



POLITECNICO DI TORINO

Corso di Laurea in Ingegneria Informatica

Tesi di Laurea

# Investigation and Implimentation of dynamic data masking in financial services

**Relatore**

prof. Antonio Lioy

Alain Divin BAHIZI

**Supervisore Aziendale**

Vincenzo Cannata

ANNO ACCADEMICO 2022-2023



*Alla mia famiglia*

# Summary

Inserire qui un breve sommario della tesi.

# Contents

<b>1</b>	<b>Data Insecurity</b>	<b>6</b>
1.1	Introduction . . . . .	6
1.2	Understanding data breaches . . . . .	6
1.2.1	Why do data breaches happen? . . . . .	7
1.3	Statistics on data breaches . . . . .	7
1.4	Impact of data breaches . . . . .	8
1.5	Case studies of major data breaches . . . . .	8
1.5.1	The First American Corporation . . . . .	8
1.5.2	The city of Calgary data breach . . . . .	8
1.5.3	US voters records 2017 . . . . .	8
1.5.4	Solar wind . . . . .	9
1.6	Legal aspects of data breaches . . . . .	9
1.6.1	History context of data breach laws . . . . .	9
1.6.2	Penalties and compensation . . . . .	10
1.7	Conclusion . . . . .	11
<b>2</b>	<b>Data Masking</b>	<b>13</b>
2.1	Definition and significance . . . . .	13
2.2	Principles and Objectives of data masking . . . . .	13
2.3	Methods of data masking . . . . .	13
2.3.1	Static Data masking . . . . .	13
2.3.2	Dynamic Data masking . . . . .	13
2.4	Deep Dive into Dynamic Data Masking (DDM) . . . . .	13
2.5	Applications and Use Cases of Data Masking . . . . .	13
2.6	Tools and Tehnologies in Data Masking . . . . .	13
2.7	Best practices in implementating data masking . . . . .	13
2.8	conclusion . . . . .	13
	<b>Bibliography</b>	<b>14</b>

# Chapter 1

## Data Insecurity

### 1.1 Introduction

As our world becomes more and more connected, data breaches happen more frequently, not only in small business with no security specialists and budget but also for government and big businesses. In this chapter we are going to analyse and focus on data breaches, what causes them and what are their consequences in our society.

Picking from the real world scenarios like the first American Corporation data breach, the city of Calgary data breach or even the solar wind data breach, statistics shows that they happen more often than we might think so and could have devastating effects on your daily life.

Even though, it seems like data breaches are unavoidable or difficult to contain, there are solutions that can decrease considerably the risk of them. One of those solutions is Data masking. Data masking is a way of producing fake but convincing versions of the organization's data to protect it from unauthorized access.

From manipulation to shuffling, data masking uses different techniques to obfuscate the data within a database. This technique is very useful when you want your data to remain useful to other parties. We could take an example of developers who need to test their application to the database, they don't need the real sensitive data but the ones they can work with.

We will also discuss the legal aspects of data breaches, data masking is of greatest importance as masking becomes more and more important. Legislation surrounding data protection are in constant change. Later in the chapter, we will study the history and the evolvement of those legislation and how it enhances our privacy.

By integrating data protection techniques like dynamic data masking, entities not only fortify their defenses against breaches but also align more closely with emerging legal standards, potentially reducing liabilities.

In the following sections, while we explore the complexity and aftermaths of data breaches, it's essential to remember that tools and strategies of data masking can serve as formidable allies in the quest for enhanced data security.

### 1.2 Understanding data breaches

Data security is one of the most important topic for a company or any organisation. When your data is compromised by an unauthorised actor, we call it a data breach. In the occurrence of a data breach, malicious actors want to access and steal sensitive information which can range from fiscal code, credit card details, passwords, personal data...

In this chapter we are going to go in depth about data breaches, the impact it has on a personal and global level, how to prevent yourself and your business from it and the regulatory aspect of data breaches.

Data breaches can occur in many different ways, an organisation must always be cautious and updated to know how to protect their data.

### 1.2.1 Why do data breaches happen?

According to the World Economic Forum of September 2022, human error causes around 95% of the security incidents.[1] There are many factors that make a human the weakest link in security such as poor password or not updating it more regularly. Passwords are not the only point of entry in data breach but users also can share information with unauthorized individuals.

Aside from the human error, there are many other ways a data breach can happen, the most common example would be a malware that steals confidential information or encrypts the victim's data.

## 1.3 Statistics on data breaches

Data breaches has a serious impact on businesses and people and here we are going to see it's consequences. According to 2022 IBM data breach report[2], the average cost of the data breach is 4.35 million dollars. This is a 2.6% increase from last year and more than 10% increase from 3 years ago.

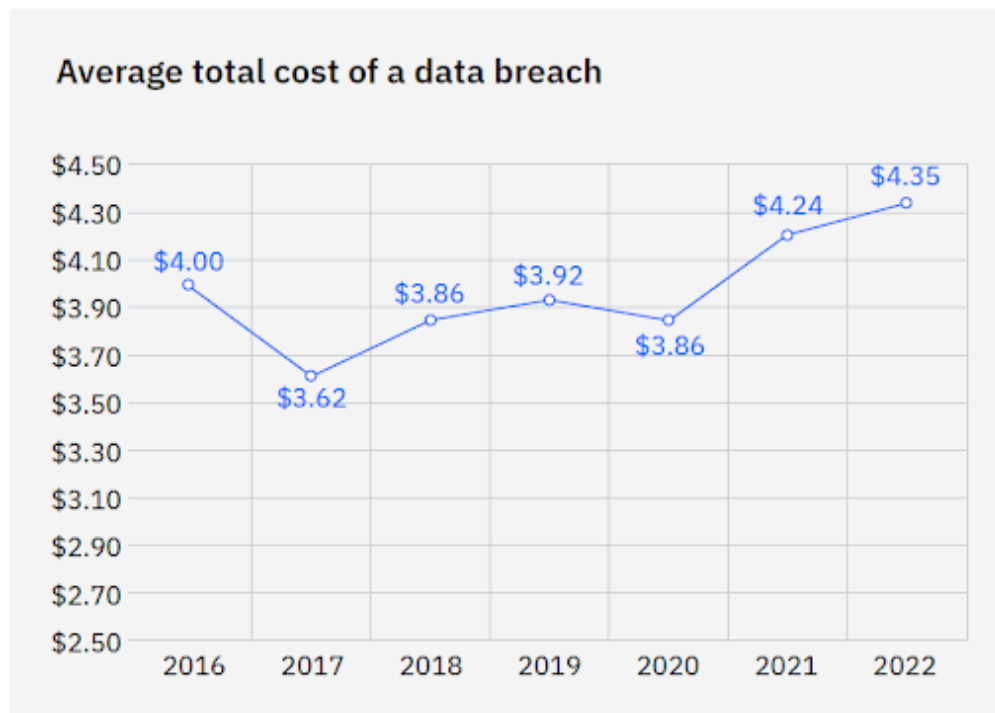


Figure 1.1. Cost of data Breach Report IBM 2022 (source: [Bitsight](#)).

Country wise, Italy is the 8th on the ranking list of countries affected by the data breach where the estimated amount of money lost is 3.74 million dollar from a value of 3.61 million dollars in the last year.

Research shows that data breach caused by poor security controls is responsible for billions of exposed records. During the year 2022, the exposed records were 19.81 billions according to the 2021 cyber Risk analytics breach review[3].

Most of these data breaches impacts 3 main sectors, the most affected industry is the healthcare industry where as of 2022, data breaches has cost 10 million dollars. Finance and technology industries follows up with a cost of 5 million dollars each.[4]

## 1.4 Impact of data breaches

Entities affected by a data breach suffers heavily. The main impact is revenue loss. These losses can be caused by downtime payment and audit fees. It can even lead to the reputation tarnish which would force clients to find reliable business offering the same service. Not only the loss of clients might lead to a revenue loss but also legal fees and regulatory fines are among the expenses caused by a data breach.

## 1.5 Case studies of major data breaches

In this section we are going to focus on different sectors affected by data breaches. Some was from human errors and others were misconfigurations.[5]

### 1.5.1 The First American Corporation

First American Financial Corporation is an American financial services company which provides title insurance and settlement services to the real estate and mortgage industries. As of May 2019, there was a data breach and compromised more than 885 million financial and personal records. The error was due to a web page not protected by an authentication process.

Among information obtained through this data leak, there was buyer's and seller's Names, ID, Social security Numbers, driver's licences, where they reside, Email Addresses and contact information. This is particularly dangerous as all these information could have been used for a phishing attack.

After this incident, The American Corporation was found to have breached cybersecurity protections regulations and The New York State Department of Financial Services fined them \$487,616.

### 1.5.2 The city of Calgary data breach

An employee from the city of Calgary has created a data breach when he unintentionally shared confidential information with another employee from a different municipality of the same province. The employee infringed a privacy violation sharing these information by email and the data was not encrypted.

This was not a malicious user though, as this leak was accidental. The employee needed technical advice and sent the data. This creates a serious risk as the data was not encrypted. The concerned individuals face a high possibility of identity theft and financial scam. [6]

### 1.5.3 US voters records 2017

This unprecedented data breach affected a data science company called Deep Root located in the United States. The company affected played a key role in President Donald Trump's electoral campaign. This data breach affected up to 198 million US citizens.

The breach was discovered by a security specialist who encountered an unprotected database containing details of those US registered voters.

Inside the database were information such as "names, birth dates, residential contacts, telecommunication details, and voter enlistment specifics". Not only personal details, the database also showed predictive models suggesting voter behaviors, policy affinities, and potential candidate support levels.

Upguard emphasized the troubling fact that this goldmine of information was not shielded in any way, making it a potential target for "anyone with online access." [7]



### 1.5.4 Solar wind

SolarWind is an American software company. It develops and manages a range of system management tools including but not limited to network and infrastructure monitoring. Its products are used world wide by a big number of organizations and companies. One of their most popular product is called Orion. It is an IT performance monitoring system tool.

As Orion is an IT monitoring system, it has privileged access to IT systems that allows it to collect logs and system performance data. These advantages make it a more profitable target for cyberattacks.

To add salt to the wound, this product is considered to be used by more than 30,000 public and private entities, including agencies to monitor and oversee their IT assets.

The data breach occurred when SolarWinds introduced a backdoor malware when updating the Orion software. This gave unauthorized access of the user's systems to threat actors.[8]

This incident compromised data, networks and system on a massive scale, affecting thousands of organizations.

The SolarWinds hack was a significant event, and its importance doesn't solely stem from the breach of one company. Instead, it had a much broader impact because it set off a chain of events that affected a vast network of organizations, including the U.S. government. This incident revealed vulnerabilities in the interconnected systems that many organizations rely on, emphasizing the need for enhanced cyber security measures and vigilance in protecting digital supply chains.

According to Solar Wind quarterly report of 2021 this data breach cost the company \$40 million.[9]

## 1.6 Legal aspects of data breaches

Data breaches often results out in legal case files. In this sub chapter we are going to analyse the history of data laws and some of the major data breach regulation that has changed the world of data protection either for the users or companies. Finally we will analyse some of the historical penalties and compensations imposed by those regulations to companies that failed to comply with them.

### 1.6.1 History context of data breach laws

Data breach law is not a new concept, it dates back in the nineteen hundreds with the U.N declaration of Human rights. In this declaration, it's article 12 states that "No one shall be subjected to arbitrary interference with his privacy,family,home or correspondence, nor to attacks upon his honour and reputation.Everyone has the right to the protection of the law against such interference or attacks".

In 1974, The US department of education elected the Family Educational Rights and Privacy acts (FERPA). This federal law protects the records of the students in all schools, universities and institutions.

Finally in 1995, the European Union adopted the Data Protection Directive that control how the companies handle personal data of their users. This is a more restrictive law compared to its counterpart of the US.

It was replaced in 2018 by the recent one, the GDPR.

The Health Insurance Portability and Accountability Act (HIPAA) from 1996 aims to simplify healthcare information processes, safeguard personal data in healthcare, and address health insurance restrictions.

California introduced laws in 2003 that made businesses and state agencies tell people when their personal data was at risk(if there have been a breach). Many other places in the U.S. and

around the world have since made similar rules based on California's example. This law was called a state data breach notification laws.

In 2018, the European union implemented the General Data Protection Regulation laws that protects the data and the privacy of in the European Union EU and the European Economic Area EEA. This law also goes into effect for the movement of data outside of EU and EEA.

In 2020, The state of California put in place a statute that controls the information and data of California residents and how businesses handle them. The statute was signed into a law and put into action on January 1, 2020. The law was called The California Consumer Privacy Act (CCPA).

The CCPA created a precedence to other states to create their own statute regarding privacy and data handling. In 2021, Virginia signed the same Consumer Data Protection Act into a law. The same was also signed in 2023 by the Colorado state.

### 1.6.2 Penalties and compensation

We are going to analyse penalties and fines given in Europe imposed by the GDPR.

The GDPR gave European Data Oversight agencies the power to charge companies up to 4% of their annual earnings for data mismanagement. Before this, the fines were inconsistent: Spain's regulatory body could charge a maximum of €600,000, France's CNIL set a cap of €150,000 for initial violations or up to €300,000 for subsequent ones, and the UK's 1998 law had a limit of £500,000, though it was later updated in 2018 to match the GDPR. Post-GDPR, the response across countries was varied: some were quick to impose hefty penalties, while others acted more conservatively. Notably, during this time, Croatia, Estonia, and Slovenia didn't publicize any GDPR-related fines.

The graph 1.2 shows how countries fines companies for data privacy non compliance or security violations.

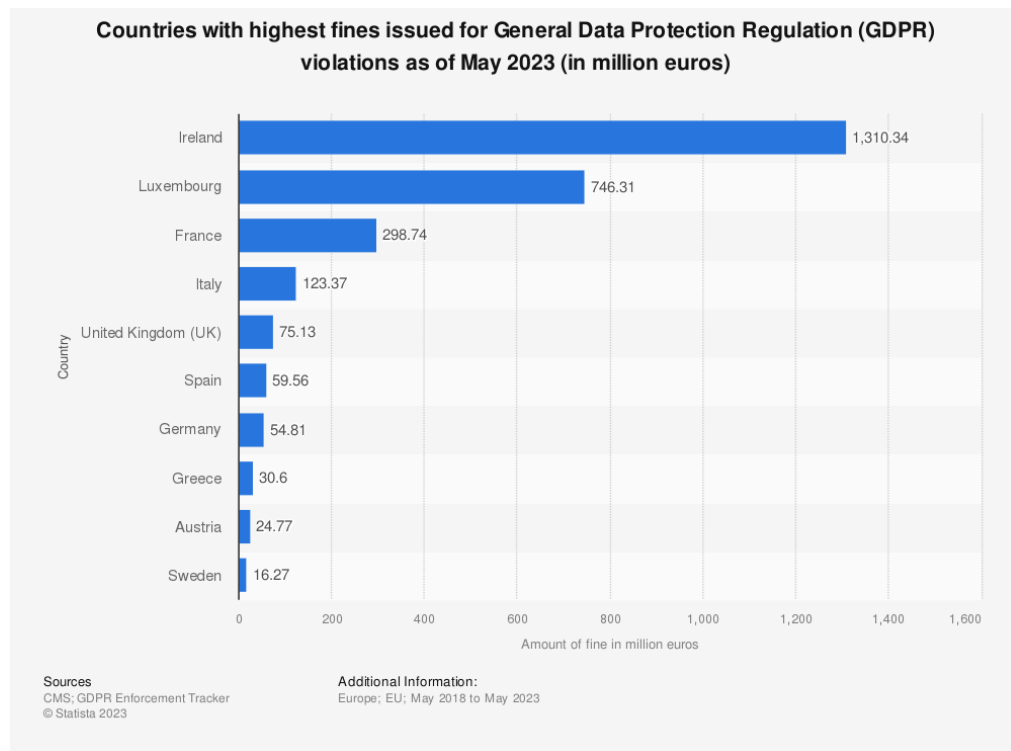


Figure 1.2. Fines issued for GDPR by country (source: [Statista](#)).

Ireland has a precedence in fining countries due to a historic €1.2 Billion on Meta (former

Facebook). This record breaking fine was imposed due to an insufficient data protection process that was used when transferring personal data of European users to the united states.

This fine is not the first one imposed to Meta as of 2022, the Ireland's Data Protection Commission imposed another €405 million due to the fact that the processing of children's personal data was not in order with the legal bases. Meta is not the only big company that is facing fines due to irregularities and failure to comply. Tiktok also faces a fine of €345 million due to the malpractice of how to manage account datas of children.

As for Luxembourg, over 95% of the fine issued by the Luxembourg National Commission for Data Protection (CNPd) was imposed to Amazon in relation to its processing of personal data and compliance with data protection laws.

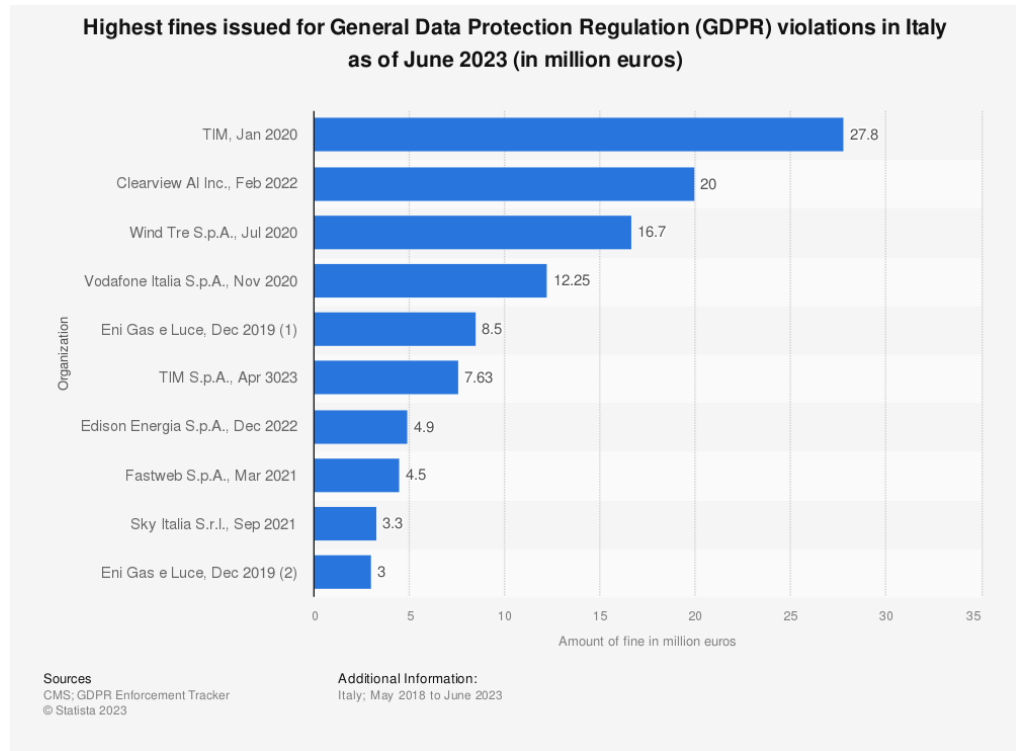


Figure 1.3. Italian historic Fines issued for GDPR (source: Statista).

As for Italy the figure below shows which company has been fined the most and how much it was fined since the implementation of the regulations of data protection.

As shown in the figure 1.3 The Italian data protection(Garante) has imposed fines on a lot of big name technology companies in the world. Recently it fined a sum of \$20 million to Clearview AI fine for non compliance. This company who owns a database of over 10 billion facial images worldwide had taken part in illegal surveillance activities with the country. Tech companies are not the only ones being targeted by this organization, as also the telecommunication company WIND had been issued by Garante a fine of €16 million. This fine was due to complaints from multiple individuals who received unsolicited marketing activities via calls and SMS.

## 1.7 Conclusion

In this chapter, we took a deep and detailed look on breaches, what they were, how they happen and its impact on both the company targeted, the clients and the society in general.

We studies major data breaches and what caused them and we took a look at the legal grounds of a data breach. Starting with the historical context up to statistics on penalties and compensation for both Europe and in Italy more specifically.

In legal actions we emphasized our study only on GDPR as it is the direction the thesis is going to take in later chapters and more importantly this thesis is the result of an internship held in the European Union where GDPR is the more prevalent data protection law.

## Chapter 2

# Data Masking

2.1 Definition and significance

2.2 Principles and Objectives of data masking

2.3 Methods of data masking

2.3.1 Static Data masking

2.3.2 Dynamic Data masking

2.4 Deep Dive into Dynamic Data Masking (DDM)

2.5 Applications and Use Cases of Data Masking

2.6 Tools and Tehnologies in Data Masking

2.7 Best practices in implementating data masking

2.8 conclusion

# Bibliography

- [1] ThriveDX, <https://thrivedx.com/resources/article/data-breach-types>
- [2] IBM, <https://www.ibm.com/downloads/cas/3R8N1DZJ>
- [3] flashpoint, <https://flashpoint.io/blog/what-are-data-breaches-how-to-prevent/>
- [4] upguard, <https://www.upguard.com/blog/cost-of-data-breach>
- [5] Biggest data breach in the financial industry, <https://www.upguard.com/blog/biggest-data-breaches-financial-services>
- [6] CalgaryHerald, <https://calgaryherald.com/news/local-news/class-action-lawsuit-claims-city-leaked-personal-information-of-3700-employees>.
- [7] DW, <https://www.dw.com/en/deep-root-analytics-behind-data-breach-on-198-million-us-voters/a-39318788>
- [8] SolarWind, <https://www.techtarget.com/whatis/feature/SolarWinds-hack-explained-Everything-you-need-to-know>
- [9] SolarWindcost, <https://www.cybersecuritydive.com/news/solarwinds-1-year-later-cyber-attack-orion/610990/#:~:text=For%20SolarWinds%2C%20the%20newly%20minted,quarterly%20report%20from%20October%20said>.