wong, 2019). This did not offer proper protection from unauthorized access, and the unauthorized access is the cause of this data breach incident according to CP (S. K.-y. Wong, 2019).

Also, database backup files that were used to facilitate database migration in the period 2016 to 2018 were not encrypted (S. K.-y. Wong, 2019). Another issue has to do with executing vulnerability scans too infrequently and using scan software incapable of finding a well-known vulnerability.

The United Kingdom Information Commissioner's Office (ICO) also investigated this data breach, since a significant number of British citizens were among the affected customers, and in relation to security errors came to the same findings as Hong Kong's Privacy Commissioner. Because of this ICO imposed a fine on CP (Venkat, 2020).

Following our findings and assessment, we consider this case to have security misconfigurations as root cause.

Parameters	Values for Cathay Pacific
Data Breach Type	Unauthorized Access
Breach Method	Hacking Malicious Software
Breach Actor	Malicious Outsider

Organization Sector	Retail
Organization Size	Large
Organization Headquarters	HK-HKG-344
Organization Component	Database Other
Data Responsibility	Data Controller

Data Subject	Customers
Data Amount (Records)	9.400.000
Data Type	PII PIFI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Mixed

Data Breach Detection	Internal Non-Auto Inadvertent
Breach Period (Days)	1309
Time between Breach and Detection	Long
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 31 Summarized overview of identified values for Cathay Pacific

5.2.21. SingHealth (2018)

Singapore Health Services (SingHealth) is Singapore's largest group of healthcare institutions, with several hospitals, national specialty centres and polyclinics (Tham, 2018). In 2018 it came to light that attackers had stolen personal information for 1.5 million patients, of which around 159 thousand also had their outpatient prescriptions stolen, from SingHealth's Sunrise Clinical Manager (SCM) patient database (Committee of Inquiry, 2019; Tham, 2018). These concerned patients who had visited SingHealth's institutions between May 1, 2015, and July 4, 2018 (Yu, 2018b).

The data involved consisted of their *"names, [National Registration Identity Card] numbers, addresses, genders, races, and dates of birth"* (Committee of Inquiry, 2019). Approximately 159 thousand of the 1.5 million patients also had their outpatient dispensed medication data stolen, including the nation's Prime Minister (Committee of Inquiry, 2019).

Investigation has shown the whole attack was stretched out over the period from August 23, 2017, and July 20, 2018, while the actual data exfiltration took place between June 27 and July 4 of 2018 (Committee of Inquiry, 2019). Around August 23, the attacker, likely having used phishing to infect front-end workstations, gained access to SingHealth's network (Committee of Inquiry, 2019). What happened next, is described by the Committee of Inquiry (2019) as follows: *"The attacker then lay dormant for several months, before commencing lateral movement in the network between December 2017 and June 2018, compromising a number of endpoints and servers, including the Citrix servers located in [Singapore General Hospital], which were connected to the SCM database."* From there the attacker tried several times to access the SCM database, by remotely connecting to the Singapore General Hospital (SGH) Citrix Servers first.

In Singapore, public healthcare has an IT provider operating under the name 'Integrated Health Information Systems', which is abbreviated to IHiS (Committee of Inquiry, 2019). In June 2018 IHiS noticed unauthorised logins to the Citrix servers and failed login attempts to the SCM database, but did not act on it immediately (Committee of Inquiry, 2019). In the meantime the attacker obtained login credentials for the SCM database on June 26, 2018 (Committee of Inquiry, 2019). A day later the attacker started querying the database and exfiltrating patient data, which remained undetected by the IHiS (Committee of Inquiry, 2019).

On July 4, 2018, unusual activity on SingHealth's database was detected by IHiS database administrators (Committee of Inquiry, 2019; Tham, 2018). Measures were put in place to prevent further intrusions (Committee of Inquiry, 2019). About this, Tham (2018) reports: *"Security measures, including the blocking of dubious connections and changing of passwords, were taken to thwart the hackers."*

After July 4, 2018, the attacker could not make successful queries to the SCM database any longer, and therefore was also unable to exfiltrate patient data from that moment (Committee of Inquiry, 2019). The breach disclosure came almost a week later, as Yu (2018a) reports: *"Six days later, on July 10, IHiS informed the Health Ministry and Cyber Security Agency of Singapore (CSA) it had suffered a cyberattack."*

Although labelled as a well-planned and sophisticated attack, the investigation by the Committee of Inquiry (2019) brought forward this finding: *"There were a number of vulnerabilities, weaknesses, and misconfigurations in the SingHealth network and SCM system that contributed to the attacker's success in obtaining and exfiltrating the data, many of which could have been remedied before the attack."*

Furthermore, the Committee of Inquiry (2019) reports: *"A significant vulnerability was the network connectivity (referred to in these proceedings as an "open network connection") between the SGH Citrix servers and the SCM database, which the attacker exploited to make queries to the database. The network connectivity was maintained for the use of administrative tools and custom applications, but there was no necessity to do so."*

In addition, the Committee of Inquiry (2019) reports: *"The SGH Citrix servers were not adequately secured against unauthorised access. Notably, the process requiring 2-factor authentication ("2FA") for administrator*

access was not enforced as the exclusive means of logging in as an administrator. This allowed the attacker to access the server through other routes that did not require 2FA."

On top of these, there was another vulnerability, as the Committee of Inquiry (2019) reports: *"There was a coding vulnerability in the SCM application which was likely exploited by the attacker to obtain credentials for accessing the SCM database."*

Key findings of the investigation were that the misconfigurations contributed to the successful attack and that this successful attack was not inevitable (Committee of Inquiry, 2019). The misconfigurations allowed the attacker to succeed in access the network and stealing the data (Davis, 2019; Yu, 2019). Remarkably, part of the vulnerabilities were already found earlier, like in 2014 by an employee (Yu, 2018b), and during a penetration test in 2017 (Davis, 2019). These vulnerabilities were not remediated yet because of mismanagement, while the remedies should have been standard practices anyway for an organization such as SingHealth (Davis, 2019).

Hence, even if the attack is considered sophisticated, would the misconfigurations not have been present, the attack would very likely not have succeeded, and the data breach would then not have occurred. Therefore, we assess this case to have a security misconfiguration as a root cause.

Parameters	Values for SingHealth
Data Breach Type	Unauthorized Access
Breach Method	Hacking Malicious Software
Breach Actor	Malicious Outsider
Organization Sector	Medical and Healthcare
Organization Size	Large
Organization Headquarters	SG-SGP-702
Organization Component	Database
Data Responsibility	Data Controller
Data Subject	Patients
Data Amount (Records)	1.500.000
Data Type	PII PHI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary
Data Breach Detection	Internal Automated Intended
Breach Period (Days)	8
Time between Breach and Detection	Short
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 32 Summarized overview of identified values for SingHealth

5.2.22. Orbitz (2018)

On its website, Orbitz offers a travel search engine with which travel fares can be aggregated in order for customers to search and book different kinds of travel related facilities and activities (Kovacs, 2018b; Schwartz, 2018b). In September 2015 Orbitz was acquired by Expedia, another online travel agency (Kovacs, 2018b; Schwartz, 2018b).

In 2017, Orbitz was notified that a portal on their legacy platform possibly was a common point of purchase in connection with fraudulent transactions (Coble, 2019). On March 20, 2018, Orbitz announced that while investigating a legacy Orbitz travel booking platform, it had found evidence on March 1 which suggested unauthorized access to personal information stored on that platform (Orbitz, 2018). Orbitz (2018) communicated the following: *". . . Orbitz determined on March 1, 2018 that there was evidence suggesting that, between October 1, 2017 and December 22, 2017, an attacker may have accessed certain personal information . . ."*

The compromised data originated from submissions done for purchases made by Orbitz's customers between January 1 and June 22 of 2016, and from submissions done for purchases made by Orbitz's partners' customers in the period from January 1, 2016 until December 22, 2017 (Orbitz, 2018). Orbitz (2018) stated the following with regard to the compromised data: *"On March 1, 2018, Orbitz determined that the personal information that was likely accessed may have included full name, payment card information, date of birth, phone number, email address, physical and/or billing address, and gender."* This concerns the personal information of 880 thousand customers (Schwartz, 2018b; Seals, 2018b).

After the acquisition, the compromised platform still functioned as both a consumer and business partner platform on which data was stored, and still contained data in the period when the breach took place (Orbitz, 2018). It apparently served as the underlying travel search and booking engine for many travel websites, and was still being used for this partner platform purpose when the breach took place (Peters, 2018). As earlier mentioned, the compromised data of Orbitz's customers were from the period January 1 and June 22 of 2016, hence the platform probably was not being used as a direct consumer platform any longer at the moment of breach.

After determining the data breach, Orbitz immediately eliminated the unauthorized access and took action to prevent any future unauthorized access to the compromised platform (Orbitz, 2018). Furthermore, Orbitz stated it would go on to investigate the incident (Orbitz, 2018), but did not share any further information with regard to how the data breach took place (Spring, 2018a).

However, the Pennsylvania Attorney General's Office also investigated this data breach, since the personal data of Pennsylvanian citizens were possibly involved (Pennsylvania Office of Attorney General, 2019). This investigation found that an actor realized unauthorized access and remained undetected, allowing the actor to implement malicious software intended to target payment card data (Coble, 2019; Pennsylvania Office of Attorney General, 2019).

Based on the investigation, the Pennsylvania Office of Attorney General (2019) stated: *"In addition, multiple Payment Card Industry Data Security Standards requirements were not in place at the time of the breach."* According to the investigation, Orbitz's security failed (Pennsylvania Office of Attorney General, 2019). The results of the investigation and the ensuing settlement showed Orbitz's liability (Pennsylvania Office of Attorney General, 2019).

The inadequate security may well have been the result of failing to integrate and update the acquired infrastructure, including legacy systems, by Expedia (Abbott & Goegan, 2018; Seals, 2018b). Furthermore, the legacy systems probably introduced risks since they often were neglected and remained without patches and updates, and often remained unmonitored (Abbott & Goegan, 2018). This made legacy systems an attractive, easy target for attackers (Abbott & Goegan, 2018).

Based on our findings, we consider this case very likely to be caused by security misconfiguration.

Parameters	Values for Orbitz
Data Breach Type	Unauthorized Access
Breach Method	Malicious Software Poor or No Security
Breach Actor	Malicious Outsider
Organization Sector	Retail
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Web Server
Data Responsibility	Data Controller
Data Subject	Customers
Data Amount (Records)	880.000
Data Type	PII PIFI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary
Data Breach Detection	Internal Non-Auto Intended
Breach Period (Days)	83
Time between Breach and Detection	Long
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 33 Summarized overview of identified values for Orbitz

5.2.23. Securus Technologies (2015)

Securus Technologies (ST) is a telecommunications company in the United States of America, with its core business being to provide phone and video visitation services to correctional institutions for paid use by people in those institutions, including the monitoring and recording of phone and video calls (J. Smith & Lee, 2015). ST's clients are the operators of prison systems, mainly local and county governments and state departments of correction (J. Smith & Lee, 2015). ST employs well over 1000 people that contribute in serving more than 3,400 governmental agencies and more than 1.2 million inmates across North America (Securus Technologies, 2020).

In November 2015, online news magazine 'The Intercept' reported that it had received a file with 37 gigabytes of data, which indicated a data breach at ST (J. Smith & Lee, 2015). The data was leaked to The Intercept through SecureDrop, a platform over which journalists and sources (especially whistle blowers) can communicate securely and sources can stay anonymous (Ball, 2014; The Intercept, n.d.).

Initially, the data in the file was spread over a large amount of tables, which The Intercept merged into one table (J. Smith & Lee, 2015). After removing the duplicates, the file contained 70 million records of unique phone calls, including metadata of the incarcerated caller, namely their first and last names, the phone numbers they called to, the date and time the time at which the calls took place, the duration of the calls and the inmates' ST account numbers (J. Smith & Lee, 2015). The 70 million phone calls were made by more than 63 thousand inmates to almost 1.3 million unique phone numbers, in the period from December 2011 until spring 2014 (J. Smith & Lee, 2015). Each phone call record included a URL link where the actual audio recording could be downloaded (J. Smith & Lee, 2015). After the incident, ST moved the audio recordings such that they could not be reached through the breached URL links any longer (J. Smith & Lee, 2016).

According to The Intercept this breach was done by a hacktivist who believed ST was violating incarcerated people's constitutional rights (J. Smith & Lee, 2015). Notably, at least 57 thousand phone call records concern recorded conversations between inmates and attorneys (J. Smith & Lee, 2015, 2016). These will, at least partly, include confidential and privileged legal communications which should have never been recorded and never kept stored, because of attorney-privilege (J. Smith & Lee, 2015).

In a press release, ST stated that the data was taken from one of ST's customers' data files, which were probably accessed through that customer's file sharing arrangement with a third-party vendor, suggesting it was a criminal attack that happened outside ST's network (Securus Technologies, 2015). ST did not release more information with regard to this breach (J. Smith & Lee, 2016). If The Intercept was informed by the hacktivist about how the data breach took place, it did not disclose this information. And we could not find any other additional information or more detailed descriptions of this data breach incident. Therefore, we cannot determine how the hacktivist gained access and eventually succeeded in exfiltrating the data. Indeed, we cannot assess whether this case had a security misconfiguration as a root cause, or was caused by a complex, sophisticated attack.

Parameters	Values for Securus Technologies
Data Breach Type	Unauthorized Disclosure
Breach Method	Inconclusive
Breach Actor	Ethical Outsider

Organization Sector	Technology
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Cloud Storage
Data Responsibility	Data Controller

Data Subject	Citizens
Data Amount (Records)	70.000.000
Data Type	PII CI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Mixed

Data Breach Detection	External Non-Auto Intended
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Inconclusive

Table 34 Summarized overview of identified values for Securus Technologies

5.2.24. National Revenue Agency of Bulgaria (2019)

Bulgaria's National Revenue Agency (NRA) is a specialized state body, falling under the Minister of Finance, with its main task being collecting and administering taxes and social security contributions, as well as other state receivables (National Revenue Agency, n.d. (a), n.d. (b)). NRA has numerous offices across Bulgaria with more than ten thousand employees (National Revenue Agency, n.d. (a), n.d. (c)).

On July 15, 2019, several media organizations received an anonymous email revealing a data breach at NRA, including a link to part of the stolen data as proof (Monitor Agency, 2019). The email was sent from a Russian email address (Cimpanu, 2019e; Krasimirov & Tsolova, 2019). The actor probably acted as a hacktivist and not with malicious intent, wanting to denounce the corruption within the Bulgarian government and the poor IT security practices that are a result of it (P. Baker, 2019; Monitor Agency, 2019; O'Neill, 2017). The actor claimed to have had access to NRA's network for more than eleven years, and a compromise of 30 gigabytes of data in 2012 remained unnoticed (Cimpanu, 2019e; Denizova, 2019). In the email it was also claimed the total amount of stolen data was 21 GB in 110 compromised folders, but the recipients of this email only got to see part of that (Cimpanu, 2019e; Monitor Agency, 2019; Stoyanov, 2019a).

This revealed part amounted to 11 gigabytes of data in the form of .csv files distributed among 57 folders, i.e., databases (Cimpanu, 2019e; Monitor Agency, 2019; Stoyanov, 2019a). About the compromised data, P. Baker (2019) states: *"Reports say that the data included the names, addresses and social security information of up to five million individuals. Further reports suggest that the data included details of taxable income, loans, health insurance payments etc."*

The news website news.bg (2019) adds that part of the trove also contained payments for pension insurance. Stoyanov (2019a) reports that there are files with *"registered users of online betting sites"*. Company related data was also part of the compromised data (Denizova, 2019; Monitor Agency, 2019; Stoyanov, 2019a). The leaked data contained more than five million data records, and affected Bulgarian and foreign citizens, as well as companies (Krasimirov & Tsolova, 2019).

Stoyanov (2019b) reports: *"The next day, more or less, the NRA also confirmed both the scale and the authenticity of the data."* According to Denizova (2019) the NRA did so after examining approximately thirty percent of the compromised data. Furthermore, the NRA admitted to having had its systems breached (Stoyanov, 2019a).

The NRA noticed abnormal activities already on June 29, 2019 (Denizova, 2019; Stoyanov, 2019b). Denizova (2019) reports: *". . . happened on June 29, when an attempt was made to gain unauthorized access to one of their servers."* Hence, NRA assumed the data was taken on that day (Stoyanov, 2019b). The NRA did not disclose this, rather this data breach was brought to public attention by the media (P. Baker, 2019).

The NRA stated that the unauthorized access was gained through a rarely used online service (Denizova, 2019; Stoyanov, 2019b). Stoyanov (2019b) reports the following on that: *"The access to the agency's server took place through one of their electronic services, which is for [Value Added Tax] refund from transactions abroad (VATrefund). It was established in 2012, benefited from a relatively small number of taxpayers and has not been renovated, making it vulnerable."* This vulnerability was eliminated shortly after the data breach became known (Denizova, 2019).

According to cybersecurity it was very likely the security was poor instead of a very sophisticated attacker being the case (O'Neill, 2017). Apparently, the attacker left various digital traces and a suspect was arrested shortly after (O'Neill, 2017).

By deploying a Structured Query Language (SQL) injection, the intruder succeeded in reaching a database in one of NRA's servers and randomly grabbed data (O'Neill, 2017; Stoyanov, 2019a). Protection against SQL injection attacks is possible and available, and should have been done correctly which would have prevented this data breach (O'Neill, 2017). The scale and scope of the breached data suggested the

operational security was not according to standards (O'Neill, 2017). Based on our findings, we consider this data breach to be caused by a security misconfiguration.

Parameters	Values for the National Revenue Agency of Bulgaria
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider
Organization Sector	Government and Military
Organization Size	Large
Organization Headquarters	BG-BGR-100
Organization Component	Database
Data Responsibility	Data Controller
Data Subject	Citizens Breachee Other
Data Amount (Records)	5.000.000
Data Type	PII PIFI OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary
Data Breach Detection	External Non-Auto Intended
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 35 Summarized overview of identified values for the National Revenue Agency of Bulgaria

5.2.25. Chtrbox (2019)

Chtrbox is an Indian marketing company based in Mumbai which connects brands with influencers to launch marketing campaigns (Chtrbox, n.d. (a), n.d. (b); Crunchbase, n.d.-e). Organizations that want to collaborate with influencers in India are Chtrbox's clients, and Chtrbox pays influencers to post promotional content on their social media accounts (Chtrbox, n.d. (a), n.d. (b); Crunchbase, n.d.-e; Kirk, 2019a; Whittaker, 2019). Chtrbox has less than 50 employees (Crunchbase, n.d.-e; Kirk, 2019a).

On May 20, 2019, a security researcher discovered a database holding sensitive information, by using the Shodan search engine, which can be used to find exposed devices and databases (Whittaker, 2019). The database was hosted by Amazon Web Services and this bucket was left unsecured without password protection, allowing unauthenticated and unauthorized access (Cluley, 2019; Whittaker, 2019). At the moment of reporting, the database was holding more than 49 million records, but this was growing (Whittaker, 2019).

The reviewed records all contained public data scraped from Instagram accounts, belonging to influencers, celebrities and brands (Whittaker, 2019). The data included their biographies, profile pictures, the number of followers, account verification status, and location information (Whittaker, 2019). In some cases there was also included their personal contact information, namely the account owners' provided email addresses and phone numbers (Whittaker, 2019). Each record also contained a calculated account valuation, based on the number of followers, engagement, likes and shares (Whittaker, 2019). This valuation served as a metric to determine the payment a influencer should receive for a sponsored post (Whittaker, 2019).

The security researcher reached out to technology news website TechCrunch in an effort to identify the owner and eventually secure the database (Whittaker, 2019). TechCrunch was able to trace back the database to Chtrbox (Whittaker, 2019). TechCrunch notified Chtrbox about this situation and shortly after the database was taken offline (Whittaker, 2019).

Chtrbox explained that they used this database only for internal purposes, in order to be able to connect brands with influencers or celebrities who will promote their product or services with quality content, in a way that offers influencers the opportunity to monetize their online social status (Cluley, 2019; Kirk, 2019b). Chtrbox stated that the database contained data available from the public domain or was self-reported by influencers, and named the a number of 350 thousand affected people (Kirk, 2019a; Whittaker, 2019). Also, Chtrbox stated that the database was only exposed for 72 hours (Whittaker, 2019). Instagram also stated that Chtrbox's database contained publicly available information relating to only 350 thousand people, and this information came from many sources and not only Instagram (Whittaker, 2019).

These statements could be refuted. TechCrunch randomly contacted multiple people whose information was in the database, and they all confirmed their personal information as well as confirmed the data was linked to their Instagram accounts (Kirk, 2019a). And not all the encountered data was publicly available, which remained unclear how that data was obtained by Chtrbox and on which Chtrbox did not respond (Kirk, 2019a; Mehta, 2019). Furthermore, the number of 350 thousand might have been to downplay the incident and was based on Chtrbox's network of influencers within India. But the database also contained data relating to people from outside of India (Kirk, 2019a). Additionally, the security researcher saw the database was indexed by the Shodan search engine already on May 14, 2019, hence the exposure period must have been longer than 72 hours (Kirk, 2019b). Finally, one could argue that publicly available data becomes more sensitive and more vulnerable when aggregated.

Leaving a database with sensitive personal information easily and openly accessible from the public Internet without even a basic for of security is clearly a mistake (Cluley, 2019; Robinson, 2019c), and we consider this to be a security misconfiguration. It does not matter whether it was an honest mistake or it was left open because it has to be accessed by other applications and services in a dynamic environment (Robinson, 2019c).

Parameters	Values for Chtrbox
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider

Organization Sector	Technology
Organization Size	Small
Organization Headquarters	IN-IND-356
Organization Component	Database Cloud Storage
Data Responsibility	Data Controller

Data Subject	Customers Other
Data Amount (Records)	49.000.000
Data Type	PII OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Mixed

Data Breach Detection	External Automated Intended
Breach Period (Days)	6
Time between Breach and Detection	Short
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 36 Summarized overview of identified values for Chtrbox

5.2.26. Cellebrite (2017)

The Israel-based digital intelligence company Cellebrite offers data access, analysis, transfer, decoding and processing solutions, and is probably best known for its mobile forensics solutions which support digital investigations (Cellebrite, n.d.; Crunchbase, n.d.-d). Related to mobile forensics, Cellebrite offers products and solution with which a large amount of different mobile phones can be accessed in order to eventually read and understand that data.

Cellebrite's customers overall vary from retailers to telecommunication companies to government agencies, but for its mobile forensics solutions it mainly has law enforcement agencies, military agencies, intelligence agencies as its clients, while also supporting corporate security (Crunchbase, n.d.-d). Cellebrite employs a several hundred people (Cellebrite, n.d.; Crunchbase, n.d.-d).

In January 2017, the online magazine Motherboard received 900 gigabytes of data from an anonymous person, who probably was involved in the hack (Cox, 2017a). This hacktivist showed contempt for the state of surveillance legislation and what it is becoming, and for the Western governments' attitude on that subject (Cox, 2017a; Muncaster, 2017).

After reviewing the data and verifying a small part of the data, Motherboard alerted Cellebrite on January 11, 2017 (Cellebrite, 2017b; Cox, 2017a). Thereupon, Cellebrite started its own investigation and confirmed unauthorized access to one of its external servers which was located in a secured data center (Cellebrite, 2017a, 2017b; Cox, 2017a). According to Cellebrite (2017b) this server was mainly used for storing back up data.

The majority of the data were log files from a legacy database backup of my.Cellebrite, Cellebrite's end-user licensing system, and from user activity (Cox, 2017b). The data also contained user contact details and some passwords, which were hashed (Cox, 2017b). Furthermore, the compromised data included technical support inquiries submitted in my.Cellebrite (Cox, 2017b). My.Celebrite is where customers logged in to access their accounts, where they could access products and software updates and (Cox, 2017a; Osborne, 2017). But also other data was encountered in the trove. Included was data collected from mobile devices by Cellebrite's customers during their digital investigations, as well as were included databases and technical information (Cox, 2017a, 2017b; Khandelwal, 2017b; Kovacs, 2017).

Neither the breach actor nor Cellebrite explained how the Cellebrite's network was accessed and how the data was exfiltrated. We did not find any other source explaining in more detail how the data breach took place. Therefore, we cannot determine the root cause for this case.

Parameters	Values for Cellebrite
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Hacking
Breach Actor	Ethical Outsider

Organization Sector	Technology
Organization Size	Large
Organization Headquarters	IL-ISR-376
Organization Component	Database Web Server
Data Responsibility	Data Controller

Data Subject	Customers Breachee
Data Amount (Records)	Inconclusive
Data Type	PII LI OBI CI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	External Non-Auto Intended
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Inconclusive

Table 37 Summarized overview of identified values for Cellebrite

5.2.27. Toyota Motor Corporation (2019)

The Toyota Motor Corporation (TMC) is an automotive company that operates globally, manufacturing and marketing its products in more than 170 countries and regions, and having its headquarters in Japan (Crunchbase, n.d.-i; Toyota Motor Corporation, n.d.). In 2019, TMC had more than 370 thousand employees (Toyota Motor Corporation, 2020).

On March 29, 2019, TMC announced that there had been detected unauthorized access to systems connected to its network (Toyota, 2019). This unauthorized access was detected eight days earlier and concerned the servers holding sales information for several of TMC's subsidiaries and affiliate enterprises (Goswami, 2019; Toyota, 2019). Because of this breach, sales information on up to 3.1 million customers was compromised (Goswami, 2019; Toyota, 2019). After the breach TMC stated it was still investigating whether data was actually read and taken (Cimpanu, 2019b; Gatlan, 2019; Orr, 2019). We could not find any other information confirming or denying that the data was read or exfiltrated. However, to be considered a data breach according to our definition, that is not relevant.

The notice by TMC said the compromised information included names, dates of birth and employment information (Goswami, 2019). TMC did not provide further details on this (Cimpanu, 2019b). It was the second security incident in five weeks, the first affected Toyota in Australia (Cimpanu, 2019b; Orr, 2019; Robinson, 2019b). Later on, TMC revealed that there also had been detected unauthorized access at subsidiaries in Vietnam and Thailand on March 19, but did not give further detailed information and did not elaborate on whether this could be a related incident (Cimpanu, 2019b; Kovacs, 2019b; Matthews, 2019).

TMC did not specify how the access took place or elaborate on who the suspected intruder was (Matthews, 2019; Orr, 2019). Some security experts stated a hacking organization named APT32, known for (industrial) cyber espionage, may have been involved in the breach (Cimpanu, 2019b; Gatlan, 2019; Ikeda, 2019a; Matthews, 2019; Robinson, 2019b; Winder, 2019). APT32 is known for sophisticated hacking attacks and were linked to earlier attacks targeting automotive companies (Carr, 2017; Ikeda, 2019a; Matthews, 2019). But these remained speculations as we could not find any evidence, nor was this confirmed or denied by one of the involved parties (Matthews, 2019; Winder, 2019).

TMC has provided minimal information with regard to this data breach (Matthews, 2019; Orr, 2019; Winder, 2019), and we did not find other descriptions or analyses of this data breach case which could provide us with additional insights. Therefore, we were not able to determine the root cause.

Parameters	Values for Toyoto Motor Corporation
Data Breach Type	Unauthorized Access
Breach Method	Inconclusive
Breach Actor	Malicious Outsider

Organization Sector	Technology
Organization Size	Large
Organization Headquarters	JP-JPN-392
Organization Component	Inconclusive
Data Responsibility	Data Controller

Data Subject	Customers
Data Amount (Records)	3.100.000
Data Type	PII
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	Internal Inconclusive Inconclusive
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Inconclusive

Table 38 Summarized overview of identified values for Toyota Motor Corporation

5.2.28. T-Mobile (2018)

T-Mobile US, better known as just T-Mobile, is a wireless network operator based in the United States (US) of America (T-Mobile U.S., n.d.). Currently it is one of the largest wireless carriers in the US in terms of customers (T-Mobile US, 2020). In 2018, T-Mobile US had around 52 thousand employees (O'Dea, 2020).

On August 23, 2018, T-Mobile publicly announced that a data breach had taken place, which involved customers' personal information (Franceschi-Bicchierai, 2018; Spring, 2018b). T-Mobile's cybersecurity team detected the unauthorized access on August 20, 2018, shortly after it started on the same day, and immediately stopped it (Corfield, 2018; Franceschi-Bicchierai, 2018; Kirk, 2018; Mathews, 2018; T-Mobile, 2018; Tech Advisor Staff, 2019).

The intruders managed to capture some customer data and approximately 2.2 million US customers were affected (Franceschi-Bicchierai, 2018; Kirk, 2018; Osborne, 2018b; Tech Advisor Staff, 2019). The involved compromised data included names, dates of birth, billing zip codes, phone numbers, email addresses, account number and account type (T-Mobile, 2018).

Afterwards, it became clear that also encrypted or hashed passwords were compromised, as a sample of the data showed and which was later confirmed by T-Mobile (Franceschi-Bicchierai, 2018; Kirk, 2018). Some security researchers stated that it may have been the weak hashing algorithm named MD5 (Franceschi-Bicchierai, 2018), which can be cracked (Whittaker, 2012). T-Mobile did not want to specify what encryption or hashing methods were used on the passwords (Franceschi-Bicchierai, 2018; Kirk, 2018). This is a bit remarkable to say the least, since as a breached organization you would want to reassure the data subjects, regulatory bodies, and the general public and therefore would disclose it if a very strong cryptographic method would have been used.

The attacker accessed T-Mobile's network and eventually its servers by exploiting a certain vulnerable application programming interface (API) linked to T-Mobile's website (Franceschi-Bicchierai, 2018; Spring, 2018b). A security researcher with ties to the hacker stated he was able to confirm this (Kirk, 2018). T-Mobile stated the intrusion was quickly detected and stopped (Franceschi-Bicchierai, 2018; Spring, 2018b).

In recent years, the usage of APIs has increased a lot (Kasun, 2020a). Inherently, this has increased the risk of APIs being vulnerable to attacks (Kasun, 2020a). There have been various cases of organizations becoming victims of security incidents because of insecure APIs (Kasun, 2020b). In fact, T-Mobile already was the victim of another customer data theft in October 2017, and this was because T-Mobile was unaware of a well-known bug in a Web API. When discovered, this API was quickly patched.

Based on the findings in the coverage of this data breach, especially the confession by T-Mobile that the unauthorized access happened because of an improperly secured API (Bagnulo, 2018; Spring, 2018b), we consider this data breach to have a security misconfiguration as root cause.

Parameters	Values for T-Mobile
Data Breach Type	Unauthorized Access
Breach Method	Poor or No Security
Breach Actor	Malicious Outsider

Organization Sector	Telecommunication
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Web Server
Data Responsibility	Data Controller

Data Subject	Customers
Data Amount (Records)	2.200.000
Data Type	PII OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	Internal Inconclusive Inconclusive
Breach Period (Days)	1
Time between Breach and Detection	Short
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 39 Summarized overview of identified values for T-Mobile

5.2.29. Marriott International (2018)

Marriott International, in short Marriott, is a globally operating company in the hospitality sector, mainly known for the broad portfolio of hotel brands they manage and franchise (Crunchbase, n.d.-f; Marriott International, n.d.). The company is based in the U.S. (DB CyberTech, 2019). In 2018, Marriott had approximately 176 thousand employees (Lock, 2020). In September 2016, Marriott acquired Starwood Hotels and Resorts Worldwide (DB CyberTech, 2019; Sorenson, 2019; West & Zentner, 2019). Starwood itself also consists of multiple brands (Fruhlinger, 2020).

On November 30, 2018, Marriott announced there had taken place a security incident involving Starwood's network and guest reservation database (Marriott International, 2018b). Marriott International (2018b) announced: *"On September 8, 2018, Marriott received an alert from an internal security tool regarding an attempt to access the Starwood guest reservation database in the United States."*

On September 7, 2018, an internal security tool monitored by Accenture, the company who managed part of Starwood's Information Technology (IT) infrastructure and its security before being acquired by Marriott and continued to do so afterwards for the legacy systems, gave an alert (Sorenson, 2019). This alert was given because of an unusual database query (Fruhlinger, 2020; Sorenson, 2019). Marriott was made aware of this situation on September 8, 2018 (Fruhlinger, 2020; Sorenson, 2019).

Sorenson (2019) testified: *"The . . . alert was triggered by a query from an administrator's account to return the count of rows from a table in the database . . . As part of our investigation into the alert, we learned that the individual whose credentials were used had not actually made the query."* This meant this account was taken control of by another party (Fruhlinger, 2020).

During investigation of Starwood's systems two hostile tools were discovered: a Remote Access Trojan (RAT), which is malicious software intended to offer access to and control of a system through a remote network connection, and MimiKatz, which is used for sniffing out username plus password combinations in system memory (Fruhlinger, 2020; Lambert, 2016; Sorenson, 2019). The RAT probably got downloaded from a phishing email (Fruhlinger, 2020).

DB CyberTech (2019) states: *"According to Marriott the stolen personal information was staged to a compromised internal server where the data was then encrypted. It's likely the attackers encrypted the data to obfuscate it so that Data Loss Prevention (DLP) systems could not identify the stolen information as it exited the Marriott network."* The attacker even tried to delete data and succeeded in deleting data that he had copied before (Marriott International, 2018b; Sorenson, 2019).

Marriott International (2018b) communicated: *"On November 19, 2018, Marriott was able to decrypt the information and determined that the contents were from the Starwood guest reservation database."* Marriott analysed the data and identified a maximum of approximately 383 million records of Starwood guests affected by the incident, although according to Marriott it is likely that less than 383 million unique guests were compromised (Marriott International, 2019).

Marriott International (2019) announced the following about the compromised data: *"the information included some combination of name, mailing address, phone number, email address, passport number, Starwood Preferred Guest ("SPG") account information, date of birth, gender, arrival and departure information, reservation date, communication preferences, and encrypted payment card numbers."* In an earlier announcement, Marriott International (2018a) has stated: *"For some, the information also includes payment card numbers and payment card expiration dates . . ."*

With regard to the passport numbers and payment card numbers, only parts of those were encrypted (Marriott International, 2019). The encryption of payment card numbers was done by using Advanced Encryption Standard (AES-218) (Marriott International, 2018b). The encryption method for the passport numbers was not explicitly shared.

Although the attack displayed some level of sophistication, Marriott still has itself to blame. Before acquiring Starwood, it was already known Starwood's IT security was not up to par and actually was breached before in 2015, remaining unnoticed for eight months (Fruhlinger, 2020). After acquiring Starwood, most of Starwood's staff managing IT security were let go, which is common for mergers because of financial reasons (Fruhlinger, 2020). Starwood's reservation infrastructure was not directly migrated into Marriott's, leaving Marriot's reservation system unable to book guests for the Starwood hotels and accommodations (Fruhlinger, 2020; Sorenson, 2019). Because of this, Starwood's legacy systems, which were already breached before and were infected with malicious software, kept being used until this data breach (Fruhlinger, 2020; Sorenson, 2019).

Marriott International (2018b) published: *"Marriott learned during the investigation that there had been unauthorized access to the Starwood network since 2014."* To be more precise, July of 2014 (Sorenson, 2019). Thus, this happened before Starwood was acquired by Marriott (Fruhlinger, 2020; Sorenson, 2019).

Notably, Marriott was not able to rule out that necessary components for decrypting were taken, since these components were stored at the same server (Fruhlinger, 2020; Marriott International, 2018b; Sorenson, 2019). Both are the result of basic security failures (Ashford, 2018b; Fruhlinger, 2020).

Marriott, as the ultimately responsible party, failed in getting the acquired Starwood IT infrastructure to have a properly performing security (Ashford, 2018b; Fruhlinger, 2020). In fact, the UK's information Commissioner's Office fined Marriott because of this (Information Commissioner's Office, 2020a). Security researchers also spoke of inferior security (Nohe, 2019).

Based on our findings, we consider this data breach to have been caused by security misconfigurations.

Parameters	Values for Marriott International
Data Breach Type	Unauthorized Access
Breach Method	Social Malicious Software
Breach Actor	Malicious Outsider

Organization Sector	Hospitality
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Database
Data Responsibility	Data Controller

Data Subject	Customers
Data Amount (Records)	383.000.000
Data Type	PII PIFI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	Internal Automated Intended
Breach Period (Days)	1514
Time between Breach and Detection	Long
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 40 Summarized overview of identified values for Marriott International

5.2.30. Australian National University (2018)

The Australian National University (ANU) is based in Australia's capital Canberra (Australian National University, n.d.). ANU is the country's sole national university (Australian National University, n.d.). In 2018, ANU had more than four thousand people staff and enrolled more than 26 thousand students (Australian National University, 2019a).

Australian National University (2019b) reports: *"Indications of an intrusion were first detected in April 2019 during a baseline threat hunting exercise. The hunt uncovered network traffic data suggesting the presence of a malicious actor . . ."* After that, the incident response team discovered and confirmed a data breach on May 17, 2019, and reported this to the University's President (Australian National University, 2019b).

Around six months earlier, in the beginning of November 2018, an unauthorized party gained access to ANU's network (Australian National University, 2019b). The actor had a dwell time of approximately six weeks and within that period managed to access a certain section of the network (Australian National University, 2019b). About that, Australian National University (2019b) reports: *"This attack resulted in the breach of part of the network known as the Enterprise Systems Domain (ESD), which houses our human resources, financial management, student administration and enterprise e-forms systems."*

The actor then succeeded in copying and exfiltrating an indeterminable quantity of data from there (Australian National University, 2019b). More specifically, Australian National University (2019b) reports: *"The actor's use of a third-party tool to extract data directly from the underlying databases of our administrative systems effectively bypassed application-level logging."* And Australian National University (2019b) also states: *"Despite our considerable forensic work, we have not been able to determine, accurately, which records were taken."*

Because the actor deleted a significant part of their traces, it could also not be determined specifically which data was taken, thus having to assume all data in those systems were compromised (Australian National University, 2019b). In the breach disclosure, Schmidt (2019) mentions the following about which data was compromised*: "Depending on the information you have provided to the University, this may include names, addresses, dates of birth, phone numbers, personal email addresses and emergency contact details, tax file numbers, payroll information, bank account details, and passport details. Student academic records were also accessed."*

Hence, we assess the compromised data to concern personally identifiable information, personally identifiable financial information and other business operations related data. ANU estimated a number of two hundred thousand people that were affected by the data breach (Martin, 2019).

The actor's attack campaign can roughly be divided into three parts: the theft of credentials, the compromising of ANU's infrastructure, and the exfiltration of data (Australian National University, 2019b). To steal credentials the actor used spearfishing emails in four instances in order to gain usernames and (hashed) passwords (Australian National University, 2019b). The goal was to obtain credentials of multiple people with access to and sufficient authorization within targeted systems (Australian National University, 2019b). The actor liked to obtain accounts from multiple people because the actor anticipated the expiration or exposure of compromised accounts (Australian National University, 2019b).

The first round of spearfishing, which took place on November 9, 2018, made use of a sophisticated email that did not require user interaction (Australian National University, 2019b). Just previewing the email could already be enough to steal the username and password of the recipient, which proved to be the case where a senior staff member's credentials were stolen (Australian National University, 2019b). Also, a network session logger was used to sniff credentials from monitored or redirected network traffic (Australian National University, 2019b).

The actor carefully constructed a “shadow ecosystem” of their own tools and compromised ANU systems, both physical and virtual, to stay undetected, maintain a foothold into the network, map the network, identify targets, run tools and to further compromise other systems (Australian National University, 2019b). The exfiltration of data (including stolen credentials), was done through email or other compromised Internet-facing systems (Australian National University, 2019b). In this way, the actor built a lateral movement infrastructure. Stilgherrian (2019) infers from the incident report by Australian National University (2019b): *“The report details how the attackers organised four separate spearphishing campaigns and built their data exfiltration infrastructure on compromised web servers and legacy systems.”*

The tactics, techniques and procedures used by the actor showed their focus, determination and sophistication (Australian National University, 2019b). Although having broad access to ANU’s network, they seemed to be specifically targeting only the ESD based on their actions and route (Australian National University, 2019b). They operated precisely and efficiently, built and extended their attack infrastructure during the campaign, used customised malicious software, avoided being detected, and exerted a high level of operational security (Australian National University, 2019b). Notably, they left behind very little evidence by wiping their traces and also encrypted data before exfiltrating them (Australian National University, 2019b). Because of the way they operated, ANU was not able to determine exactly what data was taken (Australian National University, 2019b).

Although the attack showed significant sophistication and complexity, ANU acknowledged there were certain people and process issues as well as technical vulnerabilities which contributed to the actor being successful. Notably, unsecure legacy systems were critical in enabling the actor to move through the network, building their infrastructure, and exfiltrating data. Would these legacy systems have been discovered and remediated by the responsible operators, the actor would very likely not have been successful. Therefore, this data breach case is considered to be have security misconfiguration as a root cause.

Parameters	Values for Australian National University
Data Breach Type	Unauthorized Access
Breach Method	Social Malicious Software Hacking
Breach Actor	Malicious Outsider

Organization Sector	Education
Organization Size	Large
Organization Headquarters	AU-AUS-036
Organization Component	Database
Data Responsibility	Data Controller

Data Subject	Students Employees
Data Amount (Records)	200.000
Data Type	PII PIFI OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	Internal Automated Intended
Breach Period (Days)	42
Time between Breach and Detection	Long
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 41 Summarized overview of identified values for Australian National University

5.2.31. Target (2013)

Target is a discount retailer based in Minneapolis, Minnesota, United States (Target, n.d.). Plachkinova and Maurer (2018) state: *"various acquisitions and expansions into new areas of the country, Target has become the second-largest discount retailer in the United States (behind Walmart)."* In the beginning of 2014, Target operated almost eighteen hundred retail stores in the US and employed around 360 thousand people (Plachkinova & Maurer, 2018). Target's revenue in the US in 2013 amounted to 71.3 billion US dollars (Statista Research Department, 2019).

At the end of November and the beginning of December of 2013, multiple alerts were left without action taken by Target (Plachkinova & Maurer, 2018). Target began their investigation after the US Department of Justice on December 12, 2013, contacted Target about an incident and possible data breach (Plachkinova & Maurer, 2018). On December 18, the independent security researcher Brian Krebs made a blog post about an investigation into a breach at Target (Krebs, 2013a). The next day, Target confirmed a data breach (Krebs, 2013b). Plachkinova and Maurer (2018) mentioned: *"The breach was first reported by the security journalist Brian Krebs, and Target's official response came shortly after the announcement."*

Investigation eventually learned that the personal information of 70 million customers was compromised, including names, addresses, phone numbers and email addresses (Plachkinova & Maurer, 2018; Target, 2014). Approximately 40 million customers' payment card data was also compromised (Plachkinova & Maurer, 2018). Thus, a total of 110 million customers were affected (Kassner, 2015; Krebs, 2014a). The payment card data were offered for sale on the online black market (Krebs, 2013b). The data breach was the result of an attack that started on November 27 until December 15, 2013, when it was stopped (Kassner, 2015).

The attack started at least two months earlier with a phishing email to Fazio Mechanical, a heating, ventilation and air conditioning company that worked as a third party service provider for Target (Aorato Labs, 2014; Kassner, 2015; Krebs, 2014a, 2015; Plachkinova & Maurer, 2018). Through that email, the attackers managed to get Fazio infected with the malicious software Citadel, which was used to collect web application credentials from the infected machine's browser (Aorato Labs, 2014; Kassner, 2015; Plachkinova & Maurer, 2018).

As Target's vendor, Fazio was provided access to certain administrative Target vendor systems and which Fazio used exclusively for electronic billing, contract submission and project management (Aorato Labs, 2014; Plachkinova & Maurer, 2018). Hence, the attackers were able to use the stolen credentials to access those vendor dedicated Target systems, which are all web applications hosted on Target's internal network (Aorato Labs, 2014).

The web application chosen by the attackers had a vulnerability in its web interface (Aorato Labs, 2014). This vulnerability enabled the attackers to execute code on the application's server, which they exploited (Aorato Labs, 2014). From there the attackers pursued their adventure by gathering information on Target's network and identifying relevant targets (Aorato Labs, 2014). For this, the attackers addressed the Active Directory, which contained data on all users, computers and services in that Windows domain (Aorato Labs, 2014). The Active Directory could easily and without authorization be queried for the names of relevant computers and services, and from there the attackers even obtained their locations (Aorato Labs, 2014). Furthermore, the attackers grabbed an existing domain administrator's connection token from the web application server's memory, which enabled them to access Active Directory and to create a new domain administrator account in there.

Knowing the relevant systems and having arranged access to them, the attackers continued towards their targets and had to bypass network security and had to succeed in running remote processes on various machines on their path to the relevant targets (Aorato Labs, 2014). Once propagated to the relevant computer, the attackers could use Structured Query Language (SQL) related query tools to check some of the data in its database (Aorato Labs, 2014). As soon as they determined the data was valuable, a bulk SQL copy tool was used to copy all of that database contents (Aorato Labs, 2014). Probably the attackers then realized that

the Target databases only held personal information on customers and no relevant payment card data, since Target operated their databases in accordance with Payment Card Industry Data Security Standard (PCI-DSS) (Aorato Labs, 2014). In its report, Aorato Labs (2014) states: *"Since the database is PCI-compliant, no credit cards are stored on it."*

The attackers decided to continue looking for the more valuable payment card data, and propagated to point-of-sale (POS) machines (Aorato Labs, 2014). They managed to install the malicious software Kaptoxa (pronounced "Kar-toe-sha") on the POS machines in all of Target's stores (Aorato Labs, 2014; Krebs, 2015). Aorato Labs (2014) reports about the malware: *"The malware scanned the memory of the POS machine and when identifying a credit card, it saved it to a local file."*

Plachkinova and Maurer (2018) state: *"The malware contained memory-scraping functionality that allowed the attackers to intercept cardholder information before it was sent for processing by a payment processor."* Furthermore, Plachkinova and Maurer (2018) state: *". . . [T]he configuration of point of sale terminals at Target did not provide the ability to immediately encrypt cardholder data upon registering a card swipe. Because of this, card data remained in plain text within the POS terminal's memory."* The data remained unencrypted in the POS machine's memory until it got prepared for transit (Plachkinova & Maurer, 2018). This is not according the requirements of PCI-DSS (Plachkinova & Maurer, 2018). The attackers tested the malware first on a small number of POS machines first (Shu, Tian, Ciambrone, & Danfeng, 2017; Vijayan, 2014).

The malware periodically copied this local file to a remote FTP-enabled machine inside Target's network (Aorato Labs, 2014; Kassner, 2015). In order to do so, the malware had created a remote file share by using Windows internal commands and admin credentials (Aorato Labs, 2014). The attackers also arranged that the personal information from databases got to the FTP-enabled machine (Aorato Labs, 2014). From there a script sent the data via FTP to external attacker-controlled servers (Aorato Labs, 2014; Kassner, 2015).

The attackers showed some sophistication by being able to intrude and move laterally in a large and complex network, and by succeeding in exfiltrating data from it (Jarvis & Milletary, 2014; Kassner, 2015). But the most sophistication was actually only shown by using that specific malware in that specific way to intercept and take data from the POS machines, and send it to the remote file share (Vijayan, 2014). However, as Vijayan (2014) presented: *". . . [T]he attackers would have been unable to install the malware if Target had employed proper network segmentation practices in the first place."* That would have not allowed the attackers to move freely and unnoticed to all parts of the Target network (Krebs, 2013a; Plachkinova & Maurer, 2018; Radichel, 2014).

Also, Target did not secure their POS machines properly, which allowed unauthorized software installation and configuration (Radichel, 2014; Shu et al., 2017). Sensitive data should not have been allowed within the POS environment unencrypted (Radichel, 2014).

Furthermore, Target's investigation found that several systems were outdated or unpatched (Krebs, 2015; Plachkinova & Maurer, 2018). And this all started at Fazio Mechanical according to Plachkinova and Maurer (2018): *". . . [I]t has been alleged that Fazio relied on the free, non-commercial version of Malwarebytes Anti-Malware software, which does not provide real-time protection."* That software did not present appropriate real-time protection (Kassner, 2015).

In general, vulnerable configuration allowed existing security strategies to be bypassed (Radichel, 2014). The attackers could have been slowed down or stopped at various steps of their attack if certain controls would have in place, like network segregation and segregation of the POS machines, multifactor authentication of Target administrator and third party vendor accounts, end-to-end encryption among others, as well as if proper detection was in place (Radichel, 2014).

Different security experts assessed this data breach was not all sophistication and was preventable, because of simple errors enabling the attackers to execute their more or less sophisticated activities (Field, 2014; Kassner, 2015; Krebs, 2015; Vijayan, 2014). Based on our findings, we agree with them and we determine the root cause lies in security misconfiguration.

Parameters	Values for Target
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Social Malicious Software Hacking
Breach Actor	Malicious Outsider

Organization Sector	Retail
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Database Device
Data Responsibility	Data Controller

Data Subject	Customers
Data Amount (Records)	110.000.000
Data Type	PII PIFI
Data Sensitivity	Internal
Data State	DAR DIU
Data Necessity	Necessary

Data Breach Detection	External Inconclusive Inconclusive
Breach Period (Days)	19
Time between Breach and Detection	Short
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 42 Summarized overview of identified values for Target

5.2.32. Panera Bread (2018)

Panera Bread is a United States (US) based company in the fast-casual restaurant segment, consisting of a chain of more than two thousand bakery-cafés in the US and Canada (Panera Bread, n.d.-a, n.d.-b). In 2019, Panera had 125 thousand employees (Panera Bread, 2020). Through its website Panerabread.com, customers after signing up have the possibility to order food online for delivery or for store pickup (Krebs, 2018a).

On August 2, 2017, security researcher Dylan Houlihan reported a vulnerability to Panera which left customer data publicly accessible (Houlihan, 2018; Krebs, 2018a). After first considering this a scam, a week later Panera responded with a reaction which suggested an acknowledgement of the issue and indicated Panera was working on a solution (Houlihan, 2018; Krebs, 2018a). However, after eight months the vulnerability still was not fixed and Houlihan decided to publicly disclose this issue on April 2, 2018 (Houlihan, 2018). On that same day, he also contacted well-known, influential persons from the IT security sector, notably security expert and investigative reporter Brian Krebs, for help to escalate this issue (Houlihan, 2018). After Krebs had contact with Panera, Panera's website was shortly offline and came back online seemingly fixed (Krebs, 2018a). But that so called fix by Panera consisted of requiring people to log in with a valid Panera user account and then still be able to view the exposed customer data, which obviously is not really a fix (M. Smith, 2018).

The exposed data pertained to up to 37 million Panera customers who had signed up for an account to be used for online food ordering and included full names, home addresses, email addresses, usernames, phone numbers, dates of birth, last four digits of saved credit card numbers, information related to the integration with social accounts, and saved user food preferences and dietary restrictions (Houlihan, 2018; Krebs, 2018a).

Even worse, Panera used unique sequential integers for account identifiers, which could allow simple incremental data gathering (by means of indexing, crawling and scraping) from available customer records, possibly by using automated tools (Krebs, 2018a). Furthermore, the format of the compromised database allowed anyone to search for specific customers on the basis of various parameters, for instance on phone number (Krebs, 2018a).

It appeared one of Panera's web applications (accessible through the web page delivery.panerabread.com) had a publicly available, completely unauthenticated application programming interface (API) endpoint (Houlihan, 2018). Later it became clear the same vulnerability was also present on other API endpoints, meaning the issue existed extended to all other parts of Panera's organization, for instance to Panera's commercial division available through the web application at catering.panerabread.com (Houlihan, 2018; Krebs, 2018a; M. Smith, 2018).

Implementing sound authentication and authorization is essential to properly secure APIs (George, 2018). Not doing this and leaving a hole through which anyone can reach sensitive data the customers entrusted Panera with, is considered a practice of poor security (Houlihan, 2018; Kovacs, 2018a; Muncaster, 2018b). We agree with this and establish the root cause lies in security misconfiguration.

Parameters	Values for Panera Bread
Data Breach Type	Unauthorized Disclosure
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider
Organization Sector	Retail
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Database
Data Responsibility	Data Controller
Data Subject	Customers
Data Amount (Records)	37.000.000
Data Type	PII PIFI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary
Data Breach Detection	External Non-Auto Inadvertent
Breach Period (Days)	244
Time between Breach and Detection	Long
Time between Detection and Action	Long
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 43 Summarized overview of identified values for Panera Bread

5.2.33. Suprema (2019)

Suprema is a South Korean company that provides access control, time and attendance, and biometrics solutions (Suprema, n.d.; Suprema Inc., n.d.). Suprema (n.d.) says: *"Suprema is named the world's top 50 security manufacturers and has a worldwide sales network in over 140 countries with no.1 market share in biometric access control in EMEA region."* Its customers are numerous organizations of various sizes and from different sectors (Fawkes, 2019). Although we could not find an exact number of employees working for Suprema, the sources we found indicated Suprema is a large organization in terms of number of employees (Crunchbase, n.d.-h; Suprema Inc., n.d.).

Suprema's core product is web-based comprehensive integrated security platform BioStar 2 (Fawkes, 2019). In a blog post, Fawkes (2019) says about BioStar 2: *"A centralized application, it allows admins to control access to secure areas of facilities, manage user permissions, integrate with 3rd party security apps, and record activity logs. As part of the biometric software, BioStar 2 uses facial recognition and fingerprinting technology to identify users."* Part of the BioStar 2 platform is its own centralized information storage, an Elasticsearch database where customers can store the data necessary for operating BioStar 2, so that customers do not need to store those data on premise (Fawkes, 2019; Lei, 2019). In August 2019, BioStar 2 had more than 1.5 million installations in over eighty countries (Fawkes, 2019; Lei, 2019; Townsend, 2019).

A research team formed by vpnMentor, which is an advice organization on VPN and Web Anonymity (vpnMentor, n.d.), ran a web-mapping project to find vulnerabilities that may lead to security breaches (Lei, 2019). For this they used Internet Protocol (IP) and port scanning methods to find open ports to which requests might be sent by threat actors (Lei, 2019). Fawkes (2019) reports the following in his blog post: *"The team discovered that huge parts of BioStar 2's database are unprotected and mostly unencrypted."* This discovery was done on August 5, 2019 (Fawkes, 2019).

Fawkes (2019) continues: *"The company uses an Elasticsearch database, which is ordinarily not designed for URL use. However, we were able to access it via browser and manipulate the URL search criteria into exposing huge amounts of data."* In that way it was shown that the data in that database was exposed as a result of the database being publicly accessible (Fawkes, 2019). The database consisted of data collected by Suprema's customers that utilized BioStar 2 (O'Donnell, 2019b).

Following this discovery, the research team examined the database to confirm its origin and subsequently contacted Suprema on August 7, 2019 (Fawkes, 2019). Initially the team did not receive a response nor the expected cooperation from Suprema offices (among which the office of the responsible GDPR compliance officer), until they managed to get hold of an office that was a bit more cooperative (Fawkes, 2019; Townsend, 2019). Subsequently, the breach was fixed on August 13, 2019, when Suprema closed off the database (Fawkes, 2019; O'Donnell, 2019b). It remained unknown how long the database was exposed (Lei, 2019).

The compromised data concerns personal information pertaining to customers and employees of Suprema's customers, or people related in another way to these customers, as well as other business information relevant to Suprema's customers (Fawkes, 2019). About the compromised data, Fawkes (2019) says:

> *Our team was able to access over 27.8 million records, a total of 23 gigabytes of data, which included the following information:*
>
> - *Access to client admin panels, dashboards, back end controls, and permissions*
> - *Fingerprint data*
> - *Facial recognition information and images of users*
> - *Unencrypted usernames, passwords, and user IDs*
> - *Records of entry and exit to secure areas*
> - *Employee records including start dates*
> - *Employee security levels and clearances*
> - *Personal details, including employee home address and emails*

- *Businesses' employee structures and hierarchies*
- *Mobile device and OS information* (Fawkes, 2019, Example of Entries in the Database section)

If these exposed data actually leaked, then the entire security infrastructure of each of Suprema's customers could easily have become useless (Fawkes, 2019). Fawkes (2019) states: *"This leak could have been easily avoided, had the makers of BioStar 2 taken some basic security precautions."* That would have prevented an actor from finding the vulnerable database and prevent the actor from accessing it (Fawkes, 2019; Lei, 2019). But instead, Suprema acted negligent and practiced poor security practices, including flawed access control and insecure data storage (Fawkes, 2019; Lei, 2019; Porter, 2019). The Elasticsearch database should have been password protected and not using default passwords (Fawkes, 2019; Lei, 2019). Additionally, the database could have been set up to only accept requests from trusted IP addresses (Lei, 2019). Also, the use of a properly configured firewall would have blocked requests and is an effective defence method against scanning attacks (Lei, 2019). And there are more practices available that would have made such an application less vulnerable (Fawkes, 2019; Lei, 2019).

Based on our findings, we determine that the root cause of this data breach lies in security misconfiguration.

Parameters	Values for Suprema
Data Breach Type	Unauthorized Access
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider

Organization Sector	Technology
Organization Size	Large
Organization Headquarters	KR-KOR-410
Organization Component	Database
Data Responsibility	Data Processor

Data Subject	Customers Employees Members Other
Data Amount (Records)	27.800.000
Data Type	PII BD LI OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	External Automated Intended
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 44 Summarized overview of identified values for Suprema

5.2.34. Abine (2019)

Abine Inc. is an online privacy company based in the United States which offers tools to consumers that enable them to control the level of exposure of their personal information (Abine, n.d.; Abine Inc., n.d.). Although we could not determine an exact number of employees, we could infer from different sources that the company can be considered a small company in terms of employees according to our criteria (Abine Inc., n.d.; Craft, n.d.-a; Crunchbase, n.d.-b).

Abine's flagship product is Blur, a security application that offers password management capabilities, and a range of tools related to private browsing and the concealment of email addresses, phone numbers and credit card numbers (Abine Inc., n.d.; Duffy, 2020).

On December 31, 2018 Abine Online Privacy (2018) announced in a post on their blog: *"On Thursday, December 13th 2018, we became aware that some information about Blur users had been potentially exposed . . ."* Additionally, Cimpanu (2019a) reports: *"The breach came to light last year, on December 13, when a security researcher contacted the company about a server that exposed a file containing sensitive information about Blur users, an Abine spokesperson told ZDNet via email."* Abine immediately acted to secure the exposed system and began to investigate the reported incident (Abine Online Privacy, 2018).

Cimpanu (2019a) states the following about the affected data subjects: *". . . a data breach impacting nearly 2.4 million Blur users . . ."* This concerns users who had registered for an account prior to January 6, 2018 (Abine Online Privacy, 2018). There was no information about the length of the exposure period (Muncaster, 2019).

A file was left publicly accessible in a cloud storage facility (Abine Online Privacy, 2018; Cimpanu, 2019a). Abine Online Privacy (2018) states that it contained the following exposed data:

- *Each user's email addresses*
- *Some users' first and last names*
- *Some users' password hints but only from our old MaskMe product*
- *Each user's last and second-to-last IP addresses used to login to Blur*
- *Each user's encrypted Blur password.* (Abine Online Privacy, 2018)

The data breach was the result of a misconfigured Amazon Web Services (AWS) Simple Storage Service (S3) bucket (Abrams, 2019; Kovacs, 2019a; Muncaster, 2019). Abine used this bucket for data processing (Abrams, 2019; Kovacs, 2019a).

Parameters	Values for Abine
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider

Organization Sector	Technology
Organization Size	Small
Organization Headquarters	US-USA-840
Organization Component	Cloud Storage
Data Responsibility	Data Controller

Data Subject	Customers
Data Amount (Records)	2.400.000
Data Type	PII PIFI LI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary

Data Breach Detection	External Inconclusive Inconclusive
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 45 Summarized overview of identified values for Abine

5.2.35. Texas Voter Records (2018)

On August 23, 2018, technology news website TechCrunch published a report treating the exposure of voter records containing the personal information of people residing in Texas in the United States (Whittaker, 2018c). A data breach hunter claimed to have found a file holding 14.8 million records on an unsecured server, not even password protection, and provided TechCrunch with a copy of the file (Whittaker, 2018c; Zurkus, 2018).

The file was 16 gigabytes in size and included voters' names, addresses, gender, phone numbers, ethnical and racial identity, voting history, and information relating to political affiliations and party memberships (Whittaker, 2018c). The file also included scores on individual voters' perceived views on certain societal issues, subjects or even persons (Whittaker, 2018c).

The data was analysed by another security researcher, and he concluded the data was originally compiled by data analytics firm Data Trust, or that Data Trust at least had some relation to that data (Price, 2018; Vickery, 2018; Whittaker, 2018c). Data Trust was created by the Republican Party, and provides voter and electoral data to Republican and conservative campaigns, parties, and advocacy organizations (Data Trust, n.d.; Whittaker, 2018c).

Data Trust stated the exposed data could not be confirmed as data from Data Trust (Whittaker, 2018c). Furthermore, Data Trust denied a breach of its systems as it was confident about not having hosted or transferred data on an unsecure, publicly accessible platform (Price, 2018; Whittaker, 2018c). It remained unclear who the owner was of the unprotected server where the file was found (Whittaker, 2018c). Therefore, there was no one to notify about this (Whittaker, 2018c). Therefore it also remained unclear whether the leak was fixed and whether the data remained exposed (Whittaker, 2018c). Apparently the data breach hunter only disclosed the exposed data and gave his observation of why the file was publicly accessible.

Because of the little information that accompanied the discovery of this data breach, the organization responsible for the server and the data could not be found in order to provide additional explanation about the data itself and about what happened when and for how long. This results in multiple parameters of our assessment of this case to remain inconclusive. However, based on the data breach hunter's assertion about the unsecured server, we are able to establish that the root cause of this data breach is a security misconfiguration.

Parameters	Values for the case of the Texas Voter Records
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider
Organization Sector	Inconclusive
Organization Size	Inconclusive
Organization Headquarters	US-USA-840
Organization Component	Inconclusive
Data Responsibility	Inconclusive
Data Subject	Citizens
Data Amount (Records)	14.800.000
Data Type	PII OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary
Data Breach Detection	External Inconclusive Intended
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Inconclusive
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 46 Summarized overview of identified values for the case involving Texas Voter Records

5.2.36. 500px (2018)

500px is an online community and social network for people interested in photography, operating through a web site and application with the same name enabling members to receive exposure and feedback, and to monetize their photographic work (500px, n.d.-a, n.d.-c; Crunchbase, n.d.-a). A number of 35 employees works for the Canada-based 500px (500px, n.d.-b, n.d.-c).

On February 8, 2019, 500px's engineering team was made aware of a potential security incident involving sensitive data (500px, 2019; Dunn, 2019; Williams, 2019). From the following investigation 500px learned that around July 5, 2018, an unauthorized party gained access to its systems and got hold of certain user data (Williams, 2019).

Technology news website The Register had earlier encountered a huge amount of accounts data, belonging to various organizations, being offered for sale on a darknet market (Williams, 2019). Included in this large trove of offered pilfered accounts were 500px user accounts (Williams, 2019). Allegedly, the offer concerning 500px data involved 14,870,304 accounts (Williams, 2019).

The 1.5 gigabytes of compromised data consisted mostly of data provided by users setting up their profiles (500px, 2019). In a security issue notification, 500px (2019) announced:

> *What type of user data was affected?*
>
> - *Your first and last name as entered on 500px*
> - *Your 500px username*
> - *The email address associated with your 500px login*
> - *A hash of your password, which was hashed using a one-way cryptographic algorithm*
> - *Your birth date, if provided*
> - *Your city, state/province, country, if provided*
> - *Your gender, if provided* (500px, 2019)

The vulnerability was quickly fixed (Liptak, 2019). Although 500px had promised to provide updates of findings from their investigation (completetechnology, 2019), no further explanation or details have been shared. However, the party advertising the accounts sale gave a general statement about how the user account data was obtained (with regard to all accounts of all affected organizations), namely by typically exploiting security vulnerabilities within web applications which enabled remote code execution, after which user account data could be exfiltrated (Williams, 2019).

We did not find any further information explaining in more detail what happened and how this data breach could happen. Therefore, we are unable to determine whether this data breach was the result of a complex, sophisticated attack or the result of a security misconfiguration. Hence, the root cause could not be established and remains inconclusive.

Parameters	Values for the case of the 500px
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Hacking
Breach Actor	Malicious Outsider

Organization Sector	Technology
Organization Size	Small
Organization Headquarters	CA-CAN-124
Organization Component	Web Server Database
Data Responsibility	Data Controller

Data Subject	Members
Data Amount (Records)	14.870.304
Data Type	PII LI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Mixed

Data Breach Detection	External Non-Auto Inconclusive
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Long
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Inconclusive

Table 47 Summarized overview of identified values for 500px

5.2.37. Canva (2019)

Canva is an Australia-based company that operates an online design and publishing platform with the same name (Canva, n.d.; Crunchbase, n.d.-c). It belongs to Australia's biggest companies in the technology sector (Cimpanu, 2019c). This platform provides user friendly and easy to use graphic design tools for different purposes (Cimpanu, 2019c; Crunchbase, n.d.-c). In December 2020, Canva was employing 1375 people, which has grown over time (Craft, n.d.-b). In 2019 the number of employees was approximately 550 (The Latka Agency, n.d.).

On May 24, 2019, Canva detected a security incident in their systems (Welsh, 2019). This concerned a malicious attack which Canva managed to stop (Welsh, 2019). Being interrupted during the attack and knowing that Canva is aware of the incident, the hacker organization directly decided to make the attack public the same day (Cimpanu, 2019c; Welsh, 2019). GnosticPlayers, as the infamous hacker organization is known, tweeted about the attack (Welsh, 2019). The hackers also contacted technology news website ZDNet and stated that earlier that day they had breached Canva (Cimpanu, 2019c). In his article, Cimpanu (2019c) quotes the hacker: *"'I download everything up to May 17,' the hacker said. 'They detected my breach and closed their database server.'"*

Cimpanu (2019c) further reports: *"ZDNet requested a sample of the hacked data, so we could verify the hacker's claims."* They received a sample of almost 19 thousand user accounts, even including accounts of Canva staff members and administrators (Cimpanu, 2019c). Cimpanu (2019c) continues: *"We used this information to contact Canva users, who verified the validity of the data we received."*

Quickly making the data breach public was probably done by the hackers as a sort of advertising and promotion of the stolen data and of their own skills (Christou, 2019). Also, by making the data breach public Canva had to come with a quick communication response, diverting full attention from its investigation and incident response (Welsh, 2019). Canva did respond quickly by tweeting and placing a statement on their website about the security incident within a day, and eventually explaining the data breach a few days later (Canva, 2019; Patrawala, 2019; Welsh, 2019).

The hacker were able to at least access, and therefore compromise, the following data in the database:

- Up to 139 million user records, each record containing usernames, names, email addresses and country. Also, if the user provided this, the records then also included data about their location (city) or their homepage URL (Welsh, 2019).
- For 61 million of those, there were also present password hashes (Cimpanu, 2019c). The passwords were hashed by using bcrypt (Cimpanu, 2019c; Welsh, 2019). This hashing mechanism includes salting as part of the hashing process (Arias, 2018). Currently bcrypt is one of the more robust mechanisms, providing adequate security (Cimpanu, 2019c; Csaszar, 2019).
- *"Of the total 139 million users, 78 million users had a Gmail address associated with their Canva account"* (Cimpanu, 2019c). This part of the users had signed in via Google and the Google tokens were obtained (Cimpanu, 2019c). These OAuth tokens were encrypted by means of AES128 and the encryption keys were securely stored elsewhere (Welsh, 2019). Canva stated that they did not find any evidence that these token were taken or the keys were accessed (Welsh, 2019).
- According to Canva, the hackers managed to briefly access partial payment and payment card data, but no evidence was found this data was taken (Welsh, 2019). Payment data concerned payment histories before September 16, 2017, and included dollar amounts, dates, and payment IDs (Welsh, 2019). About the payment card data, Welsh (2019) announces: *"Files contained partial credit card data from before September 28, 2016 (name, expiry date, last 4 digits, card brand and card country) . . ."*

Shortly after the breach, the captured data was already found to be offered in underground parts of the Internet (Breach Report, 2019). also reports the following: " In the beginning of 2020, it appeared that at least

two sets of data from this data breach had resurfaced and were encountered on the Internet, but in those instances containing decrypted passwords (Breach Report, 2020; Welsh, 2019).

The available sources did not specify how the attack took place. Hence, we could not establish how the hackers accessed Canva's network, how they moved through it, and how they eventually exfiltrated the data. Therefore, we were unable to determine whether the root cause of this data breach is a security misconfiguration or a complex, sophisticated attack.

Parameters	Values for the case of Canva
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Hacking
Breach Actor	Malicious Outsider
Organization Sector	Technology
Organization Size	Large
Organization Headquarters	AU-AUS-036
Organization Component	Database
Data Responsibility	Data Controller
Data Subject	Customers
Data Amount (Records)	139.000.000
Data Type	PII PIFI LI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary
Data Breach Detection	Internal Inconclusive Inconclusive
Breach Period (Days)	1
Time between Breach and Detection	Short
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Inconclusive

Table 48 Summarized overview of identified values for Canva

5.2.38. First American Financial Corporation (2019)

First American Financial Corporation (FAFC) is a company based in the United States operating in the financial and insurance sector (First American, n.d.). On its profile page, First American (n.d.) states: *"First American provides financial services through its Title Insurance and Services segment and its Specialty Insurance segment."* Krebs (2019a) says the following about FAFC: *"Santa Ana, [California] based First American is a leading provider of title insurance and settlement services to the real estate and mortgage industries."* In 2019, FAFC had more than eighteen thousand employees (First American Financial Corporation, 2019, 2020; Macrotrends, n.d.).

On May 20, 2019, the Washington real estate developer Ben Shoval had discovered a security incident related to FAFC's website (Timberg & Merle, 2019). He noticed that that simply changing the URL returned documents of others (Ikeda, 2019b). Over the next few days, Shoval reached out to FAFC multiple times but to no avail (Timberg & Merle, 2019). Therefore, on May 23, 2019, Shoval contacted investigative security journalist Brian Krebs about this security incident. Krebs (2019a) about what Shoval told him: *"He said anyone who knew the URL for a valid document at the Web site could view other documents just by modifying a single digit in the link."* In the first place, this includes every person that had ever received a document link through email from FAFC (Krebs, 2019a). Normally, the people and businesses involved in different types of transactions which were handled by the FAFC, would receive a link to the records of their transactions per email (Krebs, 2019a).

Krebs (2019a) reports: *"Shoval shared a document link he'd been given by First American from a recent transaction, which referenced a record number that was nine digits long and dated April 2019."* Krebs (2019a) further reports the following about what Shoval found out and what he could reproduce: *"Modifying the document number in his link by numbers in either direction yielded other peoples' records before or after the same date and time, indicating the document numbers may have been issued sequentially."* Everyone with access to Internet and a browser could do this and all documents could be read without being required to authenticate (Krebs, 2019a).

Hence, Krebs was able to confirm the issue and further look into it, which lead him to believe approximately 885 million files were left exposed, of which the earliest, i.e. having the lowest available record number, was from 2003 (Krebs, 2019a). These files were mainly records of transactions involving property buyers and sellers, and therefore included their submitted sensitive personal or business data (Krebs, 2019a). These digitized records contained names, marital statuses and physical addresses (Krebs, 2019a). Krebs (2019a) states the records also contained: *". . . bank account numbers and statements, mortgage and tax records, Social Security numbers, wire transaction receipts, and drivers license images . . ."*

Before publicly disclosing the data exposure, Krebs also again notified FAFC about this issue after which FAFC shut down external access to the application on May 24, 2019 (Kirk, 2019c; Krebs, 2019a). It remained unclear for how long the data was exposed (Ikeda, 2019b; Krebs, 2019a). However, as Krebs (2019a) reports: *". . . but archive.org shows documents available from the site dating back to at least March 2017."* It also stayed uncertain whether the data was extracted or even just accessed (Ikeda, 2019b; Krebs, 2019a). But according to our definition, the data exposure is already enough for this to be considered a data breach.

FAFC failed to properly secure the unique URLs to the transaction records, allowing access and making them available without authentication to unauthorized parties (Ikeda, 2019b; Krebs, 2019a). This vulnerability is also known as an Insecure Direct Object Reference (IDOR) (Padwal, Thomas, Howard, & Carr, 2019). IDOR is defined by Nidecki (2020) as: *"a cybersecurity issue that occurs when a web application developer uses an identifier for direct access to an internal implementation object but provides no additional access control and/or authorization checks."*

In the case of FAFC, upon completion of a transaction, a unique link to a web page holding the relevant information for a specific involved party was created and was meant only to be accessed by that party (Dellinger, 2019). It basically concerned a link to a specific web page holding sensitive information, but without having implemented an identity verification method (Dellinger, 2019). The resulting situation was that anyone

who would encounter a link to a certain web page holding transaction record documents, would be able to view the documents on that page (Dellinger, 2019). Furthermore, by modifying the link, other web pages holding other documents would also be freely accessible (Dellinger, 2019). And in that situation, it did not help that the record numbers were issued sequentially. Remarkably, at the time IDOR already was a known and common issue in web applications (Dellinger, 2019; Ikeda, 2019b; Kirk, 2019c; Padwal et al., 2019).

The presence of an IDOR indicates a misconfigured web server (Padwal et al., 2019). Better security hygiene and training could have made FAFC personnel more aware and could have resulted in the responsible people proactively looking for vulnerabilities like IDOR (Nair, 2020; Padwal et al., 2019). Additionally, the encountered vulnerabilities would probably have been addressed and fixed in an adequate and timely manner (Padwal et al., 2019). This did not happen in the case of FAFC, since this IDOR vulnerability already came to light earlier (Nair, 2020). All in all, this web application was not protected which shows poor security practice in this matter (Ikeda, 2019b; Krebs, 2019a). Hence, this data breach was preventable (Ikeda, 2019b; Krebs, 2019a). Based on our findings, we establish the root cause to be a security misconfiguration.

Parameters	Values for the case of First American Financial Corporation
Data Breach Type	Unauthorized Disclosure Unauthorized Access
Breach Method	Poor or No Security
Breach Actor	Inadvertent Outsider
Organization Sector	Financial and Insurance
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Web server
Data Responsibility	Data Controller
Data Subject	Customers
Data Amount (Records)	885.000.000
Data Type	PII PIFI OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary
Data Breach Detection	External Non-Auto Inadvertent
Breach Period (Days)	800
Time between Breach and Detection	Long
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 49 Summarized overview of identified values for First American Financial Corporation

5.2.39. "BreedReady" (2019)

On the Friday of March 8, 2019, security researcher and ethical hacker Victor Gevers discovered an open and publicly accessible database during his search for unprotected MongoDB databases on the Internet (Doffman, 2019; Ilascu, 2019; Kuo, 2019). MongoDB is an open source, non-relational, document oriented database management system (Boicea, Radulescu, & Agapin, 2012). Gevers could not identify the owner of the database (Ilascu, 2019; Kuo, 2019; Leung, 2019).

The exposed database contained information on more than 1.8 million women, including fields for names, ages, dates of birth, gender, addresses, phone numbers, location information, identification numbers, marital statuses, URL links to photos, education, political affiliation, and even values for a 'HasVideo' field (Doffman, 2019; Ilascu, 2019; Kuo, 2019). Remarkably, the database contained a 'BreedReady' status for every woman (Claburn, 2019; Doffman, 2019; Ilascu, 2019).

Gevers tweeted about this discovery on March 9, 2019, including a redacted screenshot of one of the database entries (Gevers, 2019a). By doing this he hoped to draw attention of the unknown database owner to the exposed database and to have that database owner take measures to stop the exposure and to remove or secure the database (Ilascu, 2019). This succeeded, because the field 'BreedReady' generated media attention and caused some controversy (Claburn, 2019; Ilascu, 2019). On March 11, 2019, Gevers tweeted that the database was no longer reachable (Gevers, 2019b).

On march 16, 2019, Gevers tweeted that he had discovered two other databases with similar database schemas and which were hosted and maintained on servers linked to a university in the Shangdong region (Gevers, 2019c). In the same tweet Gevers already suspected this concerned an accidentally exposed educational student project (Gevers, 2019c). On March 22, 2019, Gevers tweeted that he had learned from a student that these databases indeed were part of a scientific student project on how to use Big Data to find solutions for a social problem (Gevers, 2019d). The dataset originated from an official governmental source and the identifiable information was real (Gevers, 2019e). The 'BreedReady' field was automatically filled purely based on age (Gevers, 2019d). We could not find out to which specific university the databases belong.

The student also had stated that the Internet Service Provider had made a mistake by adjusting firewall settings and accidentally allowing external traffic (Gevers, 2019e). Thus, the exposure of these databases happened because of a misconfigured firewall, which resulted in allowing external parties access to the databases in the internal network (Ilascu, 2019).

Parameters	Values for the case of 'BreedReady'
Data Breach Type	Unauthorized Disclosure Unauthorized Access
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider

Organization Sector	Education
Organization Size	Unidentifiable
Organization Headquarters	CN-CHN-156
Organization Component	Database
Data Responsibility	Data Processor

Data Subject	Citizens
Data Amount (Records)	1.800.000
Data Type	PII
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Mixed

Data Breach Detection	External Automated Intended
Breach Period (Days)	Unidentifiable
Time between Breach and Detection	Unidentifiable
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 50 Summarized overview of identified values for the case known as 'BreedReady'

5.2.40. Ticketmaster (2018)

Ticketmaster is a ticket sales and distribution company based in the US and which operates globally and considers itself an industry leader (Ticketmaster, n.d.-a). Ticketmaster has more than 6,500 employees worldwide (Ticketmaster, n.d.-b).

Since April of 2018, Ticketmaster was notified on multiple instances by at least one financial organization as being the common point of purchase in relation to fraudulent activities (Vernier, 2018). In that same period, Ticketmaster was addressed by people who had discovered very suspicious code and by people whose antivirus software had identified Ticketmaster's website as malicious (Information Commissioner's Office, 2020c). But it was not until June 27, 2018, that Ticketmaster made public that it had suffered a security incident (ticketmasteruk, 2018).

On June 23, 2018, Ticketmaster discovered that the customer support product, which was supplied by third party Inbenta Technologies and ran on various Ticketmaster websites, was manipulated to steal data (Schwartz, 2018e; Ticketmaster, 2018). This concerned the data which was filled in by customers when going through Ticketmaster's payment process (Information Commissioner's Office, 2020c). The customers went through that process to buy tickets, and the incident affects customers that bought tickets between September 2017 and June 23 of 2018 (Ticketmaster, 2018). In that way personal information got compromised, including names, addresses, email addresses, telephone numbers, payment and payment card information, and Ticketmaster login details (Schwartz, 2018e; Ticketmaster, 2018).

Ticketmaster did not explicitly specify the compromised amount of data (Schwartz, 2018e).Ticketmaster first reported only around 40,000 United Kingdom (UK) customers were directly affected (Schwartz, 2018e). But Ticketmaster also stated that less than 5% of their global customer base was affected by this incident (Ticketmaster, 2018). At that time, Ticketmaster claimed to serve more than 230 million customers per year and 5% of that makes 11.5 million (Schwartz, 2018e). However, Ticketmaster stated that customers in North America had not been affected (Ticketmaster, 2018). So these should not be taken into account and we cannot assume the number of 11.5 million.

On the other hand, the Information Commissioner's Office (ICO) assessed that 9.4 million European customers, of whom 1.5 million UK customers, were potentially compromised (Information Commissioner's Office, 2020c). But the ICO only considered the period between May 25 and June 23 of 2018 (Information Commissioner's Office, 2020c). This while according to the ICO the total duration of the breach was between February 10 and June 23 of 2018 (Information Commissioner's Office, 2020c). Taking into consideration a shorter period by the ICO had to do with the General Data Protection Regulation not becoming enforceable until May 25, 2018 (GDPR, 2016). Therefore the number of 9.4 million may well be too low. Especially since Ticketmaster itself stated that customers that may have been affected are those who purchased tickets between September 2017 and June 23, 2018 (Ticketmaster, 2018). Based on above findings, we will assume a number of 10 million affected customers.

The attack was attributed by information security researchers to an organization named Magecart, which was already known for stealing from numerous organizations the personally identifiable (financial) information of their customers (Klijnsma, 2018; Kolesnikov & Parashar, 2018). The group targets websites with an insecure payment infrastructure (Perhar, 2018). Basically, the used method is the digital variant of card skimming, and by modifying the JavaScript a virtual card skimming device was implemented (Bosnell, 2018).

Inbenta provided Ticketmaster with a chatbot intended for answering user questions (Information Commissioner's Office, 2020c). The chatbot software was customized by Inbenta at Ticketmaster's request to meet its requirements (Schwartz, 2018e). Ticketmaster went on and implemented the chatbot on various pages of its website, also directly applying the script to its payment page (Information Commissioner's Office, 2020c). The attackers found the script and modified it in order to extract payment data of Ticketmaster customers making their payments (Schwartz, 2018e).

The chatbot's JavaScript was hosted on Inbenta's server (Information Commissioner's Office, 2020c). Inbenta itself was compromised after Magecart exploited multiple web server upload vulnerabilities and malicious code was injected into the JavaScript for the chatbot (Whittaker, 2018b). In that way, Magecart used Inbenta's software to load malicious JavaScript on various Ticketmaster websites (Kolesnikov & Parashar, 2018). The malicious code was designed to collect user-inputted data and to send that back to the attackers (Information Commissioner's Office, 2020c). This also applied to payments, since Ticketmaster included the chatbot on its payment page (Information Commissioner's Office, 2020c). Eventually, once the customers hit the button to submit their payment, the entered personal and payment data were extracted (i.e. copied) from specific form fields and sent to Magecart's server (Information Commissioner's Office, 2020c; Klijnsma, 2018). Ticketmaster stated that as soon as they discovered the malicious software, they disabled Inbenta's software on all their web pages (Ticketmaster, 2018).

Although Magecart displayed sophistication in this (and its other attacks), certainly some configuration vulnerabilities were present. Ticketmaster should have applied the Payment Card Industry Data Security Standard when implementing the Inbenta chatbot (Information Commissioner's Office, 2020c). Inbenta on its part should have assured that the Chatbot JavaScript was free from malicious code during operation, like it also was contractually obliged to (Information Commissioner's Office, 2020c). Ticketmaster and Inbenta could easily have done more to prevent the malicious code from being injected beforehand or from being loaded during payment processes (Information Commissioner's Office, 2020b, 2020c). The ICO did not fine Ticketmaster just like that for nothing (Information Commissioner's Office, 2020b). This means Ticketmaster is held responsible and liable, which ultimately resulted from poor security. Inbenta is not without blame either. We consider this data breach to be rooted in security misconfiguration.

Parameters	Values for Ticketmaster
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Hacking Malicious Software Poor or No Security
Breach Actor	Malicious Outsider
Organization Sector	Retail
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Website Web Server
Data Responsibility	Data Controller
Data Subject	Customers
Data Amount (Records)	10.000.000
Data Type	PII PIFI
Data Sensitivity	Internal
Data State	DIU
Data Necessity	Necessary
Data Breach Detection	External Non-Auto Inadvertent
Breach Period (Days)	282
Time between Breach and Detection	Long
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 51 Summarized overview of identified values for Ticketmaster

5.2.41. United States Postal Service (2018)

The United States Postal Service (USPS) is an independent federal agency that provides postal services in the United States (United States Postal Service, n.d.). Late 2018, the USPS reported it had more than half a million employees (United States Postal Service, n.d.).

Around mid-November 2018, a researcher, who asked to stay anonymous, contacted investigative journalist and IT security blogger Brian Krebs about a security issue at the USPS (Krebs, 2018d). More than a year earlier, that researcher discovered a security vulnerability (Krebs, 2018d). In a blog post, Krebs (2018d) reports the following about that vulnerability: *"a security weakness that allowed anyone who has an account at usps.com to view account details for some 60 million other users, and in some cases to modify account details on their behalf."*

The researcher stated that he had informed USPS about his discovery when he discovered this more than a year earlier, but that he did not receive any response during that time (Krebs, 2018d). Krebs investigated the situation and was able to confirm the researcher's findings (Krebs, 2018d). At that point, Krebs contacted the USPS which only then quickly addressed and resolved the security incident (Krebs, 2018d).

The security incident involved an application programming interface (API), basically a tool which allows different parts of an online application to communicate with each other, for instance defining how web pages and databases should interact (IBM Cloud Education, 2020; Krebs, 2018d). APIs are generally used to streamline business operations and to create solutions that provide better customer experience in an efficient and agile manner (Melo, 2019; Townsend, 2018). By enabling automated data sharing mechanisms in a scalable manner businesses are able to deliver more qualitative services and to stay relevant and modern (Kovacs, 2018a). But when implementing APIs, their security implications should be taken into account also and maybe even be one of the most important factors, if not the most important one (Kovacs, 2018a; Townsend, 2018).

The involved API in this case was attached to one of USPS services which provided business clients with access to near real-time tracking information with regard to the delivery of their mails and packages (Krebs, 2018d). However, the security vulnerability in the API allowed any user logged into usps.com to query a USPS database for account information pertaining to other users, consisting of usernames, user IDs, account numbers, email addresses, street addresses, phone numbers, users authorized to each account, mailing campaign data and other information (Krebs, 2018d).

A significant number of the involved API's functionalities even accepted wildcards as search parameters (Krebs, 2018d). The usage of wildcards makes it able to maximize the search results since the wildcards are used in search terms to represent different variations of one or more characters (Open Semantic Search, n.d.). In that way, without searching for specific terms, all records for a given data set would be returned, making it even easier to extract data without the need for more complex hacking capabilities (Krebs, 2018d).

The security vulnerability was caused by inadequate implementation of access control in the API, basically not sufficiently protecting the API through authentication and authorization (Krebs, 2018d). Being logged into usps.com apparently was enough to gain unauthenticated access to a USPS database through that API (Krebs, 2018d). The API should have been protected by way of authentication and have validated against the logged in account, as well as should have verified that an account making the query had permission to read the requested data (Kovacs, 2018a; Krebs, 2018d; Liao, 2018). However, such controls were not present (Krebs, 2018d; Liao, 2018).

Hence, this security incident is considered the result of poor security (Kovacs, 2018a). Various security experts stated proper access control is an essential part of basic security practices and the lack of adequate authentication and authorization is just poor configuration (Kovacs, 2018a; Krebs, 2018d; Muncaster, 2018d). Based on our findings, we assess this data breach to have security misconfiguration as its root cause.

Parameters	Values for United States Postal Services
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Poor or No Security
Breach Actor	Ethical Outsider
Organization Sector	Government and Military
Organization Size	Large
Organization Headquarters	US-USA-840
Organization Component	Database
Data Responsibility	Data Controller
Data Subject	Customers
Data Amount (Records)	60.000.000
Data Type	PII LI OBI
Data Sensitivity	Internal
Data State	DAR
Data Necessity	Necessary
Data Breach Detection	External Inconclusive Inconclusive
Breach Period (Days)	Inconclusive
Time between Breach and Detection	Inconclusive
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 52 Summarized overview of identified values for United States Postal Service

5.2.42. American Medical Collection Agency (2019)

American Medical Collection Agency (AMCA) was a debt collection agency based in the United States (American Medical Collection Agency, 2019). Apparently, at the end of 2018 AMCA had a staff of only 25 (Krebs, 2019b). AMCA provided billing and debt collection services to organizations in the healthcare sector, like it also did to Quest Diagnostics (Steward & Cavazos, 2019). Actually, AMCA was a subcontractor to revenue cycle management company Optum360, and Optum360 was a Quest Diagnostic contractor (McGee, 2019). Optum360 contracted AMCA to provide its services to Quest Diagnostics (Steward & Cavazos, 2019). In practice, Question Diagnostics executed the medical service requested by doctors or medical organizations, for which it had hired Optum360 to take care of the medical billing, and Optum360 on its part contracted with AMCA to collect on bad accounts (Sussman, 2019).

On February 28, 2019, researchers at Gemini Advisory encountered payment card information being offered for sale in a marketplace on the dark web (Doe, 2019b; Kovacs, 2019c). The offered data included personally identifiable information such as dates of births, Social Security Numbers and physical addresses, which was uncommon (Doe, 2019b). Seldom are these data types sold combined, for the simple reason it almost did not happen that these were stolen combined (Doe, 2019b).

Gemini Advisory further analysed the data and found the data likely was stolen from AMCA's online portal (Doe, 2019b; Kovacs, 2019c). On March 1, 2019, Gemini Advisory reached out to AMCA directly to notify them, but did not receive any response (Doe, 2019b; Kovacs, 2019c). After not receiving any response from AMCA, Gemini Advisory decided to notify federal law enforcement shortly after (Doe, 2019b; Kovacs, 2019c). Reportedly, the latter followed up by contacting AMCA (Doe, 2019b).

By the 10th of May in 2019 it was clear that AMCA knew about this situation, and there were even indications that AMCA had taken down its payment portal already in the beginning of April of 2019 for a period of time (Doe, 2019b, 2019c). in the meantime, AMCA had conducted their own investigation into the matter (McGee, 2019). On May 14, 2019, AMCA notified Optum360 and Quest Diagnostics that there had been unauthorized activity on its payment web page (Ikeda, 2019c; Sussman, 2019). This payment page was used by patients of healthcare companies such as Quest Diagnostics.

The unauthorized party had access to AMCA's systems between August 1, 2018, and March 30, 2019 (Ikeda, 2019c; Kovacs, 2019c; Lindsey, 2019; Sussman, 2019). These systems contained data which AMCA had received from the healthcare companies it provided with its services, but also information that AMCA had collected itself (Sussman, 2019).

AMCA told Optum 360 and Quest Diagnostics that this unauthorized activity resulted in a data breach compromising the personal data of 11.9 million Quest Diagnostics patients which was stored in AMCA's systems (McGee, 2019). AMCA had numerous other healthcare companies than Quest Diagnostics as customers, whose patients' information also was stored on AMCA's servers (Kovacs, 2019f). We will only consider two other healthcare organizations which were the second and third most affected in terms of number of patients, which are LabCorp and Clinical Pathology Laboratories (CPL) with respectively 7.7 and 2.2 million affected patients having their personal information exposed (Kovacs, 2019d, 2019e, 2019f).

It could not be determined precisely which individuals had their personal information compromised, thus there was no choice than to assume that every patient whose information was stored on AMCA's servers was affected (Kovacs, 2019e, 2019f). Considering only Quest Diagnostics, LabCorp and CPL this amounts to 21.8 million. The data on AMCA servers which AMCA had received from these healthcare companies, pertain to all those patients.

The compromised information included names, Social Security Numbers (SSNs), addresses, phone numbers, dates of birth, balance information, treatment provider information, and medical service information (Ikeda, 2019c; Kovacs, 2019c; Lindsey, 2019; McGee, 2019). A part of the 21.8 million had their payment card data or bank account information exposed (Kovacs, 2019f). This exposed data was probably collected by AMCA

itself, i.e., data that could have been stolen after or during collecting them by AMCA (for instance during or after a payment).

The data which included payment card data and banking information likely concerned the data offered for sale found on marketplaces on the dark web. The big data sets which did not include payment card data or banking information likely were the assumed exposed data, which were all data stored on AMCA servers.

AMCA stated they did not know how the unauthorized party gained access (Ikeda, 2019c). We did not find any other explanation for this, nor on how they moved within AMCA's network. With regard to the data extraction for at least part of the data, there were some factors which may suggest how the data extraction or theft could have taken place. In the first place, the statement that there had been unauthorized activity on the AMCA payment web page (Ikeda, 2019c; Kovacs, 2019c; Lindsey, 2019; McGee, 2019; Osborne, 2019). Secondly, the combination of data types of that were exposed. Rarely are dates of birth, SSNs and payment card data offered for sale together (Doe, 2019b). These factors suggest that part of the data were taken by way of a digital variant of skimming, like was used in the data breaches of British Airways and Ticketmaster (Osborne, 2019). Personal and payment card data entered by patients on the AMCA payment page, possibly during a payment process, could have been logged and extracted by the attackers (Ikeda, 2019c).

However, we did not find a reliable and sufficiently detailed explanation from a relevant source on how the attack could have taken place, like for instance in the British Airways data breach. Therefore we cannot establish whether this data breach was the result of a complex attack or a security misconfiguration.

Parameters	Values for American Medical Collection Agency
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Inconclusive
Breach Actor	Malicious Outsider
Organization Sector	Financial and Insurance
Organization Size	Small
Organization Headquarters	US-USA-840
Organization Component	Website Database
Data Responsibility	Data Processor
Data Subject	Patients
Data Amount (Records)	21.800.000
Data Type	PII PIFI PHI
Data Sensitivity	Internal
Data State	Inconclusive
Data Necessity	Mixed
Data Breach Detection	External Automated Intended
Breach Period (Days)	242
Time between Breach and Detection	Long
Time between Detection and Action	Short
Is the Root Cause a Complex Attack or a Security Misconfiguration?	Inconclusive

Table 53 Summarized overview of identified values for American Medical Collection Agency

5.2.43. Desjardins (2019)

Financial services cooperative Desjardins is Canada's largest credit union and one of the world's largest banks (AFP, 2019a; Cimpanu, 2019d; Desjardins, n.d.-b; Zurkus, 2019). In 2019, Desjardins had approximately 46 thousand employees (Desjardins, n.d.-a).

In December of 2018 Desjardins noticed a suspicious transaction during routine monitoring and referred this to the local police authorities (Montpetit, 2019; Shingler, 2019). It took those authorities several months to learn the scope of the scheme (Montpetit, 2019; Shingler, 2019). This made them inform Desjardins in May of 2019 that the personal information of some of its account holding customers had been compromised (Montpetit, 2019; Shingler, 2019). Desjardins launched an internal investigation which was conducted with assistance from the local police authorities (Montpetit, 2019; Shingler, 2019). The data breach was confirmed on June 14, 2019 (AFP, 2019a; Cimpanu, 2019d). On June 20, 2019, Desjardins made public the data breach (AFP, 2019a; Cimpanu, 2019d; Montpetit, 2019). Later, the Office of the Privacy Commissioner of Canada (OPC) and Quebec's Access to Information Commission (AIC) also investigated the data breach (Office of the Privacy Commissioner of Canada, 2020a).

The security incident involved approximately 9.7 million data records containing information collected by Desjardins from customers who purchased or received products from or through Desjardins (Bronskill, 2020; Solomon, 2020). This included the records of all of Desjardin's customers with an active account at that point, an amount of 4.2 million including at least 173 thousand businesses (AFP, 2019a, 2019b; Bronskill, 2020; Cimpanu, 2019d; Coble, 2020; The Canadian Press, 2019). Most of the remainder of data records pertained to people of which their Desjardins account is inactive, and to a few who had no Desjardins account at all (Bronskill, 2020).

The compromised data included first and last names, dates of birth, Canadian social insurance numbers, addresses, phone numbers, email addresses, and information with regard to banking transactions, habits and usage of Desjardins products (Bronskill, 2020; Cimpanu, 2019d; Solomon, 2020). For business clients, the compromised information included business names, business addresses, business phone numbers, business owners' names, and account users' names (Cimpanu, 2019d). Also compromised were the payment card data of 1.8 million card holders (Coble, 2020; The Canadian Press, 2019).

The data breach was the result of the actions of a malicious Desjardins marketing department employee, who was fired on June 14 (AFP, 2019a; Cimpanu, 2019d; Zurkus, 2019). In a period of at least 26 months, between March 2017 and May 2019, this employee stole the sensitive data which were collected by Desjardins during their operations (Bronskill, 2020; Office of the Privacy Commissioner of Canada, 2020b; Solomon, 2020). This data was originally stored in two data warehouses, the credit data warehouse (CDW) and the banking data warehouse (BDW). Employees with sufficient authorization could access all of the data in the CDW. Access to the BDW was segmented according to whether the data was confidential (which included personal data) or non-confidential (Office of the Privacy Commissioner of Canada, 2020b). The malicious employee could not access the confidential data in the BDW (Bronskill, 2020; Office of the Privacy Commissioner of Canada, 2020b).

However, other employees who had the necessary authorizations to access the data warehouses, including the confidential data, would sometimes have confidential data transferred or copied from the both data warehouses to the marketing department's shared directory (Office of the Privacy Commissioner of Canada, 2020b). With regard to the CDW data warehouse, the data transfer to their user folders on the marketing department's shared drive even happened automated and periodically (Office of the Privacy Commissioner of Canada, 2020b). Concerning the BDW, other employees manually copied confidential data from it to folders on a shared directory (Office of the Privacy Commissioner of Canada, 2020b). This shared directory and the concerned folders were accessible to all employees of that marketing department (Office of the Privacy Commissioner of Canada, 2020b). In that way employees who did not have the necessary authorization to access certain data in the warehouses then were enabled to freely access that data on the shared directory (Office of the Privacy Commissioner of Canada, 2020b).

The malicious employee used scripts to compile the data which his colleagues had saved in the directory (Office of the Privacy Commissioner of Canada, 2020b). After compiling, the malicious employee saved that data in his own user folder as well as in another folder in the shared directory (Office of the Privacy Commissioner of Canada, 2020b). With the use of file sharing software the data was then transferred to his work computer (Office of the Privacy Commissioner of Canada, 2020b). From his work computer he placed the data on USB storage devices (Office of the Privacy Commissioner of Canada, 2020b). Reportedly, the malicious employee offered the stolen data for sale and even sold some of them (Office of the Privacy Commissioner of Canada, 2020b; Solomon, 2020).

Having a malicious insider as the one breaching its employer already incorporates the human factor and makes this case a bit more complicated than if the one breaching was an outside attacker (Zurkus, 2019). Staff members are trusted parties, especially when having signed code of conduct attestations and confidentiality agreements like the malicious Desjardins had done (Office of the Privacy Commissioner of Canada, 2020b; Zurkus, 2019). Furthermore, the insider's malicious activity at least partly was technically indistinguishable from that employee's legitimate activities (Zurkus, 2019). Even so, this data breach could have been prevented and detected (Zurkus, 2019).

The OPC judged that Desjardins failed to implement appropriate security safeguards, especially given that Desjardins collected and held a large volume of sensitive personal data (Office of the Privacy Commissioner of Canada, 2020b). We agree with that judgment. Firstly, some of the data should not have been in Desjardins' possession any longer in terms of retention periods (Office of the Privacy Commissioner of Canada, 2020b).

Furthermore, the authorized employees should procedurally and technically have been limited or impeded to transfer data from the warehouses to the shared directory. The malicious employee should not have been able to run scripts to compile data. And the malicious employee should not have been able to use a USB storage device on his working computer, in order to block the transfer of data from the working computer to the USB storage devices. There are technical measures to arrange the above. Hence, based on our analysis we deem that this data breach was preventable and assess that the root cause is security misconfiguration.

Parameters	Values for Desjardins
Data Breach Type	Unauthorized Access Unauthorized Disclosure
Breach Method	Misuse
Breach Actor	Malicious Insider

Organization Sector	Financial and Insurance
Organization Size	Large
Organization Headquarters	CA-CAN-124
Organization Component	Directory
Data Responsibility	Data Controller

Data Subject	Customers
Data Amount (Records)	9.700.000
Data Type	PII PIFI OBI
Data Sensitivity	Confidential
Data State	DAR
Data Necessity	Mixed

Data Breach Detection	Internal Non-Auto Intended
Breach Period (Days)	792
Time between Breach and Detection	Long
Time between Detection and Action	Short

Is the Root Cause a Complex Attack or a Security Misconfiguration?	Security Misconfiguration

Table 54 Summarized overview of identified values for Desjardins

6 Results

6.1. Introduction

In the previous chapter, we applied the assessment framework to each of the selected data breaches. The assessment framework also functioned as a structure for the case descriptions that we have formulated for each case, based on the corresponding case literature. Because of the framework parameters it remained clear what to include in the case description and what to leave out, in such a way that a case description was formulated that it only contains the relevant information with regard to this research. Hence, by doing so, the previous chapter consists of the actual reporting of the results of applying the assessment framework per case, in a textual manner.

In this chapter, we analyse the data with the purpose of quantitatively describing and summarizing the data in a meaningful way. This provides us with basic information about the frequencies of parameter values in our data set. We do this by treating each parameter in its own section. Each section of this chapter shows and discusses the distribution of the concerned parameter's values across the cases from our data set. That distribution is also visually displayed in a bar chart, which is present in each section and indicates the number of occurrences of each parameter value across all cases. Furthermore, we make attempts to illuminate potential relationships between parameters.

6.2. Frequencies of Breach Type Values

The parameter Breach Type was used to indicate what kind of data breach we are dealing with by describing what happened to the compromised data. We identified five possible basic actions. When we add the value Inconclusive for cases where this cannot be determined, we get six values for this parameter. Hence, its six values are the following:

- Destruction
- Loss
- Alteration
- Unauthorized Disclosure
- Unauthorized Access
- Inconclusive

With these values we were able to characterize in what way the data got compromised. These values do not exclude each other, meaning multiple of these values can be present in a single data breach case if the data got compromised in several ways.

We see that in almost all cases the compromise of security involved unauthorized access to data (95%). This means that in almost all cases it became clear that the data breach occurred because an unauthorized party was able to actually access the data.

A significant number of the cases (65%) involved unauthorized disclosure of data, and in those instances, unauthorized disclosure was present together with unauthorized access almost every time. These parties disclosed these data in one way or another, for instance by selling them.

In two cases there was only identified unauthorized disclosure but no unauthorized access, meaning the data was left publicly exposed but access to them could not be established. Destruction, loss and alteration of data only occurred in one case and that was together with unauthorized access. This makes sense since a party needs access first to be able to cause data destruction, loss or alteration. For that matter, that we only identified destruction, loss and alteration in one case, does not necessarily mean it happened only in one case. But we did not find proof that it happened in the other cases.

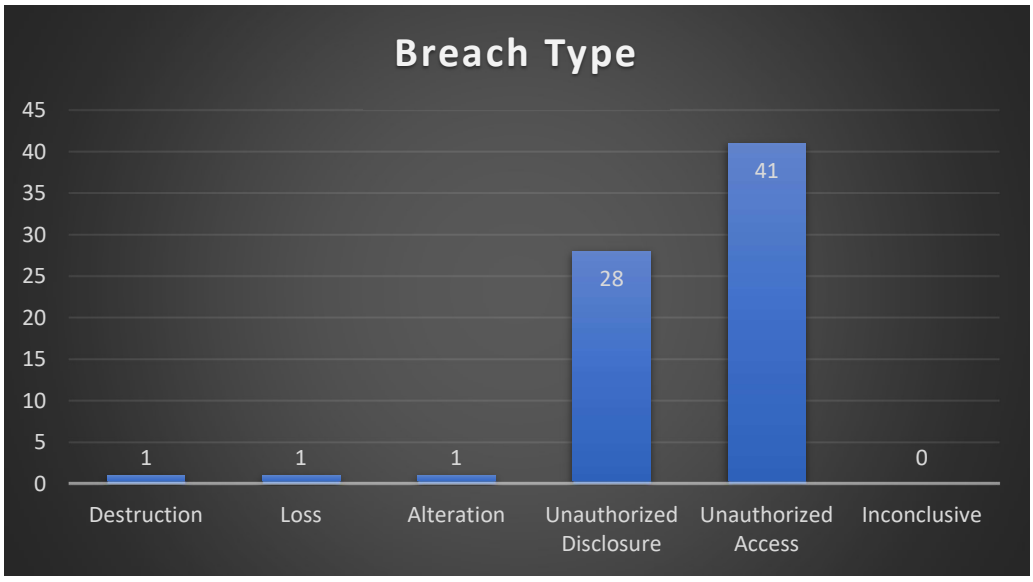


Figure 2 The number of cases in which each value of the parameter Breach Type was identified.

6.3. Frequencies of Breach Method Values

The parameter Breach Method was deployed to specify how or why the data breach happened. It consists of actions or acts which lead to the compromise of data, for which we formulated five general values. When we add the value Inconclusive for the cases where the method cannot be determined, the Breach Method parameter eventually has the following values:

- Hacking
- Malicious Software
- Physical
- Social Engineering
- Misuse
- Poor or No Security
- Inconclusive

These values enabled us to characterize the data breach with the specific methods or techniques that were involved on the way to compromising the data. Or otherwise, to label the data breach with Poor or No Security if that was the case. These values do not exclude each other, meaning a single data breach case can be characterized by multiple of these values.

In more than half of the cases, having poor or no security played a role (51%). In most of them (17 of the 22 cases) we did not identify other methods.

In 28% of the studied cases Malicious Software played a role, and this was always in combination with one or more of the other breach methods.

In 42% of the studied cases Hacking was involved. This occurred when parties gained access to the breached organisation's network, maintained that access and moved through the network in order to find and reach data. In more than half of the cases wherein hacking played a role, other values for the parameter Breach Method were also identified. This is because Hacking according to our definition (see Chapter 4 section 1) only concerns access and movement to the network, and extraction needs to be indicated by other Breach Method values. The cases in which only Hacking could be identified as a value for Breach Method, remained unclear as to how the extraction took place.

In 14% of the cases Social Engineering was used in the data breach process, but every time in combination with other breach methods. This makes sense since Social Engineering is applied to gain access to networks, but for extracting the data other methods are necessary. And in these cases, it was possible to identify Breach Method values which cover data extraction.

This is in contrast with the breach method Hacking, where one would expect the same. But we did not find breach methods explaining the extraction of data for each case in which Hacking played a role.

In only one case Misuse was the breach method. In 9% of the studied cases there could not be identified a breach method.

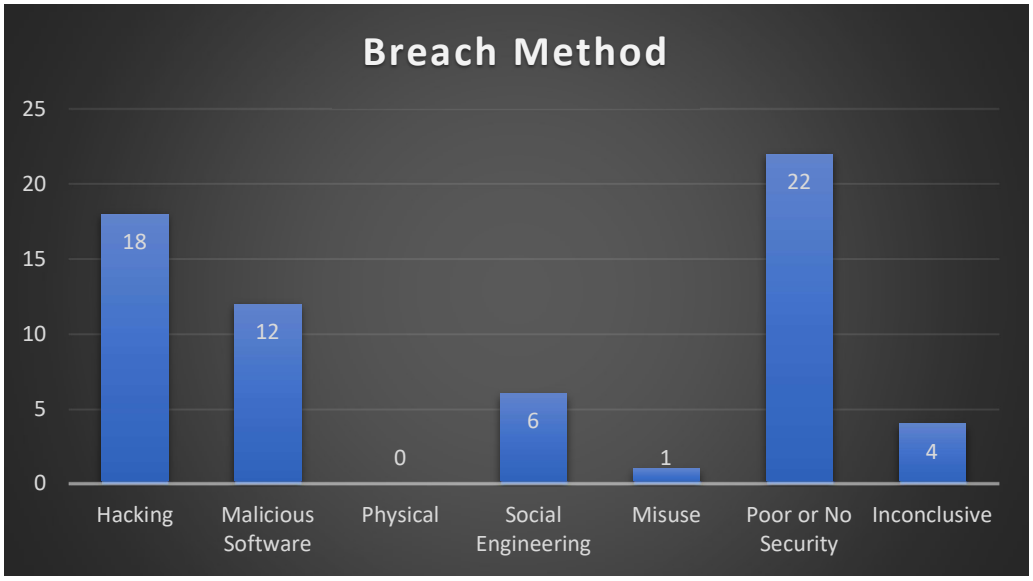


Figure 3 The number of cases in which each value of the parameter Breach Method was identified.

6.4. Frequencies of Breach Actor Values

The parameter Breach Actor was used to identify whether this actor was an insider or outsider in relation to the organization that was breached, as well as to indicate that actor's intent or motive. In order to do this, we used typifications that are represented by the following values:

- Inadvertent Insider
- Malicious Insider
- Ethical Insider
- Inadvertent Outsider
- Malicious Outsider
- Ethical Outsider
- Inconclusive (when the type of breach actor cannot be determined)

These values were only exclusively assigned, meaning the breach actor in a data breach could only be of one of these types.

In 26 of the 43 cases (60%) the breach actor was a malicious outsider. And in 35% of the data breaches the breach was discovered or caused by an ethical outsider. The ethical outsiders mostly concerned security researchers discovering a data breach, but in a few cases it was hacktivism by parties wanting to make a statement with regard to poor security standards.

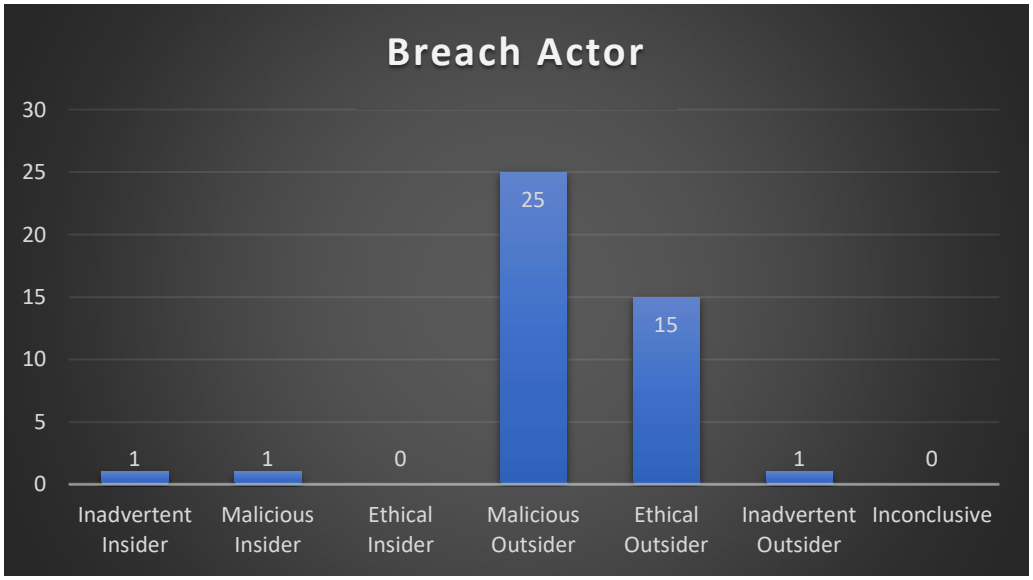


Figure 4 The number of cases in which each value of the parameter Breach Actor was identified.

6.5. Frequencies of Organization Sector Values

The parameter Organization Sector was used to indicate the main sector the affected organization was known to be operating in. The values of this nominal parameter are mutually exclusive and are the following:

- Financial and Insurance
- Retail
- Education
- Government and Military
- Medical and Healthcare
- Technology
- Other
- Inconclusive

The data breaches affected organizations across all of the sectors we defined. Most of these organizations fall in the retail (26%), technology (21%) and the financial (14%) sector. These types often have large user and customer bases, and especially banking organizations can hold a large amount of rich and veracious data. In eight of our cases, the breached organization's sector is labelled as 'Other' (19%).

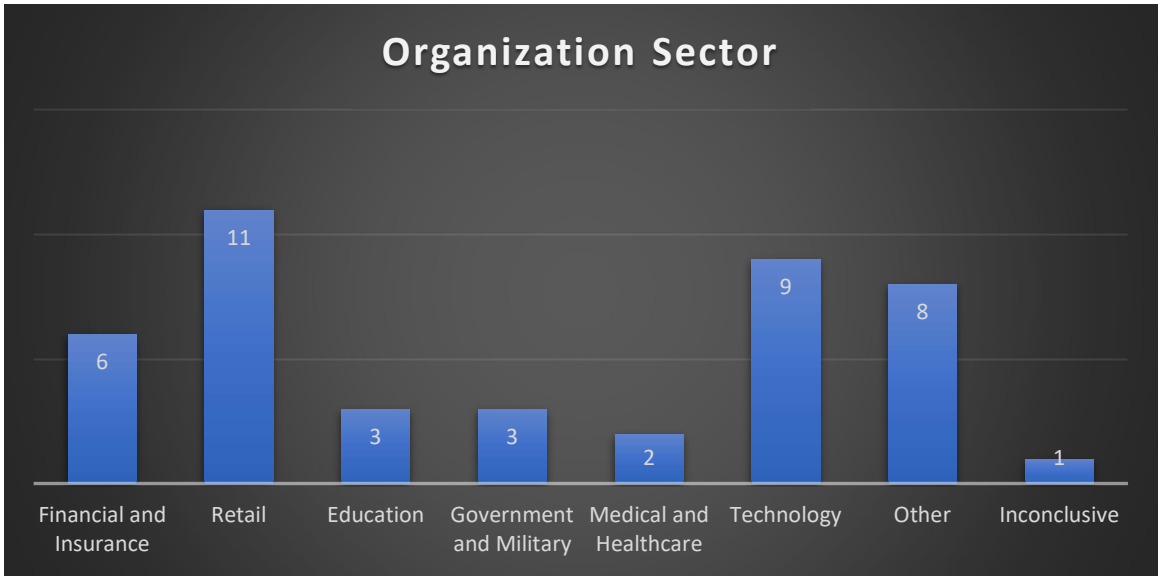


Figure 5 The number of cases in which each value of the parameter Organization Sector was identified.

6.6. Frequencies of Organization Size Values

The parameter Organization Size was used to label each affected organization as either Small, Medium or Large. This parameter is ordinal, and its values are therefore mutually exclusive, meaning that each data breach has just one victim organization and this organization only fits in one size category. Hence it is not able to label an organization with more than one these values. The value Inconclusive is intended for cases where the organization size could not be determined.

Nearly 70% of the organizations from our sample concerned a large organization according to our classification. This mainly has to do with our scope and case selection. We are interested in severe data breaches. The severity in our research is mainly determined by the amount of data which was compromised, and additionally its sensitivity. This meant that for a data breach case to qualify for our case selection, the important criterium was that a large number of people should have been affected by that data breach.

From our assessment it can be seen that if an organization holds a lot of data, which becomes apparent when having a large amount of data becoming compromised during a data breach, this will often concern a large organization. A common scenario is that an organization will hold a lot of data if it has a large customer or user base. We will elaborate on this in section 6.10.

The two cases where the organization size could not be determined concerned data breaches wherein the specific affected organization could not be identified.

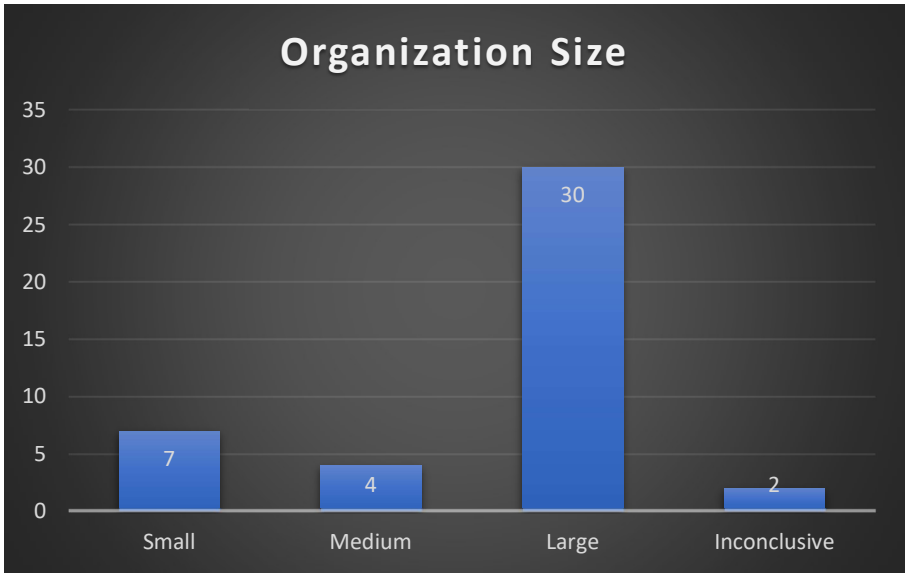


Figure 6 The number of cases in which each value of the parameter Organization Size was identified.

6.7. Frequencies of Organization Headquarters Values

The parameter Organization Headquarters was used to indicate where the affected organization was based. This is a nominal parameter and its values are mutually exclusive. Theoretically, the values we can label an organization in a data breach consist of all the countries in the world. But we did not explicitly list all these values, as this would give an unnecessary long list of values of which most eventually would not be used. Hence, in each case we identified the affected organization and determined where its headquarters is based. As the labels we used the ISO 3166 country codes. In our study sample, the following values are present:

- AU-AUS-036 (Australia)
- BG-BGR-100 (Bulgaria)
- BR-BRA-076 (Brazil)
- CA-CAN-124 (Canada)
- CN-CHN-156 (China)
- DE-DEU-276 (Germany)
- GB-GBR-826 (United Kingdom)
- HK-HKG-344 (Hong Kong)
- IL-ISR-376 (Israel)
- IN-IND-356 (India)
- JP-JPN-392 (Japan)
- KR-KOR-410 (South Korea)
- NO-NOR-578 (Norway)
- SG-SGP-702 (Singapore)
- US-USA-840 (United States of America)

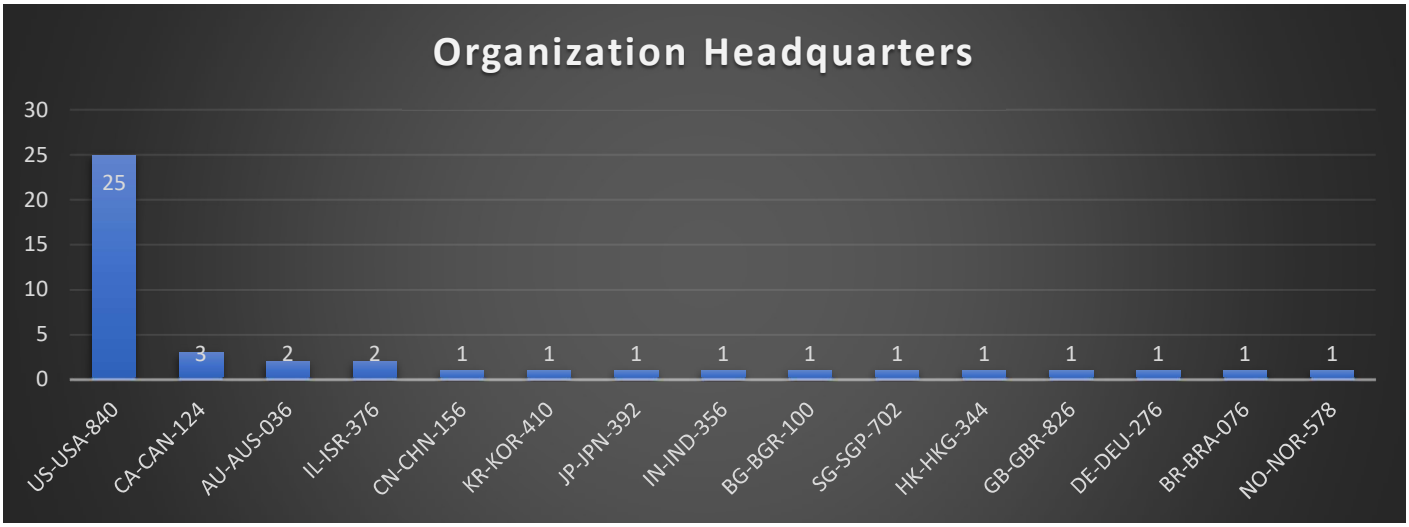


Figure 7 The number of cases in which each value of the parameter Organization Headquarters was identified.

Most of the affected organizations in our study sample (58%) are based in the United States of America (USA). The other 48% is quite evenly spread across other countries.

6.8. Frequencies of Organization Component Values

The parameter Organization Component was utilized to identify the specific component(s) from where the data were taken or where the data were on the moment they got exposed. In order to do so, we applied the following values:

- Cloud Storage
- Database
- Device
- Directory
- Mail Server
- Mailbox
- Web Server
- Website
- Other
- Inconclusive

During a data breach, it is possible that (different) data gets compromised at different parts of an organization's network and within different components. Therefore, the values for this parameter are not mutually exclusive.

In 51% of the cases, a database was involved as a compromised component concerned. This makes sense since databases are used to store large amounts of data.

Cloud storage has become more popular over the years. However, cloud storage security remained insufficient at times, vulnerabilities became known and in some cases even got exploited. Eventually, cloud storage was involved in 21% of the studied cases.

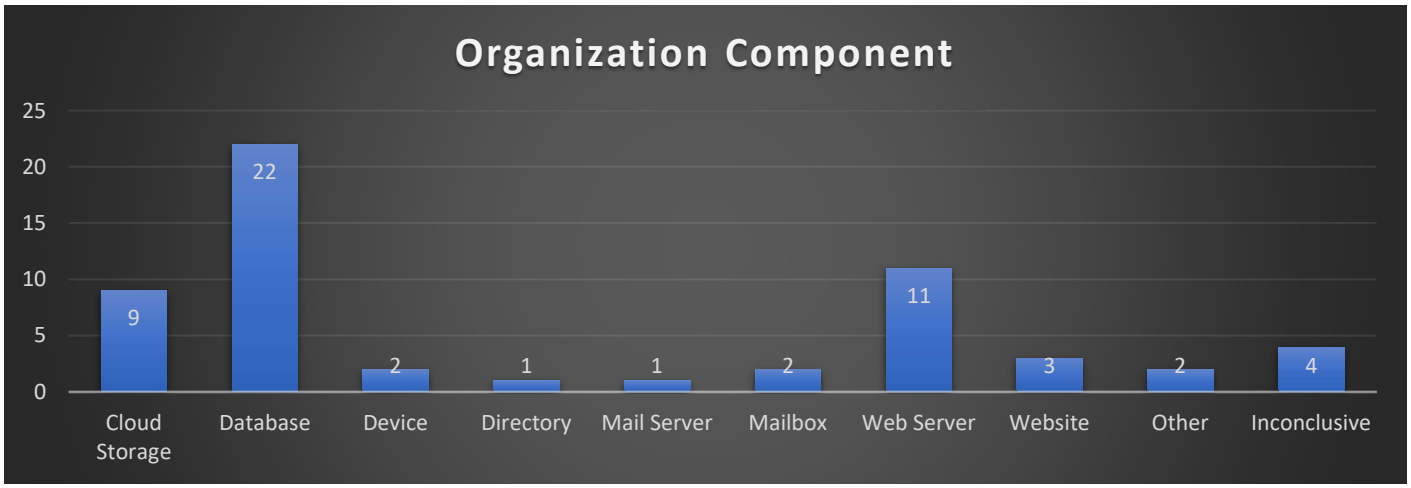


Figure 8 The number of cases in which each value of the parameter Organization Component was identified.

In 26% of the cases a web server was involved as a compromised component. This also makes sense, since a web server serves web pages and files requested over the Internet, and even can storage data itself if it is a dynamic server.

6.9. Frequencies of Data Responsibility Values

The parameter Data Responsibility was used to indicate what kind of responsibility the breached organization had over the data, in terms of whether the affected organization owned the data itself or was provided with the data by another organization on behalf of which the affected organization processed that data. This nominal parameter has the following mutually exclusive values:

- Data Controller
- Data Processor
- Inconclusive

In 91% of the cases the organization was the data controller. In one instance it could not be determined whether the breached organization was the data controller or data processor, because in that case the breached organization could not be identified at all. This does not mean it is none of both; on the contrary, it has to be one of both, only we were not able to determine which one.

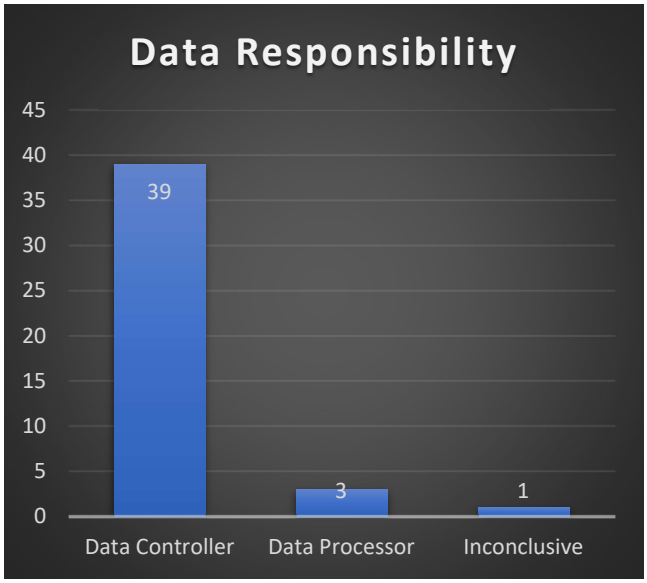


Figure 9 The number of cases in which each value of the parameter Data Responsibility was identified.

6.10. Frequencies of Data Subject Values

The parameter Data Subject was used to specify the victim or victim organization respectively to whom or to which the compromised data pertains. This could be identifiable natural persons being compromised in their personal information. But it could also be organizations having their business or government information compromised. This categorical parameter has non-mutually exclusive values, meaning a data breach can have multiple victims, i.e., multiple different data subjects. These are the values:

- Breachee
- Citizens
- Customers
- Employees
- Members
- Patients
- Students
- Trusted Business Partners
- Other
- Inconclusive

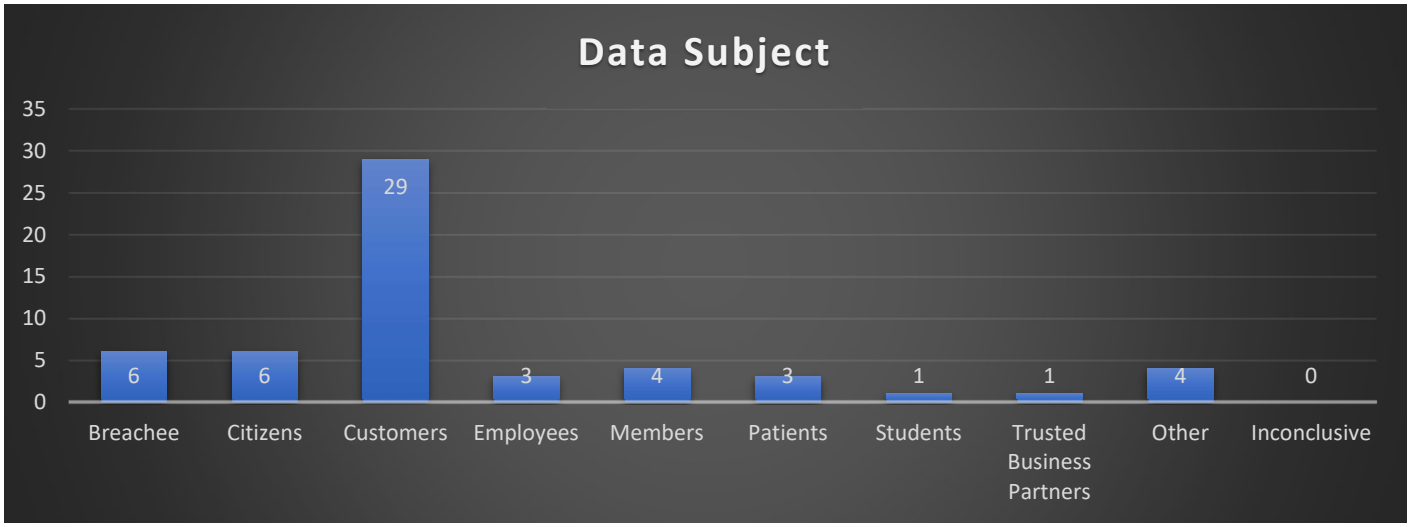


Figure 10 The number of cases in which each value of the parameter Data Subject was identified.

Customers are involved as affected data subjects in 67% of the cases, having their personal information stolen or exposed. Businesses gather, process, and keep a lot of customer data as part of the business-customer relationship, and this generally increases as the number of customers and the business grow. The business-customer relationship is often at least transactional. Generally, a transaction comprises an agreement between a provider and a customer to exchange goods, services or financial assets. The customer might have been an individual who had to register first before being able to receive goods or services, or someone who paid for goods or services using a payment card, etc. But other scenarios are also possible.

A lot of organizations operate in commercial sectors and as such are businesses with customers. This was already touched upon in the section treating the parameter Organization Sector (6.5). Hence, when these businesses leave data exposed, this will almost always involve customer data and then inherently a large amount of data is compromised. Also, in the case of an attack, attackers will often focus on organizations with a lot of customer data, especially if there are to be found vulnerabilities there.

Data belonging to citizens is compromised in 14% of the cases, while data belonging to patients and organization employees are involved in 7% of the cases. In 14% of the cases the breached organization itself, the breachee, had its own data compromised. These data concern proprietary information.

This parameter also showed that in 42 of the 43 cases it could be determined that natural persons are involved as data subjects.

6.11. Frequencies of Data Amount Values

The parameter Data Amount has been used to categorize the data breaches in terms of the number of data records that have been compromised, which provides an indication of the severity of the data breach. The mutually exclusive values of this ordinal parameter are shown in table 55 below.

R represents the number of records per million	Range explanation
R < 0.1	Less than 100 thousand records
0.1 ≤ R < 1	The number of records is equal or more than 100 thousand and less than 1 million records
1 ≤ R < 10	The number of records is equal or more than 1 million and less than 10 million records
R ≥ 10	Equal or more than 10 million records
Inconclusive	The number of records cannot be determined and therefore a range value cannot be selected

Table 55 Values for parameter Data Amount

We have found that over half of the cases (53%) have more than 10 million breached records. In 30% of the cases between 1 and 10 million data records were breached. This means that in 83% of the cases a very significant number of records are breached. In 12% of the cases, there were involved between 100 thousand and 1 million data records. Only one case had less than 100 thousand breached records. And in another case, we could not determine the number of involved records.

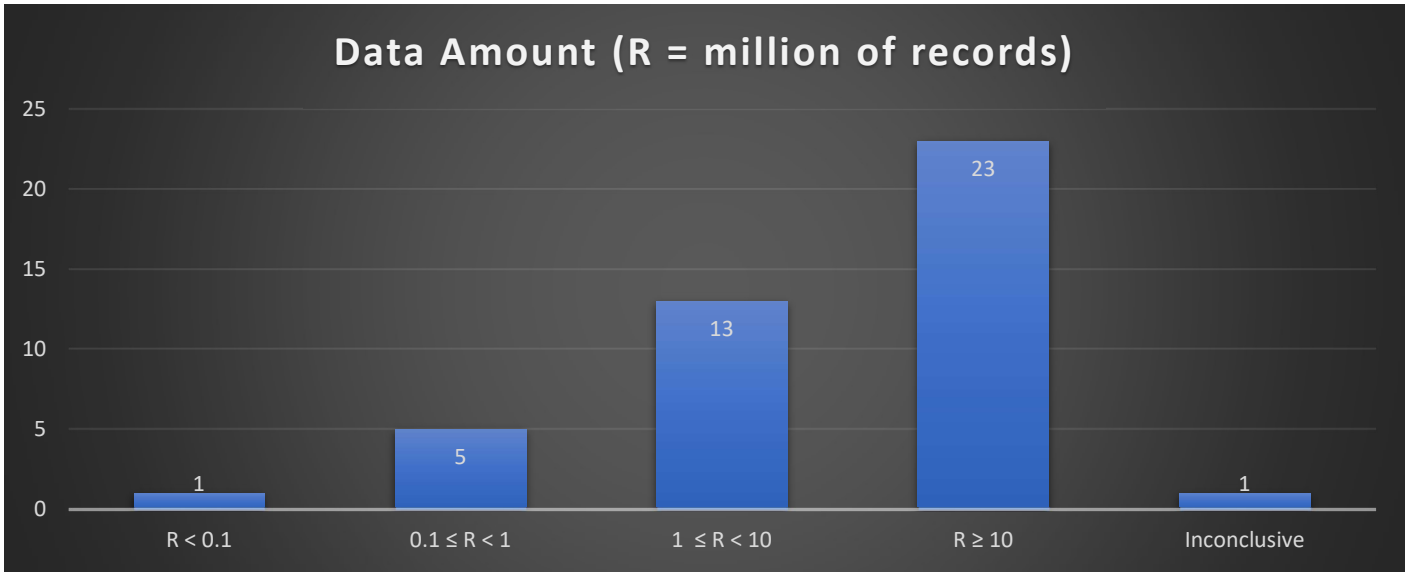


Figure 11 The number of cases in which each value of the parameter Data Amount was identified.

In previous section we saw that in almost all the studied cases natural persons were involved as data subjects. Together with the findings above, we can establish that nearly all cases concern severe data breaches. This is in accordance with what was intended by utilizing our case selection criteria, namely, to only select severe data breaches.

6.12. Frequencies of Data Type Values

The parameter Data Type has been used to identify the types of data involved in data breaches. As we saw in section 4.3.3, there are two main types, namely personal information and proprietary information, each consisting of more specific data types. This parameter's values are:

- Personal Information values (labels between brackets)
  - Personally Identifiable Information (PII)
  - Personally Identifiable Financial Information (PIFI)
  - Protected Health Information (PHI)
  - Login Information (LI)
  - Biometric Data (BD)
- Proprietary Information values
  - Trade Secrets & other Intellectual Property (TSIP)
  - Other (Business) Information (OBI)
  - Classified Information (CI)
- Inconclusive value
  - Inconclusive (INC)

In case of a data breach, it is possible that multiple of these data types may have been compromised. Thus, these values are not mutually exclusive.

Our research shows that 98% of the cases involved PII, in 53% of the cases PIFI was compromised, and LI was compromised in 33% of the cases. PHI and BD were involved in a few cases (9% and 2%). These types fall under personal information according to our categorization.

These findings correspond with what we found in section 6.10. In that section we already saw that in 42 of the 43 cases natural persons were involved as data subjects, for which we also offered an explanation.

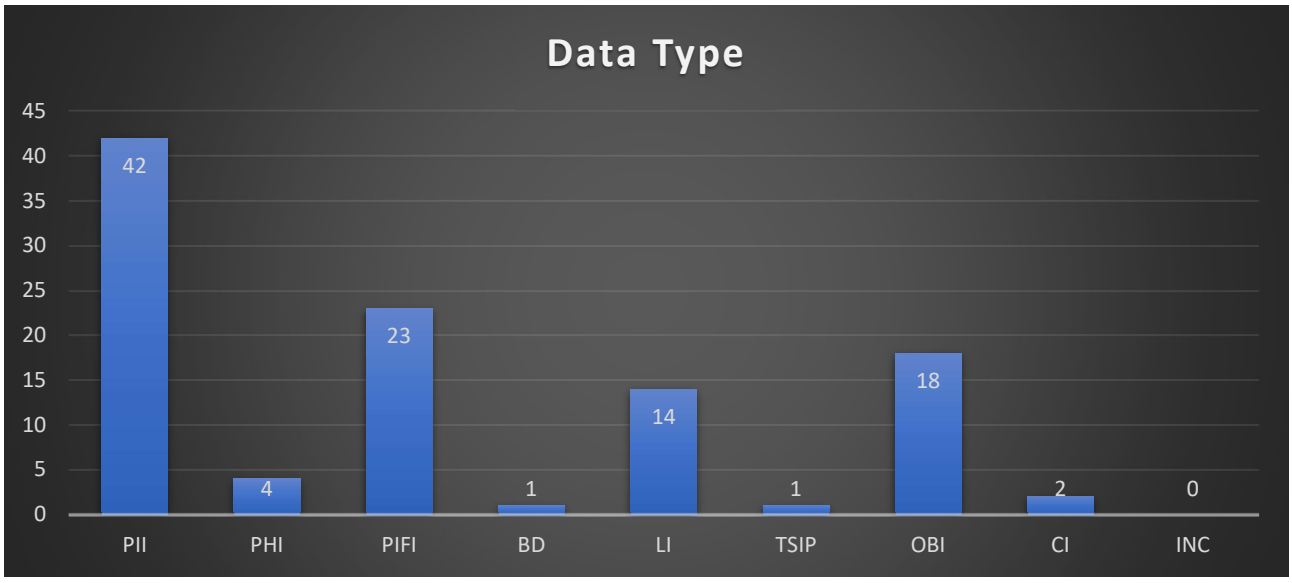


Figure 12 The number of cases in which each value of the parameter Data Type was identified.

We found that business related information (OBI and TSIP) was compromised in 44% of the cases. In one case, this even concerned intellectual property. The presence of business information as being compromised

often happened because this type of information was stored close or even together with personal information. Classified information was involved in only two cases.

Going through the case literature covering the data breaches, it became apparent that organizations are not always completely honest and transparent about what happened, how it took place, and which data were compromised in what manner. Even when there is societal and legal pressure, which is especially present when personal information is compromised, organizations try to downplay the data compromise with the purpose of coming out of the data breach and its aftermath with as little damage and implications as possible.

Regarding proprietary information, this kind of behaviour might also be strategical. Disclosing the compromise of valuable proprietary information may influence the organization's value. For instance, making this public could potentially make stock prices go down.

Also, if intellectual property is stolen, the deprived enterprise does not want to alert competitors on the potential availability of its valuable proprietary information. These competitors might scour digital underground marketplaces in search of that information (if they are not the ones that ordered the data theft in the first place).

6.13. Frequencies of Data State Values

The parameter Data State has been used to indicate for each data breach the state the data were in. Its values are the following:

- Data-at-rest (DAR)
- Data-in-motion (DIM)
- Data-in-use (DIU)
- Inconclusive

These values are not mutually exclusive, meaning a data breach can result in the compromise of a variety of data in multiple states. This could only be determined for one case though, the Target case. According to our assessment, data-at-rest and data-in-use were compromised in that case.

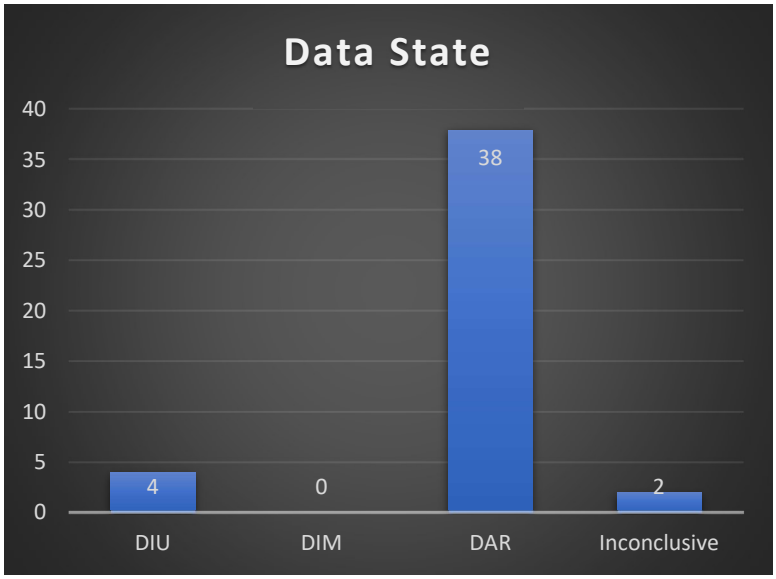


Figure 13 The number of cases in which each value of the parameter Data State was identified.

In 37 other cases we determined that the compromised data was at rest also. This comes down to a total of 88% of the cases involving data which became compromised while being at rest. Besides the Target case, 3 other cases involved data-in-use. This comes down to a total of 9% of the cases involving data which became compromised while being in use. In two cases we could not determine the state of (part of) the compromised data and in none of the cases in our sample data-in-motion was compromised.

It can be argued that in case of attacks, attackers more often find data-at-rest to be more valuable than data-in-use or data-in-motion. Although, on the one hand, this preference also depends on the data type and data sensitivity of the data. On the other hand, it also depends on the quality of the security measures that were implemented to secure the data in each state, resulting in different risk profiles for data of each state in different organizations.

In case of data compromise because of exposure, this happened far more often to data-at-rest than to data-in-use. We looked at the 28 cases having Unauthorized Disclosure as Data Breach Type and in 24 of those cases the compromised data were at rest, against 4 cases in which the compromised data were in use (the Target case involved data of both states).

6.15. Frequencies of Data Necessity Values

The parameter Data Necessity has been used to indicate whether the data compromised in a data breach was necessary for the organization holding and/or utilizing these data for the purpose of this organization's operations and offerings. This organization concerns the breached organization, which is either the data controller or data processor. The values for Data Necessity are the following:

- Necessary
- Unnecessary
- Mixed
- Inconclusive

In 77% of the cases, we assessed that the compromised data was justifiably held and processed by the breached organization for the purpose of their operations and offerings. In 23% of the cases the compromised consisted of both necessary and unnecessary data.

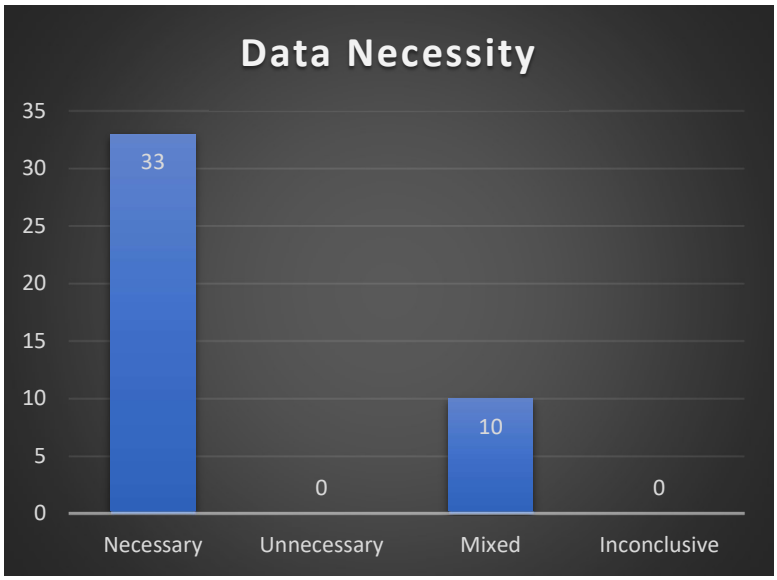


Figure 14 The number of cases in which each value of the parameter Data Necessity was identified.

There are no data breaches in our sample where all the compromised data were unnecessary. Hence, there are no organizations of which we can state that they would not have been breached if they did not hold or process that unnecessary data.

The organizations where the data breach resulted in the compromise of both necessary and unnecessary data would likely still have been breached even if the unnecessary data were not there. But the impact probably would have been less.

6.16. Frequencies of Data Breach Detection Values

Data Breach Detection has been used to describe how the breach was detected by means of three parameters, namely Detection Source, Detection Method and Detection Intent. These parameters and their mutually exclusive values (and labels) are the following:

- Detection Source
  - Internal (Internal)
  - External (External)
  - Inconclusive (Inconclusive_DS)
- Detection Method
  - Automated (Automated)
  - Manual (Non-Auto)
  - Inconclusive (Inconclusive_DM)
- Detection Intent
  - Intended (Intended)
  - Inadvertent (Inadvertent)
  - Inconclusive (Inconclusive_DI)

In 70% of the cases, the breach was detected by parties or systems external to the breached organization. Only less than one third was detected internally.

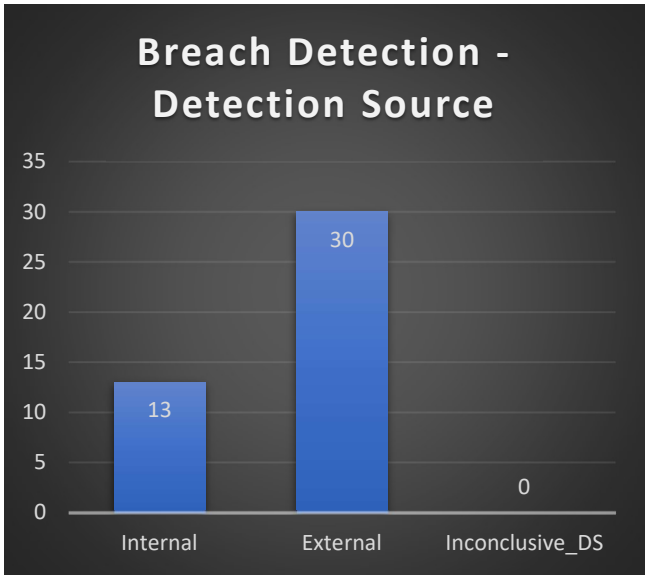


Figure 15 The number of cases in which each value of the parameter Detection Source was identified.

In 56% of the cases the breach was not detected in an automated way according to our definition. Only 23% of the studied data breaches was detected in an automated way. In 21% of the cases it could not be determined whether the detection happened by automation or manually.

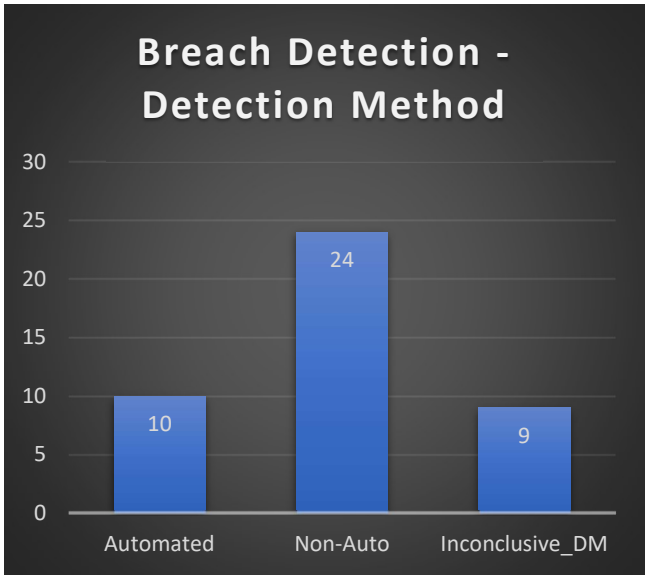


Figure 16 The number of cases in which each value of the parameter Detection Method was identified.

In 58% of the cases the data breach was detected intentionally by its detector, against 21% of the cases where the data breach was encountered unintentionally.

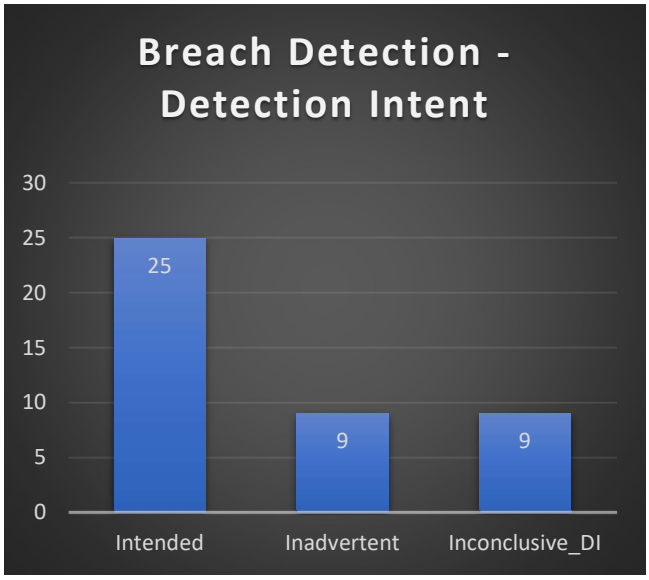


Figure 17 The number of cases in which each value of the parameter Detection Intent was identified.

6.17. Frequencies of TBBD Values

The parameter TBBD (Time between Breach and Detection) has been used to determine the time passed between the beginning of a data breach and the moment that breach gets noticed. This parameter is formulated as a categorical variable with an ordinal scale. Its mutually exclusive values are the following:

- Short (≤ 30 days)
- Medium (30 < days ≤ 100 days)
- Long (≥ 100 days)
- Inconclusive

In 42% of the cases the breach was not noticed until after more than 100 days, which may be considered meaningful. In only 19% the breach was detected within a month (i.e., 30 days). In a substantial 33% of the cases TBBD could not be determined, because the first moment of the breach could not be identified, the moment of detection could not be determined, or both.

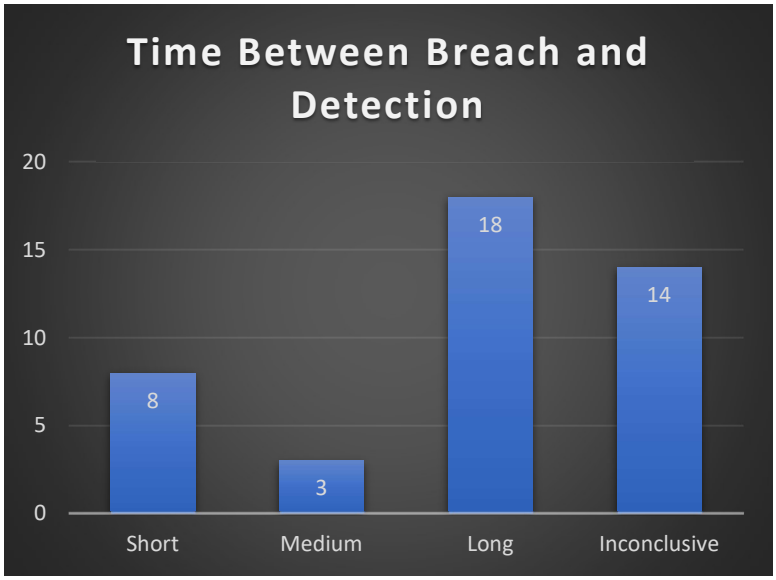


Figure 18 The number of cases in which each value of the parameter TBBD was identified.

6.18. Frequencies of Root Cause Values

The parameter Root Cause has been used to establish the root cause of data breaches, based on the selected case literature and our assessment criteria. The mutually exclusive categorical values this parameter can take are the following:

- Complex Attack
- Security Misconfiguration
- Inconclusive

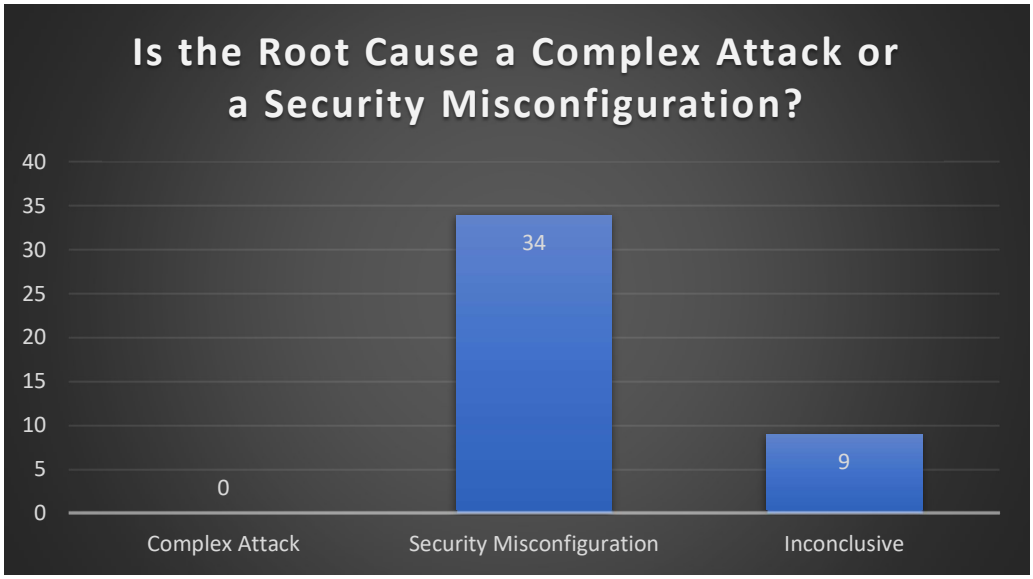


Figure 19 The number of cases in which each value of the parameter Root Cause was assessed.

We found that in 79% of the cases security misconfiguration was present as the root cause. We did not find a complex attack to be the root cause in any of the cases. In 21% of the cases we were not able to determine the root cause.

7 Discussion and Limitations

7.1. Discussion

In the past decade, numerous data breaches have occurred globally. On the other hand, because of data breaches' negative implications, organizations want to prevent data breaches from occurring within their perimeters and affecting data they are responsible for. Tobias Fiebig (2017) and Dietrich et al. (2018) claimed that data breaches frequently happen because of security misconfigurations. This would mean that organizations could easily prevent such simple errors. Nevertheless, to our knowledge, no structured research had been conducted earlier which supports their claim. The results of this research indicate that severe data breaches mostly have security misconfigurations as the root cause.

Based on the duality as described in Dietrich et al. (2018), which implicates that the root cause of a security incident is either a complex attack or a security misconfiguration, we at least expected to identify a complex attack as the root cause in some cases. But in none of the cases of the current study was the root cause found to be a complex attack. However, this does not necessarily mean there were no attacks with some level of sophistication identified in some of the cases. It only means that a complex attack was not the identified root cause in any of the cases; either a security misconfiguration was identified as the root cause or the root cause could not be determined.

In over half of the cases, having poor or no security played a role as the way in which the data breach could happen. Apparently, a lot of organizations do not have proper security in place yet. And in all the cases in which we identified poor or no security, the root cause was security misconfiguration. We can relate this to the findings by Dietrich et al. (2018), who found in their research that almost all participants in their survey encountered security misconfigurations. Although they found that these did not always lead to incidents, our study indicates that poor security and security misconfigurations are correlated.

A significant number of cases involve hacking and malicious software. This indicates there still is room for improvement in the detection and blocking of malicious software and unauthorized presences within organizations. Most of these cases still had security misconfiguration as a root cause. This could mean that malicious software and hacking were enabled by security misconfigurations, but it could also mean that hacking and malicious software enable attackers to find and exploit vulnerabilities that are the result of security misconfigurations. Better protection from malicious software and unauthorized access therefore lowers the exploitability of security misconfigurations and overall will contribute to preventing data breaches. But the technical tools to achieve this were already readily available (Tobias Fiebig, 2017). And as our research shows, still numerous data breaches are caused by security misconfigurations and involve poor security, hacking or malware. Hence, we agree with Dietrich et al. (2018) that mitigation strategies should focus more on the institutional and management domain. These should also include policies for increasing security awareness for the purpose of making people less susceptible to social engineering. Social engineering was involved in a small part of the cases, but those cases all had security misconfiguration as the root cause. Combating social engineering will therefore also result in security misconfigurations leading less often to incidents, since then they will be exploited less easily or will less often lead to inadvertent exposure.

Over half of the cases have as the breach actor an outsider with malicious intentions. This is what we expected based on literature (Widup et al., 2018). This overrepresentation may be related to the availability of cases in the pool we selected cases from. Often breached organizations are required by law to make public a suffered data breach, which then likely is to be followed by societal commotion and news coverage. In the cases of being breached by a malicious outsider, the societal commotion and coverage is even more, putting pressure on breached organizations to conduct investigations and publicly share the results. Because of this, we

conjecture that such cases will be present in the cases pool more often and have more varied literature available on them, which eventually made such cases more likely to be selected. Nonetheless, we deem these results are still valid and useful. Because our main goal is to find out the extent to which root causes are security misconfigurations. So indeed, for that purpose it is even better to have a bigger, more varied, more detailed, and more reliable package of case literature at our disposal.

We assume there are similar reasons for why more than one third of the cases involved an ethical outsider. This breach actor mainly concerns a security researcher of some sort, who scours the Internet searching for security vulnerabilities and because of that often ends up unauthorizedly accessing networks. Such security researchers sometimes want to make public what they have discovered, especially when a breached organization does not respond to or act upon their notification, or in the case they do not know the party responsible for the compromised data and want to find it by drawing its attention. This again could potentially lead to societal commotion and extended coverage of the data breach.

Since ethical outsiders discover a significant part of the data breaches, these actors should not be worked against by ignoring them or trying to quiet them, which in reality happens (e.g., the Chtrxbox case in section 5.2.25.). Organizations do not like to admit data breaches because of legal, reputational, and financial implications. But instead, organizations would do better by cherishing ethical outsiders by incentivizing them to find vulnerabilities and discover the security misconfigurations causing them. This will support organizations in mitigating data breaches or preventing them altogether.

Not once did we identify an ethical insider as the breach actor. Potentially, this could mean ethical insiders, such as whistleblowers, are silenced and the issue is solved internally without disclosing it, remaining unknown to the public and the data subjects. But people need to know if their personal information has been compromised. To achieve this, one could think of legally imposed thorough audits and placing the disclosure responsibility with governments. One could also think of the incentivization of open incident investigations, such as was proposed by T. Fiebig (2020).

Because we selected cases where the data breach resulted in the breach of large amounts of data, the involved breached organizations were more likely to be large, which is reflected by what we found. Namely, more than two-thirds of the breached organizations are large according to our size criteria. This is in agreement with the findings in Heidt, Gerlach, and Buxmann (2019). They state that even when considering the security leeway of small organizations in relation to large organizations, which exists because of lagging IT security investments, organizational IT security research tends to neglect small and medium sized organizations and focus on larger organizations (Heidt et al., 2019). That is what also happened in our research; by selecting large data breaches, we were more likely to focus on larger organizations.

Most of the cases in our study involving large organizations were found to have been caused by security misconfiguration. But for the relatively lower number of cases involving medium and small organizations, this is also the case. It should be taken into account that size affects organizational structure (Ahmady, Mehrpour, & Nikooravesh, 2016). Furthermore, complexity of the IT infrastructure is likely to increase along with increase in organization size (W. H. Baker & Wallace, 2007). Therefore size also makes a difference when it comes to implementing IT security (Chang & Ho, 2006). This implies the prevention and treatment of security misconfigurations is likely to be different for different size of organizations. Hence, our findings and recommendations are only valid for large organizations.

Almost two-thirds of the organizations in our study are from the Retail, Technology or Financial & Insurance sector, while relatively few are from the Government & Military, Education or Medical & Healthcare sector. This is not reflected by the findings of Widup et al. (2018), in which the breached organizations belonged to the government and healthcare sector much more often than to the retail and financial sector. We attribute this difference to our source for data breaches, selection bias and more importantly, the much smaller sample size.

What we did notice is that organizations from certain sectors during certain periods in the past were more likely to suffer from specific vulnerabilities. Vulnerabilities that were typical for such types of organization in

that specific period. For instance, vulnerabilities that were to be found in the e-commerce facilities of retail organizations a few years ago (Klijnsma et al., 2018; Kolesnikov & Parashar, 2018), of which our study showed they were caused by security misconfigurations. In other periods, other types of organization suffered more from other vulnerabilities. This affects the availability of such cases in our cases pool as well as selection bias. Furthermore, the distribution of different types of organizations in the cases pool is also influenced by dissimilar disclosure regulation, which can differ per sector, but also per involved data and region. However, we consider this only indirectly relevant and still deem the results valid for the main research question of this research. An organization's sector does not directly influence the presence of security misconfigurations or their prevention. Organization size and the data characteristics are the more directly important factors with regard to the presence and prevention of security misconfigurations, since those have a direct influence on the configuration and its effectiveness.

Organizations based in the United States of America (US) are overrepresented in our data set with over half of the cases involving them. Even though we have a relatively small sample size, the distribution between US and non-US organization appears to be out of proportion and is remarkable at the least. This can potentially be attributed to the legislation in the US, requiring breached organizations to disclose the incidents according to certain criteria. Subsequently, these incidents easily become widely reported by the extensive media sector in the US, mostly catering on sentiments of society and inciting public outrage. This also causes security researchers of various sorts to analyze and describe the data breach out of scientific or profit motives. The result is that various coverages on a data breach case become readily available.

In Europe there also exist data breach laws, notably the European Union's General Data Protection Regulation (GDPR) and the United Kingdom's Data Protection Act. Still, we only had very few cases that involve breached organizations from European countries (United Kingdom included). We could argue that US organizations are more often the target of attackers. This could be related to the fact that the US contains a large number of very large organizations. For instance, in 2020 at least a thousand US companies had more than 1.9 billion in revenue, and 121 of the Global 500 were US companies (Fortune, 2021a, 2021b). Combined with the large number of US inhabitants, which can be related to different organizations in different ways, we can state that large US organizations are likely to hold a lot of sensitive data.

But there are more countries with many inhabitants and a large number of sizeable organizations. And these countries only appear in a single case or few cases from our data set. Hence, we could argue that non-US organization contain less vulnerabilities. However, we do not think is the case. More than we have seen in other countries, organizations in the US are liable to compensate data subjects for damages. This incentivizes US organizations to keep up the level of their IT security, together with the incentivization coming from legislation and industry standards. Hence, we assume the overrepresentation of US organization is mainly because of them being required to disclose data breaches.

Since most breached organizations in our data set are from the US, it could be argued that the results are mostly valid for the US only. However, we consider our findings to be valid to organizations from non-US countries, since an organizations base location is not directly influencing the presence of security misconfigurations, which could lead to data breaches, or their prevention. Like earlier stated, other factors like the size of the organization and data characteristics are more interesting with regard to that. What we can take away from these specific results is that in terms of disclosure legislation, but also on the reporting about data breaches, other countries could follow the example of the US.

Because of their function, which is to store large amounts of data, we expected databases to be involved as a breached component in a data breach often. However, databases being involved in more than half of the cases, arguably also says something about the state of database security in organizations. In Dietrich et al. (2018) the abundance of unprotected database systems was already mentioned. To some extent, this currently also appears to be the case for cloud storage, which we encountered on some cases as the breached component which was poorly secured or not secured at all. Additionally, in a quarter of the cases a breached web server was involved. This indicates that web server infrastructures also contain vulnerabilities relatively often. In a few of these cases, the breached web server had to do with a vulnerable application programming interface.

The bulk of the cases that involve breached databases, cloud storages or web servers, were found to have been caused by security misconfigurations. This is important because security misconfigurations, including their treatment and prevention, inherently involve IT infrastructures. Databases, cloud storages and web servers are integral parts of IT infrastructures. In other words, the (lack of) configuring of databases, cloud storages and web servers could potentially lead to security misconfigurations and therefore affects security. Our findings emphasize that these components should be protected with extra care.

In almost all cases, the ultimate responsibility for the compromised data rested with the breached organizations themselves as the data controller. In only very few cases, the breached organization acted as the data processor and was processing, and therefore possessing, the data on behalf of another party during the breach. This distribution could be because in general most large organizations appear to deal with data which they collected themselves and hold as the owners, instead of operating with other organizations' data. But it might also be related to data processors being more careful with data they operate with on behalf another party, than data controllers operating with data they own themselves. However, there are too few cases with data processors to be able to make any kind of statements, which means we also cannot state that being either a controller or processor influences the presence of security misconfigurations. We assume our findings are valid for both data controllers and data processors, since neither has a direct influence on the presence or prevention of security misconfigurations. If we had found a significant number of cases in which a data processor suffered a data breach because of security misconfiguration, then this would have been interesting with regard to the handling of the breach, liabilities and the effect on the relationship between the data processor and the original owner of the data. Because in such cases the data processor will likely be held responsible for not implementing appropriate security measures, which sometimes will even be subject to certain legislation like the GDPR.

All cases involve natural persons as data subject in the data breach. Furthermore, we have found that customers are affected in in two thirds of the data breach cases. This is significant in relation to other data subjects, especially since most of the cases with customers as the compromised data subject have security misconfiguration as the root cause. We suggest the incentivization of data protection which is specific per data subject.

Data belonging to data subjects like customers, employees, patients or even the breached organization itself, can be of different types. Indeed, data subject shows some correlation with data type. Similar to natural persons being involved in every case as the data subject, personal information is involved in all studied cases.

Normally, data protection legislation includes instructions on how to retrieve, process and store data of natural persons. Although there exists data protection legislation in most countries worldwide (Greenleaf, 2021a, 2021b), most of the cases involving personal information have security misconfiguration as the root cause. Hence, legislation does not appear to result in personal information being protected from getting compromised due to a preventable error. Legislation seems only to be good for punishing breached organizations after the breach. But the affected data subjects are not really benefited by that and they could potentially be troubled by the consequences for a long time. Therefore, further research is needed into how legislation potentially can contribute to the decrease of security misconfigurations.

All cases are significantly large in terms of the amount of breached data, with a substantial part of the cases involving a very large amount of breached data. This means that we specifically considered large cases in our research, and inherently dismissed the smaller cases in terms of data amount. The severity and impact of a data breach, and therefore the amount of breached data, is relevant for the breached organization. However, for the data subjects whose data got compromised, it does not really matter if that data was part of a small data breach or a large one. By not including small data breach cases, we excluded a part of the real victims of the smaller data breaches.

On the other hand, in 2020 a huge increase of breached records was seen in a lower total of breaches compared to the year before (Whitney, 2021). The average data breach in 2020 was significantly bigger than it has been in almost a decade before (Whitney, 2021). From 2018 to 2019 the average of breached records per breach also increased (Whitney, 2021). Additionally, a large and heterogeneous amount of data is more

complex than a small and less heterogeneous amount of data. Therefore, the security of large amounts of data is also more complex and much more difficult to do properly. This means there is a higher data breach risk for large organizations holding a large amount of heterogeneous data. Also, the potential overall impact is higher than for a small data breach. Hence, our focus on large cases does not appear unjust.

Thus, we cannot be sure that our findings are also applicable to organizations that are more likely to fall victim to a smaller data breach. Cases concerning smaller amounts of breached data will more likely concern smaller sized breached organizations, and smaller organizations often have less sophisticated security policies and technologies. And as mentioned earlier, our findings are only valid for large organizations. So, we were not out of line by only studying large data breaches in terms of data amount.

Because of these differences, we suggest that data breaches involving a smaller amount of breached data should be studied separately. But then these should be made more visible than they are currently. Incentivizing the disclosure of data breaches and public investigations could contribute to this.

From the explanations of data states in literature (Andress & Leary, 2017; Shabtai, Elovici, & Rokach, 2012e), we derive that data-at-rest should be easier to protect than data-in-use and data-in-motion. Still, we found data-at-rest to be compromised in the vast majority of the cases. This can relate to the organization components that we saw were involved often, which are those relating to storage. The state of data depends on how and where they were situated when they got compromised. Those situations are dependent on configuration, which determines the security level of the infrastructure.

We found that most of the cases involving data-at-rest and all of the cases involving data-in-use were caused by security misconfiguration. Because of the huge amounts of data that organizations use, transfer, and especially have in storage, protecting them is essential. Organizations should know all their data and where these data reside. Additionally, by classifying the data, organizations will be able to tag assets as more or less critical, enabling them to apply suitable security to each of those assets.

In almost a quarter of the cases, the compromised data also contained unnecessary data. Unnecessarily having sensitive data in possession, means there is an unnecessary higher risk, since the potential impact is larger than if the unnecessary data was not there. Since in the majority of the cases with unnecessary data there was present security misconfiguration as the root cause, organizations should realize it is likely that risks materialize and should therefore do everything they can to minimize the risks. Specifically, organizations should know all their data and where these reside. Additionally, organizations should implement and execute comprehensive data retention policies, which includes the disposal of data which are no longer needed. We agree with Milo (2019) that data disposal policies must be developed and executed for the purpose of benefiting and protecting organizations and individuals.

Merely less than one third of the data breaches was detected by resources from within the breached organization and in less than one fourth of the cases this was done in an automated manner. Furthermore, in just over half of the cases the breach detection happened intentionally. This shows there is room for improvement of effectiveness and efficiency of organization’s own breach detection capabilities. This is important, but it is more important to prevent the data breaches. This can be done by discovering security misconfigurations before they lead to data breaches. Especially when we consider that four-fifth of all the cases involve security misconfigurations.

Considering the detection of security misconfigurations, Dietrich et al. (2018) looked at how misconfigurations are discovered in practice and identified three main cases, namely, detection due to an incident, accidental detection, and detection during an audit. Furthermore, the respondents of their qualitative study brought forward that the discovery of a security misconfiguration, either due to an incident, by accident, or during an audit, has a positive effect on the organization’s security posture.

It is disconcerting that in almost half of the cases it took a long time to detect the data breach. Especially since in nearly all those cases we identified security configuration as the root cause. In the report by Ponemon Institute (2018) it already came forward that not being able to quickly identify a data breach increases costs.

Organizations should have arrangements in place to keep the time between a potential breach and its detection as short as possible.

Hence, by improving detection capabilities, organizations are better equipped to discover data breaches timely, which enables them to respond better to data breaches. But more importantly, a lot of security misconfigurations can potentially be discovered before they even become an incident. In that way, a contribution is made to the prevention of data breaches and the overall security posture.

We realize that breached organizations sometimes out of strategic behavior are not honest about the moment and method of detection, or about the actual starting point of the breach for that matter, as they are inclined to portray a breach period that is as short as possible and to portray that they were in control of the situation. We see that here also the earlier mentioned incentivization of transparent disclosure and open investigation can contribute.

The major part of the studied cases involved security misconfiguration as root cause. This is a confirmation of what was claimed by Dietrich et al. (2018), namely that incidents are frequently the result of security misconfigurations, and therefore are preventable.

7.2. Limitations

Before formulating conclusions, there are some limitations that have to be considered. In our study we developed a novel assessment framework to assess data breaches which, to our knowledge, has not been used before in this manner. We conducted a qualitative study using a non-statistical sample of organizations that have suffered a data breach between 2013 and 2019. While we used certain criteria for selecting data breach cases to ensure a representative sample, the selection of data breaches is ultimately based on judgment and not on a statistical method. Hence, the generalizability of our results may be limited and only representative of the population of organizations we did study. We observe that, due to the availability of data for these cases, our case selection is biased towards data breaches involving large organizations and a large amount of compromised data. However, given the high impact of such organizations being breached, we believe that—despite only looking at a subset of the problem space—our results provide an important, first, observation of the issue. Nevertheless, future studies should investigate the ‘heavy tail’ of data breaches in smaller organizations. The framework we designed can be used for this, and—if applied to these cases—our results can be used as a reference point for these subsequent studies.

8 Conclusion

The research aims to identify to which extent root causes of data breaches are related to the configuration of IT infrastructure and whether data breaches can be considered commonly preventable.

In order to do so, we conduct our study on the basis of four research questions. We pose the first research question to identify known security misconfigurations. The answering of this first research question, provides us with insight in different aspects of security misconfigurations. These aspects consist mainly of common technical errors which represent corresponding security misconfigurations. Moreover, we are able to derive a general overview of known security misconfiguration types from the literature.

The literature on the human factor in IT security, strongly and repeatedly emphasizes the critical role of people in IT security. Based on that perspective, this inherently also applies to the security configuration of systems, since humans do the configuring and are responsible for it. This further establishes the involvement of humans in security misconfigurations. Because of this, we can relate the potential human factor to technical errors.

Hence, answering the first research question equipped us with better capabilities for recognizing and identifying security misconfigurations in data breach cases. Or otherwise, to recognize the lack thereof, meaning the root cause is something else than a security misconfiguration.

The second research question poses the question of how we can assess data breaches. We want to characterize them in a way that suits the purpose of our research. To structure the assessments and make sure only relevant case information is taken into account, we formulate a framework. This framework also ensures each case is assessed in the same way.

The third research question is about drawing up the key parameters and metrics to assess data breaches. The characterization happens on the basis of certain factors. To determine which factors to use in the framework, we consider the most important aspects of a data breach. We find the first relevant aspect to consider is to find out by whom the breach was done, through which actions, and by using which methods. Next, we figure that we need to know more about the organization that got breached. Naturally, we are interested in the compromised data themselves. Additionally, we want insight in the way and time period in which the breach was discovered.

Each of these four themes brings forward multiple parameters that are specifically used to characterize data breaches. By determining the applicable values for each parameter in relation to a certain data breach, our thought process is framed towards considering the relevant factors that are useful when identifying the root cause for the final parameter. But even as stand-alone parameters they bring interesting insights.

These parameters form our assessment framework, which are applied to the selected data breaches. By studying the descriptive statistics of the assessed parameter values across our data set, we work on answering the fourth research question. This question aims to determine, based on our assessment, whether the majority of severe data breaches are preventable and if there are any outstanding factors.

The results indicate that in breaches the data are mostly subject to unauthorized access by outsiders, which frequently is made possible by poor security. The organizations directly responsible for that data are large organizations which get breached especially in their storage facilities. Next to the organization which got breached, these sizeable data breaches always affect individuals since at least part of the compromised data is about them or linkable to them. Usually this is not even discovered by the breached organization itself and sometimes only after a long period of time.

Based on a qualitative assessment of data breaches by means of our newly developed assessment framework, it can be concluded that data breaches are frequently caused by security misconfigurations and therefore usually are preventable.

Bibliography

500px. (2019, February 13, 2019). Security Issue February 2019: FAQ. Retrieved from https://support.500px.com/hc/en-us/articles/360017752493-Security-Issue-February-2019-FAQ

500px. (n.d.-a). Who we are. Retrieved from https://500px.com/about

500px. (n.d.-b). We are 500px. Retrieved from https://500px.com/team

500px. (n.d.-c). 500px. Retrieved from https://nl.linkedin.com/company/500px

Abbott, C., & Goegan, S. (2018, March 29, 2018). Travel-booking site Orbitz hit with major data breach. Retrieved from https://www.natlawreview.com/article/travel-booking-site-orbitz-hit-major-data-breach

Abel, R. (2018, November 30, 2018). Sky Brasil exposes data of 32M customers on ElasticSearch. Retrieved from https://www.scmagazine.com/home/security-news/sky-brasil-one-of-the-biggest-subscription-television-services-in-brazil-is-the-latest-elasticsearch-server-user-to-leave-its-customers-exposed-after-not-securing-the-server-with-a-password/

Abine. (n.d.). About Abine, The Online Privacy Company. Retrieved from https://abine.com/about.html

Abine Inc. (n.d.). Abine, Inc. Retrieved from https://www.linkedin.com/company/abine-inc./about/

Abine Online Privacy. (2018, January 16, 2019). Blur Security Update. Retrieved from https://www.abine.com/blog/2018/blur-security-update/

Abrams, L. (2019, January 2, 2019). Abine Blur Password Manager User Data Exposed Online. Retrieved from https://www.bleepingcomputer.com/news/security/abine-blur-password-manager-user-data-exposed-online/

Adebayo, A. O., & Omotosho, Y. A. A. O. J. (2013). System and Data Capture Framework Insights into Breach Data toward Improved Feedback. *4*(3).

AFP. (2019a, June 21, 2019). Massive Data Breach at Canada Credit Union Giant Desjardins. *Security Week*. Retrieved from https://www.securityweek.com/massive-data-breach-canada-credit-union-giant-desjardins

AFP. (2019b, November 01, 2019). Canada Credit Union Data Breach Bigger Than First Thought: Desjardins. *Security Week*. Retrieved from https://www.securityweek.com/canada-credit-union-data-breach-bigger-first-thought-desjardins

Agencies. (2018, March 30, 2018). Hackers steal data of 150 million MyFitnessPal app users. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018/mar/30/hackers-steal-data-150m-myfitnesspal-app-users-under-armour

Ahmady, G. A., Mehrpour, M., & Nikooravesh, A. (2016). Organizational Structure. *Procedia - Social and Behavioral Sciences, 230*, 455-462. doi:https://doi.org/10.1016/j.sbspro.2016.09.057

Ali, Z. (2020, April 7, 2020). The world's 100 largest banks, 2020. Retrieved from https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/the-world-s-100-largest-banks-2020-57854079

American Medical Collection Agency. (2019). INFORMATION ABOUT THE COMPANY. Retrieved from http://amcaonline.com/info.html

Andress, J., & Leary, M. (2017). Chapter 6 - Protect the Data. In J. Andress & M. Leary (Eds.), *Building a Practical Information Security Program* (pp. 103-123): Syngress.

Aorato Labs. (2014). *The Untold Story of the Target Attack Step by Step*. Retrieved from https://aroundcyber.files.wordpress.com/2014/09/aorato-target-report.pdf

Arias, D. (2018, May 31, 2018). Hashing in Action: Understanding bcrypt. Retrieved from https://auth0.com/blog/hashing-in-action-understanding-bcrypt/

Article 29 Working Party. (2017). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. In.

Ashford, W. (2018a, January 22, 2018). Norwegian healthcare breach alert failed GDPR requirements. *Computer Weekly*.

Ashford, W. (2018b, November 30, 2018). Marriott data breach highlights basic failings. *Computer Weekly*.

Ashok, I. (2018a, January 19, 2018). Massive data breach hits Norway and over 3 million people's healthcare data feared stolen by hackers. *International Business Times*. Retrieved from

https://www.ibtimes.co.uk/massive-data-breach-hits-norway-over-3-million-peoples-healthcare-data-feared-stolen-by-hackers-1655856
Ashok, I. A. (2018b, March 16, 2018). Walmart Partner Exposed Personal Data Of 1.3 Million US And Canadian Shoppers. *International Business Times*. Retrieved from https://www.ibtimes.com/walmart-partner-exposed-personal-data-13-million-us-canadian-shoppers-2663307
Associated Press. (2018, April 1, 2018). Up to 5m Saks and Lord & Taylor customers at risk after data breach. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018/apr/01/saks-lord-taylor-date-breach-5-million-cards
Australian National University. (2019a). *Annual Report 2018*. Retrieved from https://www.anu.edu.au/about/strategic-planning/annual-report-2018
Australian National University. (2019b). *Incident report on the breach of the Australian National University's administrative systems*: Australian National University.
Australian National University. (n.d.). The Australian National University. Retrieved from https://www.linkedin.com/school/australian-national-university/
Ayereby, M. P.-M. (2018). Overcoming data breaches and human factors in minimizing threats to cyber-security ecosystems. *Walden Dissertations and Doctoral Studies, 6163*.
Ayyagari, R. (2012). An Exploratory Analysis of Data Breaches from 2005-2011: Trends and Insights. *Journal of Information Privacy and Security, 8*(2), 33-56. doi:10.1080/15536548.2012.10845654
Bagnulo, R. (2018). The T-Mobile Data Breach. Retrieved from https://www.akana.com/blog/t-mobile-data-breach
Baker, P. (2019). Editorial: Bulgarian Data Hack Provides a Timely Warning of Data Breaches to Come. *Intertax*, 908-909.
Baker, W. H., & Wallace, L. (2007). Is Information Security Under Control?: Investigating Quality in Information Security Management. *IEEE Security & Privacy, 5*(1), 36-44. doi:10.1109/MSP.2007.11
Ball, J. (2014, June 5, 2014). Guardian launches SecureDrop system for whistleblowers to share files. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2014/jun/05/guardian-launches-securedrop-whistleblowers-documents
Barker, E., Barker, W., Burr, W., Polk, W., & Smid, M. (2007). Nist special publication 800-57. *NIST Special publication, 800*(57), 1-142.
Baxter, G., & Sommerville, I. (2010). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers, 23*(1), 4-17. doi:10.1016/j.intcom.2010.07.003 %J Interacting with Computers
Baxter, P., & Jack, S. M. (2008). Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. *The Qualitative Report, 13*(4), 544-559. doi:10.46743/2160-3715/2008.1573
BBC News. (2020, April 28, 2020). British Airways to cut up to 12,000 jobs as air travel collapses. Retrieved from https://www.bbc.com/news/business-52462660
Bodenheim, R., Butts, J., Dunlap, S., & Mullins, B. (2014). Evaluation of the ability of the Shodan search engine to identify Internet-facing industrial control devices. *International Journal of Critical Infrastructure Protection, 7*(2), 114-123. doi:https://doi.org/10.1016/j.ijcip.2014.03.001
Boicea, A., Radulescu, F., & Agapin, L. I. (2012). *MongoDB vs Oracle -- Database Comparison*. Paper presented at the Proceedings of the 2012 Third International Conference on Emerging Intelligent Data and Web Technologies. https://doi.org/10.1109/EIDWT.2012.32
Bojanova, I., Yesha, Y., Black, P. E., & Wu, Y. (2019, 15-19 Jul 2019). *Information Exposure (IEX): A New Class in the Bugs Framework (BF)*. Paper presented at the 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC).
Bosnell, J. (2018). *British Airways suffers data breach compromising information on over 429,000 customer cards*. Retrieved from https://managingrisktogether.orx.org/orx-news/orx-news-deep-dive-british-airways-data-breach
Bradley, T. (2018, March 30, 2018). Security Experts Weigh In On Massive Data Breach Of 150 Million MyFitnessPal Accounts. *Forbes*.
Brandom, R. (2017, Oct 3, 2017). Former Equifax CEO blames breach on a single person who failed to deploy patch. *The Verge*.
Breach Report. (2019). Canva Data Breach Information: Millions User Details Stolen Within a Week. Retrieved from https://breachreport.com/news/post/canva-data-breach-millions-user-details-stolen

Breach Report. (2020). Breach Catalogue - Canva.com. Retrieved from https://breachreport.com/catalogue?q=Canva.com#
Bressers, J. (2019). Tips to secure Elasticsearch clusters for free with encryption, users, and more. Retrieved from https://www.elastic.co/blog/tips-to-secure-elasticsearch-clusters-for-free-with-encryption-users-and-more
Bronskill, J. (2020, December 14, 2020). Data breach at Desjardins caused by series of gaps, privacy watchdog says. Retrieved from https://globalnews.ca/news/7520414/desjardins-data-breach-privacy-watchdog-probe/
Cameron, D. (2017, September 21, 2017). Passwords to Over a Half Million Car Tracking Devices Leaked Online. Retrieved from https://gizmodo.com/passwords-to-access-over-a-half-million-car-tracking-de-1818624272
Camtepe, S. A., & Yener, B. (2007, 17-21 Sept. 2007). *Modeling and detection of complex attacks.* Paper presented at the 2007 Third International Conference on Security and Privacy in Communications Networks and the Workshops - SecureComm 2007.
Canva [@canva]. (2019, May 25, 2019). This morning we've been alerted to a security incident... [Twitter Tweet]. Retrieved December 14, 2020 from https://twitter.com/canva/status/1132086889408749573
Canva. (n.d.). Canva. Retrieved from https://www.linkedin.com/company/canva/about/
Capital One. (2019a, September 23, 2019). Information on the Capital One Cyber Incident. Retrieved from https://www.capitalone.com/facts2019/
Capital One. (2019b, September 23, 2019). Frequently Asked Questions. Retrieved from https://www.capitalone.com/facts2019/2/
Carr, N. (2017). Cyber Espionage is Alive and Well: APT32 and the Threat to Global Corporations. Retrieved from https://www.fireeye.com/blog/threat-research/2017/05/cyber-espionage-apt32.html
Carter, A. B. (2019). Considerations for Genomic Data Privacy and Security when Working in the Cloud. *The Journal of Molecular Diagnostics, 21*(4), 542-552. doi:https://doi.org/10.1016/j.jmoldx.2018.07.009
Cathay Pacific. (2018a, October 24, 2018). Cathay Pacific announces data security event affecting passenger data. Retrieved from https://news.cathaypacific.com/cathay-pacific-announces-data-security-event-affecting-passenger-data
Cathay Pacific. (2018b). *Interim Report 2018*. Retrieved from https://www.cathaypacific.com/content/dam/cx/about-us/investor-relations/interim-annual-reports/en/2018_cx_interim_report_en.pdf
Cathay Pacific. (2018c). *Written submission by Cathay Pacific Airways Limited for the Joint Meeting on Wednesday, 14 November, 2018 of the Panel on Constitutional Affairs, Panel on Information Technology and Broadcasting, Panel on Security*. Retrieved from https://www.legco.gov.hk/yr18-19/english/panels/ca/papers/caitbse20181114cb2-222-2-e.pdf
Cathay Pacific. (2019). *Annual Report 2018*. Retrieved from https://www.cathaypacific.com/content/dam/cx/about-us/investor-relations/interim-annual-reports/en/2018_annual_report_en.pdf
Cellebrite. (2017a). Cellebrite statement on information security breach [Press release]. Retrieved from https://www.cellebrite.com/en/resources/press/cellebrite-statement-on-information-security-breach/
Cellebrite. (2017b). Cellebrite provides update on information security investigation to forensic customers [Press release]. Retrieved from https://www.cellebrite.com/en/resources/in-the-news/update-on-information-security-investigation-to-forensic-customers/
Cellebrite. (n.d.). Cellebrite. Retrieved from https://www.linkedin.com/company/cellebrite
Chang, S. E., & Ho, C. B. (2006). Organizational factors to the effectiveness of implementing information security management. *Industrial Management & Data Systems, 106*(3), 345-361. doi:10.1108/02635570610653498
Cheng, L., Liu, F., & Yao, D. (2017). Enterprise data breach: causes, challenges, prevention, and future directions. *7*(5), e1211. doi:https://doi.org/10.1002/widm.1211
Cherdantseva, Y., & Hilton, J. (2013). Information Security and Information Assurance. The Discussion about the Meaning, Scope and Goals,". In.
Christou, L. (2019, June 3, 2019). Gnosticplayers: Why the hacker behind the Canva data breach boasted to the media. Retrieved from https://www.verdict.co.uk/canva-data-breach-gnosticplayers/
Chtrbox. (n.d. (a)). WHAT IS CHTRBOX? Retrieved from https://www.chtrbox.com/#whatIsChtrbox
Chtrbox. (n.d. (b)). HOW DOES IT WORK? Retrieved from https://www.chtrbox.com/#howDoesItWork

Cichonski, P., Millar, T., Grance, T., & Scarfone, K. (2012). *NIST Special Publication 800-61 Revision 2, Computer Security Incident Handling Guide*.
Cimpanu, C. (2018a, January 18, 2018). Hacker Might Have Stolen the Healthcare Data for Half of Norway's Population. Retrieved from https://www.bleepingcomputer.com/news/security/hacker-might-have-stolen-the-healthcare-data-for-half-of-norways-population/
Cimpanu, C. (2018b, November 29, 2018). Sky Brasil exposes data of 32 million subscribers. Retrieved from https://www.zdnet.com/article/sky-brasil-exposes-data-of-32-million-subscribers/
Cimpanu, C. (2019a, January 2, 2019). Data of 2.4 million Blur password manager users left exposed online. Retrieved from https://www.zdnet.com/article/data-of-2-4-million-blur-password-manager-users-left-exposed-online/
Cimpanu, C. (2019b, March 30, 2019). Toyota announces second security breach in the last five weeks. Retrieved from https://www.zdnet.com/article/toyota-announces-second-security-breach-in-the-last-five-weeks/
Cimpanu, C. (2019c, May 24, 2019). Australian tech unicorn Canva suffers security breach. Retrieved from https://www.zdnet.com/article/australian-tech-unicorn-canva-suffers-security-breach/
Cimpanu, C. (2019d, June 20, 2019). Desjardins, Canada's largest credit union, announces security breach. Retrieved from https://www.zdnet.com/article/desjardins-canadas-largest-credit-union-announces-security-breach/
Cimpanu, C. (2019e, July 16, 2019). Hacker steals data of millions of Bulgarians, emails it to local media. Retrieved from https://www.zdnet.com/article/hacker-steals-data-of-millions-of-bulgarians-emails-it-to-local-media/
CipherCloud. (2019, February 19, 2019). 92 Million MyHeritage Genealogy Accounts Breached. Now What? Retrieved from https://securityboulevard.com/2019/02/92-million-myheritage-genealogy-accounts-breached-now-what-2/
Claburn, T. (2018, April 19, 2018). Millions of scraped public social net profiles left in open AWS S3 box. Retrieved from https://www.theregister.com/2018/04/19/48_million_personal_profiles_left_exposed_by_data_firm_localblox/
Claburn, T. (2019, March 11, 2019). The Handmaid's Tale or Man-made Fail? Exposed DB of 'BreedReady' women probably not as bad as it sounds. Retrieved from https://www.theregister.com/2019/03/11/exposed_database_breedready/
CloudSploit. (2019, August 2, 2019). A Technical Analysis of the Capital One Hack. Retrieved from https://blog.cloudsploit.com/a-technical-analysis-of-the-capital-one-hack-a9b43d7c8aea
Cluley, G. (2018). Security – it Shouldn't Just Be the Jewel in Your Crown, but Your Partners and Suppliers Too. Retrieved from https://securityboulevard.com/2018/03/security-it-shouldnt-just-be-the-jewel-in-your-crown-but-your-partners-and-suppliers-too/
Cluley, G. (2019, May 22, 2019). Data on millions of Instagram accounts spills onto the internet. Retrieved from https://www.tripwire.com/state-of-security/featured/data-millions-instagram-accounts-internet/
Cobb, M. (2017, February 10, 2017). What caused the ClixSense privacy breach that exposed user data? Retrieved from https://searchsecurity.techtarget.com/answer/What-caused-the-ClixSense-privacy-breach-that-exposed-user-data
Coble, S. (2019, December 16, 2019). Orbitz and Expedia Agree to Data Breach Settlement with Pennsylvania. *Infosecurity Magazine*.
Coble, S. (2020, February 27, 2020). Desjardins Group Breach Cost $38m Higher Than Expected. *Infosecurity Magazine*.
Coffee Meets Bagel (2019, February 14, 2019). [Important Information About Your Coffee Meets Bagel Account Security].
Committee of Inquiry. (2019). *Public Report of the Committee of Inquiry (COI) into the cyber attack on Singapore Health Services Private Limited Patient Database on or around 27 June 2018*. Retrieved from https://www.mci.gov.sg/-/media/mcicorp/doc/report-of-the-coi-into-the-cyber-attack-on-singhealth-10-jan-2019.ashx
completetechnology. (2019). Photo Site 500PX Hit With Data Breach Recently. Retrieved from https://www.ctsglobaltech.com/2019/03/11/photo-site-500px-hit-with-data-breach-recently/
Constantin, L. (2014, August 8, 2014). Payment cards with chips aren't perfect, so encrypt everything, experts say. *PC World*.

Cooper, D. L. (2015). *Data security: data breaches*. Paper presented at the Proceedings of the 2015 Information Security Curriculum Development Conference, Kennesaw, Georgia. https://doi.org/10.1145/2885990.2886003
Corfield, G. (2018, August 24, 2018). Hackers clock personal deets on 'two million' T-Mobile US subscribers. Retrieved from https://www.theregister.com/2018/08/24/t_mobile_us_data_breach/
Cox, J. (2017a, January 12, 2017). Hacker Steals 900 GB of Cellebrite Data. Retrieved from https://www.vice.com/en/article/3daywj/hacker-steals-900-gb-of-cellebrite-data
Cox, J. (2017b, January 12, 2017). Cellebrite Sold Phone Hacking Tech to Repressive Regimes, Data Suggests. Retrieved from https://www.vice.com/en/article/aekqjj/cellebrite-sold-phone-hacking-tech-to-repressive-regimes-data-suggests
Craft. (n.d.-a). Abine. Retrieved from https://craft.co/abine
Craft. (n.d.-b, December 2020). Canva. Retrieved from https://craft.co/canva
Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*: Sage Publications.
Crunchbase. (2020). Coffee Meets Bagel. Retrieved from https://www.crunchbase.com/organization/coffee-meets-bagel
Crunchbase. (n.d.-a). 500px. Retrieved from https://www.crunchbase.com/organization/500px
Crunchbase. (n.d.-b). Abine. Retrieved from https://www.crunchbase.com/organization/abine
Crunchbase. (n.d.-c). Canva. Retrieved from https://www.crunchbase.com/organization/canva
Crunchbase. (n.d.-d). Cellebrite. Retrieved from https://www.crunchbase.com/organization/cellebrite
Crunchbase. (n.d.-e). Chtrbox. Retrieved from https://www.crunchbase.com/organization/chtrbox
Crunchbase. (n.d.-f). Marriott International. Retrieved from https://www.crunchbase.com/organization/marriott-international
Crunchbase. (n.d.-g). Sky Brasil. Retrieved from https://www.crunchbase.com/organization/sky-brasil
Crunchbase. (n.d.-h). Suprema. Retrieved from https://www.crunchbase.com/organization/suprema-2
Crunchbase. (n.d.-i). Toyota Motor Corporation. Retrieved from https://www.crunchbase.com/organization/toyota
Csaszar, K. (2019). How secure is bcrypt? Retrieved from https://synkre.com/how-secure-is-bcrypt/
Daitch, H. (2016, November 15, 2016). AdultFriendFinder Data Breach Exposes Information from 412 Million Accounts. Retrieved from https://www.identityforce.com/blog/adultfriendfinder-data-breach
Daitch, H. (2018, September 18, 2018). Government Payment Service Leaks Personal Information of More Than 14 Millionn Customers. Retrieved from https://www.identityforce.com/blog/govpaynet-leaks-personal-information-14-million-customers
Daniel, L. G., & Onwuegbuzie, A. J. (2002). Reliability and Qualitative Data: Are Psychometric Concepts Relevant within an Interpretivist Research Paradigm? In.
Data Trust. (n.d.). Data Trust. Retrieved from https://www.linkedin.com/company/thedatatrust/about/
Davis, J. (2019, January 10, 2019). Massive SingHealth Data Breach Caused by Lack of Basic Security. Retrieved from https://healthitsecurity.com/news/massive-singhealth-data-breach-caused-by-lack-of-basic-security
DB CyberTech. (2019). *The Marriott Breach - A Classic Insider Threat that Behavioral Analysis Immediately Identifies*. Retrieved from https://www.dbcybertech.com/pdf/Marriot-Breach-White-Paper.pdf
de Looper, C. (2019, February 14, 2019). Happy Valentine's Day! Coffee Meets Bagel dating app data may have been breached. Retrieved from https://www.digitaltrends.com/mobile/coffee-meets-bagel-data-breach/
Death, D. (2017). *Information Security Handbook: Develop a threat model and incident response strategy to build a strong information security framework*: Packt Publishing.
Dellinger, A. (2019, May 26, 2019). Understanding The First American Financial Data Leak: How Did It Happen And What Does It Mean? *Forbes*.
Denizova, V. (2019, July 16, 2019). NRA: The hacked data is authentic. They are about 3% of our arrays. *Capital*. Retrieved from https://www.capital.bg/politika_i_ikonomika/bulgaria/2019/07/16/3938976_nap_haknatite_danni_sa_avtentichni_te_sa_okolo_3_ot/
Desjardins. (n.d.-a). Quick facts about Desjardins. Retrieved from http://web.archive.org/web/20190402222513/https://www.desjardins.com/ca/about-us/desjardins/who-we-are/quick-facts/index.jsp

Desjardins. (n.d.-b). Quick facts about Desjardins. Retrieved from https://www.desjardins.com/ca/about-us/desjardins/who-we-are/quick-facts/index.jsp
Deutsch, O. (2018, June 4, 2018). MyHeritage Statement About a Cybersecurity Incident. Retrieved from https://blog.myheritage.com/2018/06/myheritage-statement-about-a-cybersecurity-incident/
Dickey, M. R. (2016, November 13, 2016). FriendFinder Networks hack reportedly exposed over 412 million accounts. Retrieved from https://techcrunch.com/2016/11/13/friendfinder-hack-412-million-accounts-breached/
Dietrich, C., Krombholz, K., Borgolte, K., & Fiebig, T. (2018). *Investigating System Operators' Perspective on Security Misconfigurations*. Paper presented at the Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, Toronto, Canada.
Diogenes, Y., & Ozkaya, E. (2019). *Cybersecurity-Attack and Defense Strategies*: Packt Publishing.
Doe, D. (2019a, February 14, 2019). Coffee Meets Bagel notifies users of breach. Retrieved from https://www.databreaches.net/coffee-meets-bagel-notifies-users-of-breach/
Doe, D. (2019b, May 10, 2019). American Medical Collection Agency breach impacted 200,000 patients – Gemini Advisory. Retrieved from https://www.databreaches.net/american-medical-collection-agency-breach-impacted-200000-patients-gemini-advisory/
Doe, D. (2019c, June 3, 2019). Update on American Medical Collection Agency breach: Almost 12 million Quest Diagnostic patients impacted. Retrieved from https://www.databreaches.net/update-on-american-medical-collection-agency-breach-almost-12-million-quest-diagnostic-patients-impacted/
Doffman, Z. (2019, March 11, 2019). Chinese Data Breach Exposes 'Breed Ready' Status Of Almost 2 Million Women. *Forbes*.
Duckett, C. (2019, June 7, 2019). Cathay Pacific's unpatched decade-old vulnerability led to 2018 breach. Retrieved from https://www.zdnet.com/article/cathay-pacifics-unpatched-decade-old-vulnerability-led-to-2018-breach/
Duffy, J. (2020, June 22, 2020). How to Protect Yourself Online With Disposable Credit Card Numbers. *PC Magazine*.
Dunn, J. E. (2019, February 15, 2019). Photography site 500px resets 14.8 million passwords after data breach. Retrieved from https://nakedsecurity.sophos.com/2019/02/15/photography-site-500px-resets-14-8-million-passwords-after-data-breach/
Eisner, E. W. (2017). *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice, Reissued with a New Prologue and Foreword*: Teachers College Press.
Elastic. (2020). Elasticsearch. Retrieved from https://www.elastic.co/elasticsearch/
Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1-4. doi:10.11648/j.ajtas.20160501.11
European Central Bank. (2014). ECB announces theft of contact information [Press release]. Retrieved from https://www.ecb.europa.eu/press/pr/date/2014/html/pr140724.en.html
European Central Bank. (2015). *Annual Report 2014*. Retrieved from https://www.ecb.europa.eu/pub/pdf/annrep/ar2014en.pdf
European Central Bank. (2020). About. Retrieved from https://www.ecb.europa.eu/ecb/html/index.en.html
European Commission. (2016). User guide to the SME Definition. In.
European Data Protection Board. (2021). Guidelines 01/2021 on Examples regarding Data Breach Notification. In.
European Network and Information Security Agency. (2020). *Physical manipulation/ damage/ theft/ loss*. Retrieved from https://www.enisa.europa.eu/publications/physical-manipulation-damage-theft-loss
Fater, T. (2014, February 25, 2014). Student data at risk for disclosure. *Indiana Daily Student*. Retrieved from https://www.idsnews.com/article/2014/02/student-data-at-risk-for-disclosure
Fawkes, G. (2019). *Report: Data Breach in Biometric Security Platform Affecting Millions of Users*. Retrieved from https://www.vpnmentor.com/blog/report-biostar2-leak/
Fiebig, T. (2017). *An empirical evaluation of misconfiguration in Internet services.* Retrieved from http://nbn-resolving.de/urn:nbn:de:101:1-201804165184
http://d-nb.info/1156186447
Fiebig, T. (2020, 7-11 Sept. 2020). *How to stop crashing more than twice: A Clean-Slate Governance Approach to IT Security.* Paper presented at the 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW).

Field, T. (2014, February 26, 2014). Why Target Breach Was Preventable. Retrieved from https://www.bankinfosecurity.com/target-breach-was-preventable-a-6543
First American. (n.d.). First American. Retrieved from https://www.linkedin.com/company/first-american/about/
First American Financial Corporation. (2019). *2018 Annual Report*. Retrieved from http://s21.q4cdn.com/992793803/files/doc_financials/2018/Annual/2018-FAF-Annual-Report.pdf
First American Financial Corporation. (2020). *2019 Annual Report*. Retrieved from http://s21.q4cdn.com/992793803/files/doc_financials/2019/ar/2019_FAF_Annual_Report_BMK_vWEB.pdf
Forman, I. (2018). Security risks of DNA testing. In: Comp 116: Security.
Fortune. (2021a). Fortune 500. Retrieved from https://fortune.com/fortune500/2020/search/
Fortune. (2021b). Global 500. Retrieved from https://fortune.com/global500/2020/search/
Franceschi-Bicchierai, L. (2018, August 24, 2018). Hackers Stole Personal Data of 2 Million T-Mobile Customers. Retrieved from https://www.vice.com/en/article/a3qpk5/t-mobile-hack-data-breach-api-customer-data
Fraser, B. (1997). *Site Security Handbook*: RFC Editor.
Fruhlinger, J. (2020, February 12, 2020). Marriott data breach FAQ: How did it happen and what was the impact? *CSO Online*.
Fu, D. (2018). Inside Information Data Breach [Press release]. Retrieved from https://www1.hkexnews.hk/listedco/listconews/sehk/2018/1024/ltn20181024757.pdf
Furnell, S., & Clarke, N. (2012). Power to the people? The evolving recognition of human aspects of security. *Comput. Secur., 31*(8), 983-988. doi:10.1016/j.cose.2012.08.004
Galan Manso, C., Górniak, S., & European Network and Information Security Agency. (2013). Recommendations for a methodology of the assessment of severity of personal data breaches. In. Heraklion: European Network and Information Security Agency (ENISA).
Gamer, N. (2015, December 8, 2015). Defending against new POS malware with EMV technology. Retrieved from https://blog.trendmicro.com/%E2%80%8Bdefending-against-new-pos-malware-with-emv-technology/
Gatlan, S. (2019, March 29, 2019). Toyota Security Breach Exposes Personal Info of 3.1 Million Clients. Retrieved from https://www.bleepingcomputer.com/news/security/toyota-security-breach-exposes-personal-info-of-31-million-clients/
GDPR. (2016). {Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)}. *Official Journal of the European Union, L119*, 1-88. doi:citeulike-article-id:14071352
Gemalto. (2019). Breach Level Index. from Gemalto https://breachlevelindex.com/
Gemini Advisory. (2018, April 1, 2018). Fin7 Syndicate Hacks Saks Fifth Avenue and Lord & Taylor Stores. Retrieved from https://geminiadvisory.io/fin7-syndicate-hacks-saks-fifth-avenue-and-lord-taylor/
George, T. (2018, June 28, 2018). The Next Big Cyber-Attack Vector: APIs. *Security Week*. Retrieved from https://www.securityweek.com/next-big-cyber-attack-vector-apis
Gevers, V. [@0xDUDE]. (2019a, March 9, 2019). In China, they have a shortage of women. ... [Twitter Tweet]. Retrieved December 30, 2020 from https://twitter.com/0xDUDE/status/1104482014202351616
Gevers, V. [@0xDUDE]. (2019b, March 11, 2019). The unprotected database is not reachable anymore for... [Twitter Tweet]. Retrieved December 30, 2020 from https://twitter.com/0xDUDE/status/1105095049937932288
Gevers, V. [@0xDUDE]. (2019c, March 16, 2019). Two databases with almost the same database schema... [Twitter Tweet]. Retrieved December 30, 2020 from https://twitter.com/0xDUDE/status/1106909141040353281
Gevers, V. [@0xDUDE]. (2019d, March 22, 2019). Last night we shared on @i24NEWS a new insight. ... [Twitter Tweet]. Retrieved December 30, 2020 from https://twitter.com/0xDUDE/status/1109043562992332801
Gevers, V. [@0xDUDE]. (2019e, March 22, 2019). The original dataset came from an official source ... [Twitter Tweet]. Retrieved December 30, 2020 from https://twitter.com/0xDUDE/status/1109044765063024640
Gilbert, H., & Handschuh, H. (2004). *Security Analysis of SHA-256 and Sisters*, Berlin, Heidelberg.

Gillis, A. S. (2020, August 2020). Definition: biometrics. Retrieved from https://searchsecurity.techtarget.com/definition/biometrics
Goel, V., & Abrams, R. (2018, April 1, 2018). Card Data Stolen From 5 Million Saks and Lord & Taylor Customers. *The New York Times*. Retrieved from https://www.nytimes.com/2018/04/01/technology/saks-lord-taylor-credit-cards.html
Goldstein, M., Perlroth, N., & Corkery, M. (2014, December 22, 2014). Neglected Server Provided Entry for JPMorgan Hackers. *The New York Times*. Retrieved from https://dealbook.nytimes.com/2014/12/22/entry-point-of-jpmorgan-data-breach-is-identified/
Gonzalez, C. V., & Jung, G. (2019). *Teaching Cyber Security Topics Effectively in a College or University with Limited Resources.* Paper presented at the 2019 International Conference on Computational Science and Computational Intelligence (CSCI).
Goodin, D. (2016, September 13, 2016). 6.6 million plaintext passwords exposed as site gets hacked to the bone. *Ars Technica*.
Goswami, S. (2019, April 1, 2019). Toyota Reveals a Second Data Breach. Retrieved from https://www.bankinfosecurity.com/toyota-reveals-second-data-breach-a-12303
Gramm-Leach-Bliley, (1999).
Grance, T., Kim, B., & Scarfone, K. (2004). *NIST Special Publication 800-61, Computer Security Incident Handling Guide*.
Greenleaf, G. (2021a). Global Data Privacy Laws 2021: Despite COVID Delays, 145 Laws Show GDPR Dominance. *Privacy Laws & Business International Report*(169), 3-5. doi:10.2139/ssrn.3836348
Greenleaf, G. (2021b). Global Tables of Data Privacy Laws and Bills (7th Ed, January 2021). *Privacy Laws & Business International Report*(169), 6-19.
Greenwood, M. G. (2018, September 7, 2018). So, about that BA hack …. Retrieved from https://medium.com/the-automator/so-about-that-ba-hack-a82e5701f095
Gregorio-de Souza, I., Berk, V. H., Giani, A., Bakos, G., Bates, M., Cybenko, G., & Madory, D. (2006). *Detection of complex cyber attacks.* Paper presented at the Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense V.
Gregorio, N., Mathanamohan, J., Mahmoud, Q. H., & AlTaei, M. (2019). Hacking in the cloud. *2*(1), e84. doi:https://doi.org/10.1002/itl2.84
Greve, M., Masuch, K., & Trang, S. (2020). *The More, the Better? Compensation and Remorse as Data Breach Recovery Actions – An Experimental Scenario-based Investigation*. Paper presented at the 15th International Conference on Wirtschaftsinformatik, Potsdam, Germany. https://www.researchgate.net/publication/339797664_The_More_the_Better_Compensation_and_Remorse_as_Data_Breach_Recovery_Actions_-_An_Experimental_Scenario-based_Investigation
Gustafsson, J. (2017). Single case studies vs. multiple case studies: A comparative study. In.
Hadlington, L. (2018). The "human factor" in cybersecurity: Exploring the accidental insider. In (pp. 46-63).
Hauer, B. (2015). Data and Information Leakage Prevention Within the Scope of Information Security. *IEEE Access, 3*, 2554-2565. doi:10.1109/ACCESS.2015.2506185
Hawkins, B. (2015). *Case Study: The Home Depot Data Breach*. Retrieved from https://www.sans.org/reading-room/whitepapers/casestudies/paper/36367
Heidt, M., Gerlach, J. P., & Buxmann, P. (2019). Investigating the Security Divide between SME and Large Companies: How SME Characteristics Influence Organizational IT Security Investments. *Information Systems Frontiers, 21*(6), 1285-1305. doi:10.1007/s10796-019-09959-1
Helse Sør-Øst. (2018, January 30, 2018). Innbrudd i datasystemene til Sykehuspartner i Helse Sør-Øst. Retrieved from https://www.helse-sorost.no/nyheter/innbrudd-i-datasystemene-til-sykehuspartner-i-helse-sor-ost
Honan, B. (2014, July 31, 2014). European Central Bank hacked. *CSO Online*.
Hoog, A., & Strzempka, K. (2011). Chapter 4 - iPhone and iPad data security. In A. Hoog & K. Strzempka (Eds.), *iPhone and iOS Forensics* (pp. 79-105). Boston: Syngress.
Houlihan, D. (2018, April 3, 2018). No, Panera Bread Doesn't Take Security Seriously. Retrieved from https://medium.com/@djhoulihan/no-panera-bread-doesnt-take-security-seriously-bf078027f815
Hughes, M. (2018, March 14, 2018). Jewelry site accidentally leaks personal details (and plaintext passwords!) of 1.3M users. Retrieved from https://thenextweb.com/security/2018/03/14/jewelry-site-accidentally-leaks-personal-details-plaintext-passwords-1-3m-users/
Hunt, T. (2016). ClixSense. Retrieved from https://haveibeenpwned.com/PwnedWebsites#ClixSense

IBM Cloud Education. (2020). Application Programming Interface (API). Retrieved from https://www.ibm.com/cloud/learn/api
IEEE Reliability Society. (2017). IEEE Standard Framework for Prognostics and Health Management of Electronic Systems. In *IEEE Std 1856-2017* (pp. 1-31).
Ikeda, S. (2019a, April 9, 2019). New Toyota Data Breach Exposes Personal Information of 3.1 Million Customers. *CPO Magazine*.
Ikeda, S. (2019b, June 10, 2019). Security Oversight at First American Causes Data Leak of 900 Million Records. *CPO Magazine*.
Ikeda, S. (2019c, June 11, 2019). Third Party Data Breach Hits Quest Diagnostics with 12 Million Confidential Patient Records Exposed. *CPO Magazine*.
Ilascu, I. (2018, November 29, 2018). SKY Brasil Exposes 32 Million Customer Records. Retrieved from https://www.bleepingcomputer.com/news/security/sky-brasil-exposes-32-million-customer-records/
Ilascu, I. (2019, March 22, 2019). Creepy Database Lists 'BreedReady' Status for 1.8 Million Women. Retrieved from https://www.bleepingcomputer.com/news/security/creepy-database-lists-breedready-status-for-18-million-women/
Imran, M., Faisal, M., & Islam, N. (2019, November 13-14, 2019). *Problems and Vulnerabilities of Ethical Hacking in Pakistan.* Paper presented at the 2019 Second International Conference on Latest trends in Electrical Engineering and Computing Technologies (INTELLECT).
Indiana University USSS. (2014). FREQUENTLY ASKED QUESTIONS (FAQ): USSS DATA EXPOSURE. Retrieved from https://apps.usss.iu.edu/usss-data-exposure/faq.cfm
Information Commissioner's Office. (2019, July 08, 2019). Intention to fine British Airways £183.39m under GDPR for data breach. Retrieved from https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/07/ico-announces-intention-to-fine-british-airways/
Information Commissioner's Office. (2020a, October 30, 2020). ICO fines Marriott International Inc £18.4million for failing to keep customers' personal data secure. Retrieved from https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/10/ico-fines-marriott-international-inc-184million-for-failing-to-keep-customers-personal-data-secure/
Information Commissioner's Office. (2020b, November 13, 2020). ICO fines Ticketmaster UK Limited £1.25million for failing to protect customers' payment details. Retrieved from https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/11/ico-fines-ticketmaster-uk-limited-125million-for-failing-to-protect-customers-payment-details/
Information Commissioner's Office. (2020c). PENALTY NOTICE. In.
ISO. (2012). ISO/IEC 27032. In.
ISO. (2015a). Information technology — Security techniques — Storage security. In (1 ed.): International Organization for Standardization.
ISO. (2015b). Space systems -- Closed loop problem solving management. In: International Organization for Standardization.
ISO. (2017a). ISO 10646 Information technology -- Universal Coded Character Set. In: International Organization for Standardization.
ISO. (2017b). ISO/IEC/IEEE 24765 Systems and software engineering — Vocabulary. In.
ISO. (2018). ISO/IEC 27000:2018. In.
ISO. (2020). Codes for the representation of names of countries and their subdivisions — Part 1: Country code. In: International Organization for Standardization.
J.P. Morgan. (2020). About Us. Retrieved from https://www.jpmorgan.com/about
Jarvis, K., & Milletary, J. (2014). *Inside a Targeted Point-of-Sale Data Breach*. Retrieved from https://portal.secureworks.com/intel/mva?Task=ShowThreat&ThreatId=773
Jeng, A. (2015). Minimizing damage from JP Morgan's data breach. In: SANS Institute.
Jimenez-Gomez, C. E. (2017). *Data Breaches 2004-2017 (EN)*. Retrieved from: https://www.kaggle.com/estratic/data-breaches-2004-2017-en-20180218
Johnson, S. (2018, April 18, 2018). Data firm left detailed profiles of 48 million people on a publicly accessible website. Retrieved from https://bigthink.com/stephen-johnson/data-firm-left-detailed-profiles-of-48-million-people-on-a-publicly-accessible-website
Jøsang, A. (2017). *A Consistent Definition of Authorization*, Cham.
Kassner, M. (2015, February 2, 2015). Anatomy of the Target data breach: Missed opportunities and lessons learned. Retrieved from https://www.zdnet.com/article/anatomy-of-the-target-data-breach-missed-opportunities-and-lessons-learned/

Kasun, C. (2020a, June 26, 2020). API Security Part #1: Introduction. Retrieved from https://medium.com/@chamikakasun/api-security-part-1-introduction-b7160e9cc071
Kasun, C. (2020b, July 1, 2020). API Security Part #2: API Security Vulnerabilities. Retrieved from https://medium.com/@chamikakasun/api-security-part-2-api-security-vulnerabilities-b33c08c44127
Khandelwal, S. (2016, September 14, 2016). Massive Data Breach Exposes 6.6 Million Plaintext Passwords from Ad Company. Retrieved from https://thehackernews.com/2016/09/plaintext-passwords-leaked.html
Khandelwal, S. (2017a, September 22, 2017). Passwords For 540,000 Car Tracking Devices Leaked Online. Retrieved from https://thehackernews.com/2017/09/hacker-track-car.html
Khandelwal, S. (2017b, January 12, 2017). Phone-Hacking Firm Cellebrite Got Hacked; 900GB Of Data Stolen. Retrieved from https://thehackernews.com/2017/01/mobile-hacking-cellebrite.html
Kirk, J. (2016, November 14, 2016). Alleged Adult Website Breach May Affect 412 Million Accounts. Retrieved from https://www.bankinfosecurity.com/alleged-adult-website-breach-may-affect-412-million-accounts-a-9519
Kirk, J. (2018, August 27, 2018). T-Mobile Database Breach Exposes 2 Million Customers' Data. Retrieved from https://www.bankinfosecurity.com/t-mobile-database-breach-affects-two-million-customers-a-11420
Kirk, J. (2019a, May 21, 2019). Database May Have Exposed Instagram Data for 49 Million. Retrieved from https://www.bankinfosecurity.com/database-may-have-exposed-instagram-personal-data-a-12503
Kirk, J. (2019b, May 24, 2019). Instagram Bans Social Media Company After Data Exposure. *Data Breach Today*. Retrieved from https://www.databreachtoday.com/instagram-bans-social-media-company-after-data-exposure-a-12518
Kirk, J. (2019c, May 27, 2019). Title Company Exposes 16 Years of US Mortgage Data. *Data Breach Today*. Retrieved from https://www.databreachtoday.com/title-company-exposes-16-years-us-mortgage-data-a-12524
Kissel, R. (2013). *Glossary of key information security terms*. Retrieved from https://nvlpubs.nist.gov/nistpubs/ir/2013/NIST.IR.7298r2.pdf
Klijnsma, Y. (2018, September 11, 2018). Inside the Magecart Breach of British Airways: How 22 Lines of Code Claimed 380,000 Victims. Retrieved from https://www.riskiq.com/blog/labs/magecart-british-airways-breach/
Klijnsma, Y., Kremez, V., & Herman, J. (2018). *Inside Magecart: Profiling the Groups Behind the Front Page Credit Card Breaches and the Criminal Underworld that Harbors Them*. Retrieved from https://cdn.riskiq.com/wp-content/uploads/2018/11/RiskIQ-Flashpoint-Inside-MageCart-Report.pdf
Kolesnikov, O., & Parashar, H. (2018). *BRITISH AIRWAYS BREACH: MAGECART FORMGRABBING SUPPLY CHAIN ATTACK DETECTION*. Retrieved from https://www.securonix.com/resources/securonix-threat-research-british-airways-breach-magecart-formgrabbing-supply-chain-attack-detection/
Kostadinov, D. (2019, June 19, 2019). Information and Asset Classification. Retrieved from https://resources.infosecinstitute.com/certification/information-and-asset-classification/
Kovacs, E. (2016, September 15, 2016). 6.6 Million Users Affected by ClixSense Breach. *Security Week*. Retrieved from https://www.securityweek.com/66-million-users-affected-clixsense-breach
Kovacs, E. (2017, January 13, 2017). Mobile Forensics Firm Cellebrite Hacked. *Security Week*. Retrieved from https://www.securityweek.com/mobile-forensics-firm-cellebrite-hacked
Kovacs, E. (2018a, November 28, 2018). Industry Reactions to USPS Exposing User Data. *Security Week*. Retrieved from https://www.securityweek.com/industry-reactions-usps-exposing-user-data
Kovacs, E. (2018b, March 20, 2018). Orbitz Data Breach Impacts 880,000 Payment Cards. *Security Week*. Retrieved from https://www.securityweek.com/orbitz-data-breach-impacts-880000-payment-cards
Kovacs, E. (2019a, January 03, 2019). Blur Exposes Information of 2.4 Million Users. *Security Week*. Retrieved from https://www.securityweek.com/blur-exposes-information-24-million-users
Kovacs, E. (2019b, March 29, 2019). Millions of Toyota Customers in Japan Hit by Data Breach. *Security Week*. Retrieved from https://www.securityweek.com/millions-toyota-customers-japan-hit-data-breach
Kovacs, E. (2019c, June 03, 2019). AMCA Breach Hits 12 Million Quest Diagnostics Patients. *Security Week*. Retrieved from https://www.securityweek.com/amca-breach-hits-12-million-quest-diagnostics-patients

Kovacs, E. (2019d, June 05, 2019). LabCorp Says 7.7 Million Patients Caught in AMCA Data Breach. *Security Week*. Retrieved from https://www.securityweek.com/labcorp-says-77-million-patients-caught-amca-data-breach
Kovacs, E. (2019e, July 17, 2019). AMCA Breach Impacts 2.2 Million Patients of Clinical Pathology Laboratories. *Security Week*. Retrieved from https://www.securityweek.com/amca-breach-impacts-22-million-patients-clinical-pathology-laboratories
Kovacs, E. (2019f, July 22, 2019). AMCA Breach: Many More Impacted Healthcare Firms Come Forward. *Security Week*. Retrieved from https://www.securityweek.com/amca-breach-many-more-impacted-healthcare-firms-come-forward
Kraemer, S., Carayon, P., & Clem, J. (2009). Human and organizational factors in computer and information security: Pathways to vulnerabilities. *Computers & Security, 28*(7), 509-520. doi:https://doi.org/10.1016/j.cose.2009.04.006
Krasimirov, A., & Tsolova, T. (2019, October 26, 2019). In systemic breach, hackers steal millions of Bulgarians' financial data. Retrieved from https://www.reuters.com/article/us-bulgaria-cybersecurity/hackers-steal-millions-of-bulgarians-financial-records-tax-agency-idUSKCN1UB0MA
Krebs, B. (2013a, December 19, 2013). Sources: Target Investigating Data Breach. Retrieved from https://krebsonsecurity.com/2013/12/sources-target-investigating-data-breach/
Krebs, B. (2013b). Cards Stolen in Target Breach Flood Underground Markets. Retrieved from https://krebsonsecurity.com/2013/12/cards-stolen-in-target-breach-flood-underground-markets/
Krebs, B. (2014a). Email Attack on Vendor Set Up Breach at Target. Retrieved from https://krebsonsecurity.com/2014/02/email-attack-on-vendor-set-up-breach-at-target/
Krebs, B. (2014b, September 2, 2014). Banks: Credit Card Breach at Home Depot. Retrieved from https://krebsonsecurity.com/2014/09/banks-credit-card-breach-at-home-depot/
Krebs, B. (2014c, September 18, 2014). Home Depot: 56M Cards Impacted, Malware Contained. Retrieved from https://krebsonsecurity.com/2014/09/home-depot-56m-cards-impacted-malware-contained/
Krebs, B. (2014d, September 18, 2014). In Home Depot Breach, Investigation Focuses on Self-Checkout Lanes. Retrieved from https://krebsonsecurity.com/2014/09/in-home-depot-breach-investigation-focuses-on-self-checkout-lanes/
Krebs, B. (2015). Inside Target Corp., Days After 2013 Breach. Retrieved from https://krebsonsecurity.com/2015/09/inside-target-corp-days-after-2013-breach/
Krebs, B. (2018a, April 2, 2018). Panerabread.com Leaks Millions of Customer Records. Retrieved from https://krebsonsecurity.com/2018/04/panerabread-com-leaks-millions-of-customer-records/#more-43216
Krebs, B. (2018b, June 6, 2018). Researcher Finds Credentials for 92 Million Users of DNA Testing Firm MyHeritage. Retrieved from https://krebsonsecurity.com/2018/06/researcher-finds-credentials-for-92-million-users-of-dna-testing-firm-myheritage/
Krebs, B. (2018c, September 17, 2018). GovPayNow.com Leaks 14M+ Records. Retrieved from https://krebsonsecurity.com/2018/09/govpaynow-com-leaks-14m-records/
Krebs, B. (2018d). USPS Site Exposed Data on 60 Million Users. Retrieved from https://krebsonsecurity.com/2018/11/usps-site-exposed-data-on-60-million-users/
Krebs, B. (2019a). First American Financial Corp. Leaked Hundreds of Millions of Title Insurance Records. Retrieved from https://krebsonsecurity.com/2019/05/first-american-financial-corp-leaked-hundreds-of-millions-of-title-insurance-records/
Krebs, B. (2019b). Collections Firm Behind LabCorp, Quest Breaches Files for Bankruptcy. Retrieved from https://krebsonsecurity.com/2019/06/collections-firm-behind-labcorp-quest-breaches-files-for-bankruptcy/
Krebs, B. (2019d, August 2, 2019). What We Can Learn from the Capital One Hack. Retrieved from https://krebsonsecurity.com/2019/08/what-we-can-learn-from-the-capital-one-hack/
Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015). Advanced social engineering attacks. *Journal of Information Security and Applications, 22*, 113-122. doi:https://doi.org/10.1016/j.jisa.2014.09.005
Kuo, L. (2019, March 11, 2019). China database lists 'breedready' status of 1.8 million women. *The Guardian*. Retrieved from https://www.theguardian.com/world/2019/mar/11/china-database-lists-breedready-status-of-18-million-women
Kurane, S. (2014, December 23, 2014). JPMorgan data breach entry point identified: NYT. Retrieved from https://www.reuters.com/article/us-jpmorgan-cybersecurity/jpmorgan-data-breach-entry-point-identified-nyt-idUSKBN0K105R20141223

Lambert, L. K. (2016, June 22, 2016). Catching a RAT by the tail. *CSO Online*.
Lei, J. (2019, December 11, 2019). Suprema Casts a Cloud on Biometric Security. Retrieved from https://medium.com/@jinghul.lei/suprema-casts-a-cloud-on-biometric-security-383374e95ca0
Leung, H. (2019, March 11, 2019). A Researcher Uncovered a Chinese Database That Lists the "BreedReady" Status of 1.8 Million Women. *Time*.
Liao, S. (2018, November 22, 2018). USPS took a year to fix a vulnerability that exposed all 60 million users' data. Retrieved from https://www.theverge.com/2018/11/22/18107945/usps-postal-service-data-vulnerability-security-patch-60-million-users
Lindsey, N. (2019, June 26, 2019). AMCA Healthcare Data Breach Could Set a New Precedent for Health IT Security. *CPO Magazine*.
Liptak, A. (2019, February 13, 2019). Personal information of 14.8 million 500px users leaked in security breach. Retrieved from https://www.theverge.com/2019/2/13/18223660/500px-security-breach-14-8-million-users-personal-information-stolen-cybersecurity
Lock, S. (2020, March 4, 2020). Number of Marriott International employees worldwide from 2007 to 2019. Retrieved from https://www.statista.com/statistics/297269/number-of-marriott-international-employees-worldwide/
Lord, N. (2019). Data Protection: Data In transit vs. Data At Rest. Retrieved from https://digitalguardian.com/blog/data-protection-data-in-transit-vs-data-at-rest
Lu, J. (2019). Assessing The Cost, Legal Fallout Of Capital One Data Breach.
Macrotrends. (n.d.). First American Financial: Number of Employees 2009-2020. Retrieved from https://www.macrotrends.net/stocks/charts/FAF/first-american-financial/number-of-employees
Mares, O. (2018, December 4, 2018). SKY BRAZIL HACKED; OVER 30 MILLION USERS AFFECTED. *Information Security Newspaper*. Retrieved from https://www.securitynewspaper.com/2018/12/04/sky-brazil-hacked-over-30-million-users-affected/
Marriott International. (2018a). Original Notice from November 30, 2018. In *[First Announcement]*.
Marriott International. (2018b). Marriott Announces Starwood Guest Reservation Database Security Incident [Press release]. Retrieved from https://news.marriott.com/news/2018/11/30/marriott-announces-starwood-guest-reservation-database-security-incident
Marriott International. (2019). Updated: March 4, 2019. In *[Second Announcement]*.
Marriott International. (n.d.). Marriott International. Retrieved from https://nl.linkedin.com/company/marriott-international
Martin, L. (2019, June 4, 2019). Australian National University hit by huge data breach. *The Guardian*. Retrieved from https://www.theguardian.com/australia-news/2019/jun/04/australian-national-university-hit-by-huge-data-breach
Mathews, L. (2018, August 24, 2018). Hackers Swipe Data On 2 Million T-Mobile Subscribers. *Forbes*.
Matthews, K. (2019, April 5, 2019). Incident Of The Week: Toyota's Second Data Breach Affects Millions Of Drivers. Retrieved from https://www.cshub.com/attacks/articles/incident-of-the-week-toyotas-second-data-breach-affects-millions-of-drivers
Mazareanu, E. (2020, July 7, 2020). Number of passengers traveling with British Airways and its subsidiaries from 2011 to 2019. Retrieved from https://www.statista.com/statistics/734311/british-airways-passenger-figures/#:~:text=British%20Airways%20and%20subsidiaries%3A%20number%20of%20passengers%20carried%202011%2D2019&text=In%202019%2C%20British%20Airways%2C%20including,subsidiaries%2C%20uplifted%2047.7%20million%20passengers.
McCallister, E., Grance, T., & Scarfone, K. A. (2010). *SP 800-122. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*. Retrieved from https://dl.acm.org/doi/10.5555/2206206
McCandless, D., Quick, M., Hollowood, E., Miles, C., & Hampson, D. (2019). World's Biggest Data Breaches & Hacks. from Information is Beautiful https://informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks
McDaniel, D. (2019). Data Breaches: Who is behind them, why they do it, and how to protect your data David McDaniel East Carolina University.
McGee, M. K. (2018, March 30, 2018). Under Armour Reports Massive Breach of MyFitnessPal App. Retrieved from https://www.bankinfosecurity.com/under-armour-reports-massive-breach-myfitnesspal-app-a-10756

McGee, M. K. (2019, June 3, 2019). Quest Diagnostics: Data on 12 Million Patients Exposed. Retrieved from https://www.bankinfosecurity.com/quest-diagnostics-data-on-12-million-patients-exposed-a-12560
Mehta, I. (2019, May 21, 2019). (Updated) Private data (including rates) of 49M Instagram influencers leaked due to agency's malpractice. Retrieved from https://thenextweb.com/security/2019/05/21/private-data-including-rates-of-49m-instagram-influencers-leaked-due-to-agencys-malpractice/
Melo, S. (2019). API Integrations: 5 Ways they can benefit your business. Retrieved from https://mydatascope.com/blog/en/api-integrations-5-ways-they-can-benefit-your-business/
Metalidou, E., Marinagi, C., Trivellas, P., Eberhagen, N., Skourlas, C., & Giannakopoulos, G. (2014). *The Human Factor of Information Security: Unintentional Damage Perspective* (Vol. 147).
Mikalsen, K. S. (2018, January 18, 2018). Mistenker at en fremmed stat står bak hackerangrep. Retrieved from https://www.nrk.no/osloogviken/mistenker-at-en-fremmed-stat-star-bak-hackerangrep-1.13869547
Milo, T. (2019). Getting Rid of Data. *Journal of Data and Information Quality, 12*(1), Article 1. doi:10.1145/3326920
Mishra, S. (2019). *Forensic Investigation Framework for Complex Cyber Attack on Cyber Physical System by Using Goals/Sub-goals of an Attack and Epidemics of Malware in a System*, Singapore.
Monitor Agency. (2019, July 15, 2019). The NRA is investigating a data leak of 5 million Bulgarians and foreigners. *Monitor*. Retrieved from https://www.monitor.bg/bg/a/view/nap-razsledva-tech-na-danni-na-5-miliona-bylgari-i-chujdenci-170594
Montpetit, J. (2019, June 21, 2019). Personal data of 2.7 million people leaked from Desjardins. Retrieved from https://www.cbc.ca/news/canada/montreal/desjardins-data-breach-1.5183297
Muncaster, P. (2017, February 3, 2017). Cellebrite Hack Highlights Dangers of State Snooping. *Infosecurity Magazine*.
Muncaster, P. (2018a, September 18, 2018). Government Payment Service Exposes 14m Records. *Infosecurity Magazine*.
Muncaster, P. (2018b, April 3, 2018). Panera Bread Data Leak May Have Hit Millions: Report. *Infosecurity Magazine*.
Muncaster, P. (2018c, November 15, 2018). Skimmed BA and Newegg Customer Card Details Up for Sale. *Infosecurity Magazine*.
Muncaster, P. (2018d, November 22, 2018). US Postal Service Exposes 60 Million Users in API Snafu. *Infosecurity Magazine*.
Muncaster, P. (2019, January 3, 2019). Password Manager Users Exposed After Privacy Snafu. *Infosecurity Magazine*.
MyFitnessPal. (2018). MyFitnessPal Account Security Issue: Frequently Asked Questions [Press release]. Retrieved from https://content.myfitnesspal.com/security-information/FAQ.html
Naiakshina, A., Danilova, A., Tiefenau, C., Herzog, M., Dechand, S., & Smith, M. (2017). *Why Do Developers Get Password Storage Wrong?: A Qualitative Usability Study*. Paper presented at the Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, Texas, USA.
Nair, P. (2020, July 23, 2020). First American Title Insurance Co. Faces Charges in NY. Retrieved from https://www.bankinfosecurity.com/first-american-mortgage-faces-charges-in-ny-a-14692
Namanya, A. P., Cullen, A., Awan, I. U., & Disso, J. P. (2018, 6-8 Aug. 2018). *The World of Malware: An Overview.* Paper presented at the 2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud).
National Revenue Agency. (n.d. (a)). National Revenue Agency - Republic of Bulgaria. Retrieved from https://www.linkedin.com/company/national-revenue-agency-republic-of-bulgaria
National Revenue Agency. (n.d. (b)). About the NRA. Retrieved from https://nra.bg/page?id=680
National Revenue Agency. (n.d. (c)). List of NRA offices. Retrieved from https://nra.bg/page?id=546
news.bg. (2019). NRA, SANS, Ministry of Interior are checking for a possible hacker attack. In.
Nidecki, T. A. (2020). What Are Insecure Direct Object References. Retrieved from https://www.acunetix.com/blog/web-security-zone/what-are-insecure-direct-object-references/
Nieles, M., Dempsey, K., & Pillitteri, V. (2017). NIST special publication 800-12 revision 1. from National Institute of Standards and Technology: https://nvlpubs. nist. gov/nistpubs/SpecialPublications/NIST. SP. 800-12r1. pdf. *Accessed August, 7*, 2018.

Nieles, M., Dempsey, K., & Pillitteri, V. Y. (2017). An introduction to information security. In.
NJCCIC. (2018, March 23, 2018). Walmart Partner MBM Company Exposes Data on 1.3 Million Customers. Retrieved from https://www.cyber.nj.gov/public-data-breaches/walmart-partner-mbm-company-exposes-data-on-13-million-customers
Nohe, P. (2019). Autopsying the Marriott Data Breach: This is why insurance matters. Retrieved from https://www.thesslstore.com/blog/autopsying-the-marriott-data-breach-this-is-why-insurance-matters/
Novaes Neto, N., Madnick, S., de Paula, M. G., & Malara Borges, N. (2020). A Case Study of the Capital One Data Breach (Revised). *SSRN Electronic Journal*. doi:10.2139/ssrn.3570138
O'Dea, S. (2020, February 27, 2020). Number of employees of T-Mobile US from 2013 to 2019. Retrieved from https://www.statista.com/statistics/483653/t-mobile-us-employees/
O'Donnell, L. (2018, March 15, 2018). Walmart Jewelry Partner Exposes Personal Data Of 1.3M Customers. Retrieved from https://threatpost.com/walmart-jewelry-partner-exposes-personal-data-of-1-3m-customers/130486/
O'Donnell, L. (2019a, February 14, 2019). Coffee Meets Bagel Dating App Warns Users of Breach. Retrieved from https://threatpost.com/coffee-meets-bagel-breach/141850/
O'Donnell, L. (2019b, August 15, 2019). Fingerprints of 1M Exposed in Public Biometrics Database. Retrieved from https://threatpost.com/fingerprints-of-1m-exposed-in-public-biometrics-database/147345/
O'Neill, P. H. (2017, July 17, 2019). What happens when a country's entire adult population is hacked? *MIT Technology Review*.
Office of the Privacy Commissioner of Canada. (2020a). Desjardins breach: Quebec's Commission d'accès à l'information and the Office of the Privacy Commissioner of Canada to release results of investigations [Press release]. Retrieved from https://www.priv.gc.ca/en/opc-news/news-and-announcements/2020/ma_201213/
Office of the Privacy Commissioner of Canada. (2020b, December 14, 2020). Investigation into Desjardins' compliance with PIPEDA following a breach of personal information between 2017 and 2019. Retrieved from https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2020/pipeda-2020-005/
Open Semantic Search. (n.d.). Search operators and wildcards. Retrieved from https://opensemanticsearch.org/doc/search/operators
Orbitz. (2018). INFORMATION ABOUT ORBITZ DATA SECURITY INCIDENT. Retrieved from https://orbitz.allclearid.com/additionalinformation.html
Orr, J. (2019, May 6, 2019). Top Cyber Security Breaches So Far. Retrieved from https://www.cshub.com/attacks/articles/top-5-cyber-security-breaches-of-2019-so-far
Osborne, C. (2016, September 14, 2016). ClixSense data breach exposes personal information of millions of subscribers. Retrieved from https://www.zdnet.com/article/clixsense-data-breach-exposes-personal-information-of-million-of-subscribers/
Osborne, C. (2017, January 13, 2017). Mobile hacking firm Cellebrite confirms server breach. Retrieved from https://www.zdnet.com/article/mobile-hacking-firm-cellebrite-confirms-server-breach/
Osborne, C. (2018a, September 18, 2018). GovPayNow payment portal may have exposed over 14 million customer records. Retrieved from https://www.zdnet.com/article/govpaynow-data-breach-leaks-over-14-million-customer-records/
Osborne, C. (2018b, August 24, 2018). Hackers help themselves to data belonging to 2 million T-Mobile customers. Retrieved from https://www.zdnet.com/article/international-hackers-help-themselves-to-data-belonging-to-2-million-t-mobile-customers/
Osborne, C. (2019, June 4, 2019). Massive Quest Diagnostics data breach impacts 12 million patients. Retrieved from https://www.zdnet.com/article/massive-quest-diagnostics-data-breach-impacts-12-million-patients/
Padwal, K., Thomas, A., Howard, T., & Carr, M. (2019). Common Lessons from Disparate Information Security Incidents.
Paganini, P. (2017, September 24, 2017). Passwords and much more for 540,000 SVR Tracking accounts leaked online. Retrieved from https://securityaffairs.co/wordpress/63343/data-breach/svr-tracking-data-leak.html

Paganini, P. (2018a, March 18, 2018). Unsecured AWS S3 bucket managed by Walmart jewelry partner exposes data of 1.3M customers. Retrieved from https://securityaffairs.co/wordpress/70381/data-breach/walmart-jewelry-partner-leak.html
Paganini, P. (2018b, April 19, 2018). Private Intelligence agency LocalBlox leaked 48 Million personal data records. Retrieved from https://securityaffairs.co/wordpress/71534/data-breach/localblox-data-leak.html
Pagano Dritto, G. (2019, March 27, 2019). An Overview on Elasticsearch and its usage. Retrieved from https://towardsdatascience.com/an-overview-on-elasticsearch-and-its-usage-e26df1d1d24a
Panera Bread. (2020). *Panera Responsibility Report (2017-2019)*. Retrieved from https://www-beta.panerabread.com/content/dam/panerabread/integrated-web-content/documents/press/2020/panera-bread-csr-2017-2019.pdf
Panera Bread. (n.d.-a). Our History. Retrieved from https://www.panerabread.com/en-us/company/our-history.html
Panera Bread. (n.d.-b). Panera Bread. Retrieved from https://www.linkedin.com/company/panera-bread
Park, M., & Chai, S. (2018). The Value of Personal Information : An Exploratory Study for Types of Personal Information and Its Value. *Asia Pacific Journal of Information Systems, 28*, 154-166. doi:10.14329/apjis.2018.28.3.154
Parsons, M., Lin, M., & Phillips, B. (2019). *The Cathay Pacific Breach: Is Data Protection and Cyber Security Law in Hong Kong about to receive an upgrade?* Retrieved from https://f.datasrvr.com/fr1/119/29799/Client_Alert_-_The_Cathay_Pacific_Breach-Is_Data_Protection_and_Cyber_Security_Law_in_Hong_Kong_about_to_receive_an_upgrade.pdf
Pascu, L. (2018, December 5, 2018). Data of 32 million SKY Brasil customers easily accessible on unprotected ElasticSearch server. Retrieved from https://securityboulevard.com/2018/12/data-of-32-million-sky-brasil-customers-easily-accessible-on-unprotected-elasticsearch-server/
Patrawala, F. (2019). Canva faced security breach, 139 million users data hacked: ZDNet reports. Retrieved from https://securityboulevard.com/2019/05/canva-faced-security-breach-139-million-users-data-hacked-zdnet-reports/
Pennsylvania Office of Attorney General. (2019). AG Shapiro Announces Settlement with Orbitz and Expedia in Data Breach Affecting Pennsylvania Consumers [Press release]. Retrieved from https://www.attorneygeneral.gov/taking-action/press-releases/ag-shapiro-announces-settlement-with-orbitz-and-expedia-in-data-breach-affecting-pennsylvania-consumers/
Perez, S. (2019, February 14, 2019). Happy Valentine's Day: your dating app account was hacked, says Coffee Meets Bagel. Retrieved from https://techcrunch.com/2019/02/14/happy-valentines-day-your-dating-app-account-was-hacked-says-coffee-meets-bagel/
Perhar, J. (2018). How an inexpensive iFrame could have saved British Airways from a potential £500 million fine. Retrieved from https://www.pcibooking.net/british-airways/
Perlroth, N., & Goldstein, M. (2014, September 12, 2014). After Breach, JPMorgan Still Seeks to Determine Extent of Attack. *The New York Times*. Retrieved from https://www.nytimes.com/2014/09/13/technology/after-breach-jpmorgan-still-seeks-to-determine-extent-of-attack.html
Peters, J. (2018). Weekly Cyber Risk Roundup: Orbitz Breach, Facebook Privacy Fallout. Retrieved from https://securityboulevard.com/2018/03/weekly-cyber-risk-roundup-orbitz-breach-facebook-privacy-fallout/
Pieters, W. (2013, 2-5 Dec. 2013). *Defining "The Weakest Link" Comparative Security in Complex Systems of Systems.* Paper presented at the 2013 IEEE 5th International Conference on Cloud Computing Technology and Science.
Plachkinova, M., & Maurer, C. (2018). Teaching Case: Security Breach at Target. *Journal of Information Systems Education, 29*(1), 11-20.
Ponemon Institute. (2018). *2018 Cost of a Data Breach Study: Global Overview*. Retrieved from
Ponemon Institute. (2020). Cost of a data breach report 2019. In.
Porter, J. (2019, Augustus 14, 2019). Huge security flaw exposes biometric data of more than a million users. Retrieved from https://www.theverge.com/2019/8/14/20805194/suprema-biostar-2-security-system-hack-breach-biometric-info-personal-data
Posey Garrison, C., & Ncube, M. (2011). A longitudinal analysis of data breaches. *Information Management & Computer Security, 19*(4), 216-230. doi:10.1108/09685221111173049

Price, A. (2018, August 25, 2018). Report: Texas voter records exposed online. *Austin American-Statesman*. Retrieved from https://www.statesman.com/news/20180825/report-texas-voter-records-exposed-online
Privacy Rights Clearinghouse. (2019). Data Breaches. Retrieved 25-04-2019, from Privacy Rights Clearinghouse https://www.privacyrights.org/data-breaches
Radichel, T. (2014). Case study: Critical controls that could have prevented target breach. In: SANS Institute InfoSec Reading Room.
Ragan, S. (2014a, September 18, 2014). What you need to know about the Home Depot data breach. *CSO Online*.
Ragan, S. (2016b, October 20, 2016). Penthouse, Adult FriendFinder databases leak, at least 100 million accounts impacted. *CSO Online*.
Ragan, S. (2016c, November 13, 2016). 412 million FriendFinder accounts exposed by hackers. *CSO Online*.
Reeve, T. (2018, September 10, 2018). An analysis of the British Airways payment page shows that the site is loading files from seven external domains that have little or nothing to do with payment processing, according to a security expert. *SC Media*.
Reuters. (2019, May 20, 2019). Israel's MyHeritage expands DNA testing to health. Retrieved from https://www.reuters.com/article/us-healthcare-myheritage-idUSKCN1SQ1EG
Robinson, T. (2018a, March 16, 2018). Open AWS S3 bucket managed by Walmart jewelry partner exposes info on 1.3M customers. *SC Media*.
Robinson, T. (2018b, September 18, 2018). 14 million customer records exposed in GovPayNow leak. *SC Media*.
Robinson, T. (2019b, March 29, 2019). Toyota reports second breach in five weeks. *SC Media*.
Robinson, T. (2019c, May 21, 2019). Unsecure Chtrbox AWS database exposes data on 49 million Instagram influencers, accounts. *SC Media*.
Roman, J. (2014a, February 27, 2014). Indiana University Reports Breach. *Data Breach Today*. Retrieved from https://www.databreachtoday.com/indiana-university-reports-breach-a-6579
Roman, J. (2014b, September 15, 2014). JPMorgan Chase Confirms Cyber-Attack. Retrieved from https://www.bankinfosecurity.com/jpmorgan-a-7319
Rosenblum, P. (2018). Confusion Reigns In The Wake Of Saks, Lord And Taylor Data Breach. *Forbes*.
Sachowski, J. (2016). Chapter 9 - Establish Secure Storage and Handling. In J. Sachowski (Ed.), *Implementing Digital Forensic Readiness* (pp. 95-104). Boston: Syngress.
Sæther, A. S. (2018, January 15, 2018). Datasystemene til Helse Sør-Øst angrepet. *Verdens Gang*. Retrieved from https://www.vg.no/nyheter/innenriks/i/0E4W4G/datasystemene-til-helse-soer-oest-angrepet
Sæther, A. S., Bugge, S., & Sarmadawy, H. (2018, January 15, 2018). Kan ikke utelukke at pasientjournaler er på avveie. *Verdens Gang*. Retrieved from https://www.vg.no/nyheter/innenriks/i/8wBO0G/kan-ikke-utelukke-at-pasientjournaler-er-paa-avveie
Saleem, H., & Naveed, M. (2020). SoK: Anatomy of Data Breaches. *Proceedings on Privacy Enhancing Technologies, 2020*, 153-174. doi:10.2478/popets-2020-0067
Schmidt, B. (2019). Message from the Vice-Chancellor. In.
Schwalb, E. M. (2003). *ITV Handbook: Technologies and Standards*: Prentice Hall PTR.
Schwartz, M. J. (2014, July 24, 2014). European Central Bank Breached. Retrieved from https://www.bankinfosecurity.com/euro-central-bank-suffers-data-breach-a-7101
Schwartz, M. J. (2015, April 6, 2015). Why POS Malware Still Works. Retrieved from https://www.bankinfosecurity.com/pos-malware-still-works-a-8044
Schwartz, M. J. (2018a, October 26, 2018). British Airways Finds Hackers Stole More Payment Card Data. Retrieved from https://www.bankinfosecurity.com/british-airways-finds-hackers-stole-more-payment-card-data-a-11645
Schwartz, M. J. (2018b, March 20, 2018). Expedia's Orbitz Suspects 880,000 Payment Cards Stolen. Retrieved from https://www.bankinfosecurity.com/expedias-orbitz-suspects-880000-payment-cards-stolen-a-10729
Schwartz, M. J. (2018c, September 11, 2018). RiskIQ: British Airways Breach Ties to Cybercrime Group. Retrieved from https://www.bankinfosecurity.com/riskiq-british-airways-breach-ties-to-cybercrime-group-a-11481

Schwartz, M. J. (2018d, April 2, 2018). Saks, Lord & Taylor Suffer Payment Card Data Breach. Retrieved from https://www.bankinfosecurity.com/saks-lord-taylor-suffer-payment-card-data-breach-a-10757
Schwartz, M. J. (2018e, June 28, 2018). Ticketmaster Breach Traces to Embedded Chatbot Software. Retrieved from https://www.bankinfosecurity.com/ticketmaster-breach-traces-to-embedded-chatbot-software-a-11144
Seals, T. (2014, November 7, 2014). Home Depot: Massive Breach Happened Via Third-Party Vendor Credentials. *Infosecurity Magazine*.
Seals, T. (2017, March 13, 2017). Home Depot to Pay $27.25m in Latest Data Breach Settlement. *Infosecurity Magazine*.
Seals, T. (2018b, March 20, 2018). Orbitz Attack Impacts Hundreds of Thousands of Consumers. *Infosecurity Magazine*.
Securus Technologies. (2015). Securus Provides Updates on Investigation into Stolen Data Records [Press release]. Retrieved from https://www.prnewswire.com/news-releases/securus-provides-updates-on-investigation-into-stolen-data-records-300178709.html
Securus Technologies. (2020). Securus Technologies. Retrieved from https://www.linkedin.com/company/securus-technologies
Sen, R., & Borle, S. (2015). Estimating the Contextual Risk of Data Breach: An Empirical Approach. *Journal of Management Information Systems, 32*(2), 314-341. doi:10.1080/07421222.2015.1063315
Shabtai, A., Elovici, Y., & Rokach, L. (2012a). Data Leakage Detection/Prevention Solutions. In *A Survey of Data Leakage Detection and Prevention Solutions* (pp. 17-37). Boston, MA: Springer US.
Shabtai, A., Elovici, Y., & Rokach, L. (2012b). Data Leakage/Misuse Scenarios. In A. Shabtai, Y. Elovici, & L. Rokach (Eds.), *A Survey of Data Leakage Detection and Prevention Solutions* (pp. 39-46). Boston, MA: Springer US.
Shabtai, A., Elovici, Y., & Rokach, L. (2012c). Introduction to Information Security. In *A Survey of Data Leakage Detection and Prevention Solutions* (pp. 1-4). Boston, MA: Springer US.
Shabtai, A., Elovici, Y., & Rokach, L. (2012d). *A Survey of Data Leakage Detection and Prevention Solutions*.
Shabtai, A., Elovici, Y., & Rokach, L. (2012e). A Taxonomy of Data Leakage Prevention Solutions. In *A Survey of Data Leakage Detection and Prevention Solutions* (pp. 11-15). Boston, MA: Springer US.
Shackleford, D. (2007). *Regulations and Standards: Where Encryption Applies*. Retrieved from https://www.sans.org/reading-room/whitepapers/vpns/paper/34675
Shingler, B. (2019, July 23, 2019). What you need to know about the Desjardins data breach. Retrieved from https://www.cbc.ca/news/canada/montreal/desjardins-data-breach-explain-1.5185163
Shu, X., Tian, K., Ciambrone, A., & Danfeng, Y. (2017). Breaking the Target: An Analysis of Target Data Breach and Lessons Learned.
Silver-Greenberg, J., Goldstein, M., & Perlroth, N. (2014, October 2, 2014). JPMorgan Chase Hacking Affects 76 Million Households. *The New York Times*. Retrieved from https://dealbook.nytimes.com/2014/10/02/jpmorgan-discovers-further-cyber-security-issues/
SKY Brasil. (2020). SKY Brasil Internet. Retrieved from https://assine.sky.com.br/banda-larga
Smith, J., & Lee, M. (2015, November 11, 2015). NOT SO SECURUS. *The Intercept*.
Smith, J., & Lee, M. (2016, February 12 2016). NOT SO SECURUS: PART 2. *The Intercept*.
Smith, M. (2018, April 3, 2018). Panera Bread blew off breach report for 8 months, leaked millions of customer records. *CSO Online*.
Solomon, H. (2020, December 14th, 2020). Desjardins at fault for huge data breach, say privacy commissioners. Retrieved from https://www.itworldcanada.com/article/breaking-desjardins-at-fault-for-huge-data-breach-say-privacy-commissioners/439581
Sommerville, I. (1998). Systems engineering for software engineers. *Annals of Software Engineering, 6*(1), 111-129. doi:10.1023/A:1018901230131
Sorenson, A. (2019). Testimony of Arne Sorenson, President & CEO, Marriott International Before the Senate Committee on Homeland Security & Governmental Affairs Permanent Subcommittee on Investigations. In.
Sprecher, A. M. (2018, March 30, 2018). The Under Armour Hack Was Even Worse Than It Had To Be. *Wired*.
Spring, T. (2018a, March 21, 2018). Orbitz Warns 880,000 Payment Cards Suspected Stolen. Retrieved from https://threatpost.com/orbitz-warns-880000-payment-cards-suspected-stolen/130601/

Spring, T. (2018b, August 24, 2018). T-Mobile Alerts 2.3 Million Customers of Data Breach Tied to Leaky API. Retrieved from https://threatpost.com/t-mobile-alerts-2-3-million-customers-of-data-breach-tied-to-leaky-api/136896/
Statista Research Department. (2019, July 20, 2020). Total revenue of Target in the United States from 2012 to 2025. Retrieved from https://www.statista.com/statistics/299541/revenue-of-target-worldwide/
Steward, D., & Cavazos, R. (2019). Future Challenges and Recommendations. In *Big Data Analytics in U.S. Courts: Uses, Challenges, and Implications* (pp. 75-84). Cham: Springer International Publishing.
Stiennon, R. (2013). Categorizing data breach severity with a breach level index. In.
Stilgherrian. (2019, October 2, 2019). ANU incident report on massive data breach is a must-read. Retrieved from https://www.zdnet.com/article/anu-incident-report-on-massive-data-breach-a-must-read/
Storm, D. (2016, November 14, 2016). Biggest hack of 2016: 412 million FriendFinder Networks accounts exposed. *Computerworld*.
Stouffer, K., Pillitteri, V., Lightman, S., Abrams, M., & Hahn, A. (2015). 800-82 Rev 2. *Guide to industrial control systems (ICS) security*. doi:http://dx.doi.org/10.6028/NIST.SP.800-82r2
Stoyanov, N. (2019a, July 15, 2019). Personal data of millions of Bulgarian citizens and companies have leaked from the National Revenue Agency. *Capital*. Retrieved from https://www.capital.bg/politika_i_ikonomika/bulgaria/2019/07/15/3938624_ot_nap_sa_iztekli_lichni_danni_na_milioni_bulgarski/
Stoyanov, N. (2019b, July 19, 2019). The country with the most open data in the world. *Capital*. Retrieved from https://www.capital.bg/politika_i_ikonomika/bulgaria/2019/07/19/3940270_durjavata_s_nai-otvorenite_danni_v_sveta/
Suen, L.-J. W., Huang, H.-M., & Lee, H.-H. (2014). A Comparison of Convenience Sampling and Purposive Sampling. *Journal of Nursing, 61*(3), 105. doi:10.6224/JN.61.3.105
Suprema. (n.d.). Who We Are. Retrieved from https://www.supremainc.com/en/about/suprema.asp
Suprema Inc. (n.d.). Suprema Inc. Retrieved from https://www.linkedin.com/company/suprema-inc-
Sushko, O. (2017, September 21, 2017). Auto Tracking Company Leaks Hundreds of Thousands of Records Online. Retrieved from https://mackeeper.com/blog/post/auto-tracking-company-leaks-hundreds-of-thousands-of-records-online/
Sussman, B. (2019, June 3, 2019). Third-Party Risk Strikes Again: Hackers Access Millions of Medical & Billing Records. Retrieved from https://www.secureworldexpo.com/industry-news/third-party-breach-example-2019
SVR Tracking. (n.d.-a). About Us. Retrieved from https://www.svrtracking.com/about-us.html
T-Mobile. (2018). [Incident Announcement]. In.
T-Mobile U.S. (n.d.). T-Mobile. Retrieved from https://www.linkedin.com/company/t-mobile
T-Mobile US. (2020). T-Mobile Overtakes AT&T as America's #2 Wireless Provider and Continues to Deliver Industry-Leading Customer Growth with Strong Financial Results in Q2 2020 [Press release]. Retrieved from https://investor.t-mobile.com/news-and-events/t-mobile-us-press-releases/press-release-details/2020/T-Mobile-Overtakes-ATT-as-Americas-2-Wireless-Provider-and-Continues-to-Deliver-Industry-Leading-Customer-Growth-with-Strong-Financial-Results-in-Q2-2020/default.aspx
Target. (2014). Target Provides Update on Data Breach and Financial Performance [Press release]. Retrieved from https://corporate.target.com/press/releases/2014/01/target-provides-update-on-data-breach-and-financia
Target. (n.d.). all about Target. Retrieved from https://corporate.target.com/about
Taylor, H. (2018, September 18, 2018). NEWS & COMMENT: GOVPAYNOW.COM LEAKS 14M+ RECORDS — KREBS ON SECURITY. Retrieved from https://journalofcyberpolicy.com/2018/09/18/news-comment-govpaynow-com-leaks-14m-records-krebs-security/
Tech Advisor Staff. (2019, April 16, 2019). The Biggest Data Breaches. *Tech Advisor*.
Techopedia. (2019). Data Loss. Retrieved from https://www.techopedia.com/definition/29863/data-loss
TechTarget. (2008, February 2008). Definition: personally identifiable financial information (PIFI). Retrieved from https://whatis.techtarget.com/definition/personally-identifiable-financial-information-PIFI
Tham, I. (2018, July 20, 2018). Personal info of 1.5m SingHealth patients, including PM Lee, stolen in Singapore's worst cyber attack. *The Straits Times*. Retrieved from

https://www.straitstimes.com/singapore/personal-info-of-15m-singhealth-patients-including-pm-lee-stolen-in-singapores-most
The Canadian Press. (2019, December 10, 2019). Desjardins says employee who stole personal data also accessed credit card info. Retrieved from https://www.bnnbloomberg.ca/desjardins-says-employee-who-stole-personal-data-also-accessed-credit-card-info-1.1360652
The Home Depot. (2014). The Home Depot Completes Malware Elimination and Enhanced
Encryption of Payment Data in All U.S. Stores [Press release]. Retrieved from http://media.corporate-ir.net/media_files/IROL/63/63646/HD_Data_Update_II_9-18-14.pdf
The Intercept. (n.d., January 15, 2020). The Intercept Welcomes Whistleblowers. Retrieved from https://theintercept.com/source/
The Latka Agency. (n.d.). Canva. Retrieved from https://getlatka.com/companies/canva
Ticketmaster. (2018). INFORMATION ABOUT DATA SECURITY INCIDENT BY THIRD-PARTY SUPPLIER. Retrieved from https://security.ticketmaster.co.uk/
Ticketmaster. (n.d.-a). Ticketmaster. Retrieved from https://www.linkedin.com/company/ticketmaster/about/
Ticketmaster. (n.d.-b). Our Story. Retrieved from https://business.ticketmaster.com/our-story/
ticketmasteruk [@TicketmasterUK]. (2018, June 27, 2018). We have created a dedicated website about the recent data security incident, ... [Twitter Tweet]. Retrieved December 31, 2020 from https://twitter.com/TicketmasterUK/status/1011998001844379648
Timberg, C., & Merle, R. (2019, May 25, 2019). Security blog reports that First American left hundreds of millions of records exposed. *The Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2019/05/24/security-blog-reports-that-first-american-left-hundreds-millions-records-exposed/
Townsend, K. (2018, November 26, 2018). U.S. Postal Service API Flaw Exposes Data of 60 Million Customers. *Security Week*.
Townsend, K. (2019, August 15, 2019). Millions of Unencrypted Fingerprint and Facial Biometrics Found on Unsecured Database. *Security Week*. Retrieved from https://www.securityweek.com/millions-unencrypted-fingerprint-and-facial-biometrics-found-unsecured-database
Toyota. (2019). Notice regarding the possibility of customer information leakage at our Tokyo area dealers [Press release]. Retrieved from https://global.toyota/jp/newsroom/corporate/27465617.html
Toyota Motor Corporation. (2020). *FINANCIAL SUMMARY*. Retrieved from https://global.toyota/pages/global_toyota/ir/financial-results/2020_4q_summary_en.pdf
Toyota Motor Corporation. (n.d.). Toyota Motor Corporation. Retrieved from https://www.linkedin.com/company/toyota
Under Armour. (2018). Under Armour Notifies MyFitnessPal Users Of Data Security Issue [Press release]. Retrieved from http://investor.underarmour.com/news-releases/news-release-details/under-armour-notifies-myfitnesspal-users-data-security-issue?ReleaseID=1062368
United States Postal Service. (n.d.). About the United States Postal Service. Retrieved from http://web.archive.org/web/20181225051223/https://about.usps.com/who/profile/
UpGuard. (2018, April 18, 2018). Block Buster: How A Private Intelligence Platform Leaked 48 Million Personal Data Records. Retrieved from https://www.upguard.com/breaches/s3-localblox
US District Court at Seattle. (2019a). *Complaint for Violation of 18 U.S.C. § 1030(a)(2)*. Retrieved from https://www.justice.gov/usao-wdwa/press-release/file/1188626/download
US District Court at Seattle. (2019b). *Indictment No. CR19-159*. Retrieved from https://www.justice.gov/usao-wdwa/press-release/file/1198481/download
Venkat, A. (2020, March 5, 2020). Cathay Pacific Airlines Fined Over Data Breach. Retrieved from https://www.bankinfosecurity.com/cathay-pacific-airlines-fined-over-data-breach-a-13879
Vernier, N. (2018). Protecting customers from the Ticketmaster breach: Monzo's story. Retrieved from https://monzo.com/blog/2018/06/28/ticketmaster-breach
Vickery, C. [@VickerySec]. (2018, August 23, 2018). 14 million Texans exposed in new Voter Data breach discovery by @s7nsins [Twitter Tweet]. Retrieved December 7, 2020 from https://twitter.com/VickerySec/status/1032744727257395201
Vijayan, J. (2014, February 6, 2014). Target breach happened because of a basic network segmentation error. *Computerworld*.
vpnMentor. (n.d.). About vpnMentor. Retrieved from https://www.linkedin.com/company/vpnmentor/about/

Wagner, J. (2016, September 13, 2016). Reset those passwords — again: Over 6 million ClixSense users compromised by data breach. Retrieved from https://www.digitaltrends.com/computing/clixsense-hacked/
Waqas. (2017, September 23, 2017). Over Half a Million Vehicle Records from SVR Tracking Leaked Online. Retrieved from https://www.hackread.com/over-half-a-million-vehicle-records-from-svr-tracking-leaked-online/
Wells, M. (2017). Security Incident Response from SVR. Retrieved from https://www.svrtracking.com/security-incident-response.html
Welsh, S. (2019, January 17, 2020). Canva Security Incident – May 24 FAQs. Retrieved from https://www.canva.com/help/article/incident-may24
West, T., & Zentner, A. (2019). Threats and Major Data Breaches: Securing Third-Party Vendors.
Whitney, L. (2021, January 21, 2021). 2020 sees huge increase in records exposed in data breaches. *TechRepublic*.
Whittaker, Z. (2012, June 7, 2012). MD5 password scrambler 'no longer safe'. Retrieved from https://www.zdnet.com/article/md5-password-scrambler-no-longer-safe/
Whittaker, Z. (2016, November 13, 2016). AdultFriendFinder network hack exposes 412 million accounts. Retrieved from https://www.zdnet.com/article/adultfriendfinder-network-hack-exposes-secrets-of-412-million-users/
Whittaker, Z. (2018a, April 18, 2018). Data firm leaks 48 million user profiles it scraped from Facebook, LinkedIn, others. Retrieved from https://www.zdnet.com/article/data-firm-leaks-48-million-user-profiles-it-scraped-from-facebook-linkedin-others/
Whittaker, Z. (2018b, June 28, 2018). Inbenta, blamed for Ticketmaster breach, admits it was hacked. Retrieved from https://www.zdnet.com/article/inbenta-blamed-for-ticketmaster-breach-says-other-sites-not-affected/
Whittaker, Z. (2018c, August 27, 2018). Millions of Texas voter records exposed online. Retrieved from https://techcrunch.com/2018/08/23/millions-of-texas-voter-records-exposed-online/
Whittaker, Z. (2019, May 23, 2019). Millions of Instagram influencers had their contact data scraped and exposed. Retrieved from https://techcrunch.com/2019/05/20/instagram-influencer-celebrity-accounts-scraped/
Widup, S., Spitler, M., Hylender, D., & Bassett, G. (2018). *2018 Verizon Data Breach Investigations Report*. Retrieved from https://www.researchgate.net/publication/324455350_2018_Verizon_Data_Breach_Investigations_Report
Williams, C. (2019, February 11, 2019). 620 million accounts stolen from 16 hacked websites now for sale on dark web, seller boasts. Retrieved from https://www.theregister.com/2019/02/11/620_million_hacked_accounts_dark_web/
Wilson, Y., & Hingnikar, A. (2019). Failures. In *Solving Identity Management in Modern Applications: Demystifying OAuth 2.0, OpenID Connect, and SAML 2.0* (pp. 229-240). Berkeley, CA: Apress.
Winder, D. (2019, March 30, 2019). Toyota And Lexus Dealerships Hacked Leaving Millions At Risk -- What You Need To Do Now. *Forbes*.
Wong, S. K.-y. (2019). *Cathay Pacific Airways Limited and Hong Kong Dragon Airlines Limited - Unauthorised access to personal data of passengers* (R19-15281). Retrieved from https://www.pcpd.org.hk/english/enforcement/commissioners_findings/investigation_reports/files/PCPD_Investigation_Report_R19_15281_Eng.pdf
Wong, T. (2019). Characterizing the Harms of Compromised Genetic Information for Article III Standing in Data Breach Litigation. *Columbia Journal of Law and Social Problems, 53*, 48.
Xu, T., Zhang, J., Huang, P., Zheng, J., Sheng, T., Yuan, D., . . . Pasupathy, S. (2013). *Do not blame users for misconfigurations*. Paper presented at the Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles, Farminton, Pennsylvania.
Yin, R. K. (2013). *Case Study Research: Design and Methods*: SAGE Publications.
Yu, E. (2018a, July 20, 2018). Singapore suffers 'most serious' data breach, affecting 1.5M healthcare patients including Prime Minister. Retrieved from https://www.zdnet.com/article/singapore-suffers-most-serious-data-breach-affecting-1-5m-healthcare-patients-including-prime/
Yu, E. (2018b, September 21, 2018). SingHealth data breach reveals several 'inadequate' security measures. Retrieved from https://www.zdnet.com/article/singhealth-data-breach-reveals-several-inadequate-security-measures/

Yu, E. (2019, January 10, 2019). SingHealth breach review recommends remedies that should already be basic security policies. Retrieved from https://www.zdnet.com/article/singhealth-breach-review-recommends-remedies-that-should-already-be-basic-security-policies/
Zurkus, K. (2018, August 24, 2018). 99% of Texas Voter Records Exposed. *Infosecurity Magazine*.
Zurkus, K. (2019, June 21, 2019). Desjardins Insider Accessed Data of 2.9m Members. *Infosecurity Magazine*.