University of Zurich[UZH]

Burkhard Stiller,
Muriel Franco, Christian Killer, Sina Rafati,
Bruno Rodrigues, Eder Scheid, Rafael Ribeiro, Alberto Huertas,
Jan von der Assen, Eryk Schiller (Edts).

Internet Economics XV

TECHNICAL REPORT – No. IFI-2022.01

January 2022

University of Zurich
Department of Informatics (IFI)
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

ifi

B. Stiller et al. (Eds.): Internet Economics XV
Technical Report No. IFI-2022.01, January 2022
Communication Systems Group (CSG)
Department of Informatics (IFI)
University of Zurich
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland
URL: http://www.csg.uzh.ch/

Introduction

The Department of Informatics (IFI) of the University of Zurich, Switzerland works on research and teaching in the area of computer networks and communication systems. Communication systems include a wide range of topics and drive many research and development activities. Therefore, during the autumn term HS 2021 a new instance of the Internet Economics seminar has been prepared and students as well as supervisors worked on this topic.
Even today, Internet Economics are run rarely as a teaching unit. This observation seems to be a little in contrast to the fact that research on Internet Economics has been established as an important area in the center of technology and economics on networked environments. After some careful investigations it can be found that during the last ten years, the underlying communication technology applied for the Internet and the way electronic business transactions are performed on top of the network have changed. Although,a variety of support functionality has been developed for the Internet case, the core functionality of delivering data, bits, and bytes remained unchanged. Nevertheless, changes and updates occur with respect to the use, the application area, and the technology itself.Therefore, another review of a selected number of topics has been undertaken.

Content

This new edition of the seminar entitled "Internet Economics XV" discusses a number of selected topics in the area of Internet Economics.
The first talk, Talk 4, provides an overview of malware affecting the Internet of Everything (IoE), highlighting its economic and physical impacts. Talk 5 introduces the concepts of crypto vaults and explores the best practices for designing a vault for crypto exchange centers. Furthermore, it shows the security measures that reputable exchanges offer their customers. Talk 6 explores the cutting-edge topic of Decentralized Finance (DeFi), explaining how DeFi is gaining relevance as well as challenges and opportunities in this field. Talk 7 provides an overview of Decentralized Exchange (DEX) platforms, explaining the flow of funds in different order types and comparing DEX with traditional Centralized Exchange (CEX) platforms. Talk 8 discusses the growing market of digital espionage and forensics, analyzing forensic solutions on the Internet of Things (IoT), cloud computing, and mobile devices. In addition, it presents current players in this market, their products, and customers. Talk 9 brings the state of the General Data Protection Regulation (GDPR), highlighting its direct costs, regulations, and fines, focusing on how the companies experience the result of the data breaches and the GDPR. Talk 10 elaborates on the impact of the remote working economy on cybersecurity. It highlights the first problems companies faced when the COVID-19 pandemic started, lists major cybercrimi-

nal types and attacker techniques during the pandemic, and discusses the countermeasures from a technical level. Finally, Talk 11 discusses public surveillance, assessing the pros and cons as well as the associated economics and enabling technology for different areas of public surveillance.

Seminar Operation

Based on well-developed experiences of former seminars, held in different academic environments, all interested students worked on an initially offered set of papers and book chapters. Those relate to the topic titles as presented in the Table of Content below. They prepared a written essay as a clearly focused presentation, an evaluation, and a summary of those topics. Each of these essays is included in this technical report as a separate section and allows for an overview on important areas of concern, sometimes business models in operation, and problems encountered.
In addition, every group of students prepared a slide presentation of approximately 45 minutes to present its findings and summaries to the audience of students attending the seminar and other interested students, research assistants, and professors. Following a general question and answer phase, a student-lead discussion debated open issues and critical statements with the audience.
Local IFI support for preparing talks, reports, and their preparation by students had been granted by Muriel Franco, Christian Killer, Sina Rafati, Bruno Rodrigues, Eder John Scheid, Eryk Schiller, Rafael Hengen Ribeiro, Jan von der Assen, and Burkhard Stiller. In particular, many thanks are addressed to Rafael Hengen Ribeiro and Jan von der Assen for organizing the seminar and for their strong commitment on getting this technical report ready and quickly published. A larger number of pre-presentation discussions have provided valuable insights in the emerging and moving field of communication systems, both for all groups of students and supervisors. Many thanks to all people contributing to the success of this event, which has happened in a lively group of highly motivated and technically qualified students and people.

Zurich, January 2022

Contents

Chapter 4

Malware Affecting the Internet of Everything (IoE): Economic and Physical Impact

Janik Lüchinger, Marinja Principe

The Internet of Everything (IoE) and the Internet of Things (IoT) are hot topics in today's research. Over 20 billion devices are already interconnected, using consistent data flow to provide smart solutions. There are already different reviews which are looking into IoT and related cybersecurity issues as well as analyzing IoE and how to provide a more secure environment. Most of them agree that IoE will be a security challenge. The complexity of the large scale of devices needs to be thought through. However, the physical and economic impacts resulting from a malware infection of the IoE device are not yet addressed. This work analyzes the physical and economic impacts different malware has on IoE devices based on four promising use cases: Continuous Glucose Monitoring, Brain-Controlled Artificial Limbs, Brain-to-Vehicle, and Brain/Cloud Interface. Each use case will describe how the system is set up, investigate if the most common types of malware can affect it, and, if yes, what the physical or economic impacts are. This review shows that various types of malware can affect the use cases, which can lead to significant physical and economic impacts for the user and the company providing the IoE device.

Contents

4.1 Introduction

The usage of the Internet is expanding daily. The Internet of Things (IoT) is an approach to include the Internet in our professional, social, and personal lives [1]. IoT devices are beyond machine-to-machine connections. Holler et al. [2] describe it as an array of applications, domains, and networking protocols. Physical devices are linked to the Internet and interconnect devices with each other. Smart homes, eHealth, smart manufacturing, and self-modifying mechanisms are some examples that will have a significant impact on the industry and the personal life of everybody [3]. This rapid evolution had resulted in a constant communication flow between devices via networks and other communication systems.
The inclusion of the IoT into our daily life opened new cybersecurity challenges. Lu and Xu [3] claim that attackers could take advantage of the interconnectivity and data flow if these challenges are not appropriately faced. According to Duan et al. [4], there exist 20 billion devices around the world that are part of an IoT-based connection. This number shows the importance of the security of IoT. The more devices of people are involved, the greater is the risk of the individual, and the higher is the cybersecurity risk for the network [1]. Miraz et al. [1] believe that the challenges of IoT contain a standardized protocol, the implementation of IPv6, and power supply. Some countries have started paying attention to IoT, implementing standards, and adjusting the laws to increase cybersecurity.
There is a slow transition from the IoT into the Internet of Everything (IoE). IoE does not only include things but also people, data, and processes [1]. These four pillars determine how the Internet is used in our daily life. People are interconnected with devices, using many data and optimized processes. This new extension from IoT to IoE enriches the life of the user. As a result, the number of devices, including IoE, is increasing. A large number of interconnected devices exist, which can extract and analyze data in real-time to provide automated processes. A necessity of IoE is cloud computing. "Everything" will be connected online [1], with the Internet as the "heart and center of IoE" [5]. With new possibilities, there are new challenges. With the extension of data and the cooperation of humans with machines, the security risk is getting bigger. The Internet is the "heart of the IoE," [5] which means that all security threats lie within it. The more devices are connected, the more vulnerable entry points it has. Chinchawade claims that IoE devices' characteristics include "extremely large scale, low-cost design, resource constraints, device heterogeneity, preference of functions over security, higher privacy requirements, and harder trust managements" [5]. Not only people but also cities can benefit from IoE. The connections in "smart cities" and information from "big data" can be used to help with city-specific concerns [1]. Examples are smart streets and agricultural growth monitoring. IoE also helps in the mining industry of fossil fuels in monitoring and improving safety.
However, although there is a big request for IoE, there is not yet enough research to answer how to ensure security for the user when using it. Some works are looking at Brain-Computer Interface (BCI) and how to maintain security [6], and others analyze different security frameworks as the four layers of cybersecurity discussed by Lu and Xu [3]. Nevertheless, only a few works are looking at the physical or economic impact of an affected IoE device, and even fewer are looking at specific use cases. These use cases are beneficial for showing in detail what risks and impacts could appear in this particular situation.
In order to cover the previous limitations and challenges, this work analyses different malware affecting the IoE and their physical as well as economic impacts on involved stakeholders. Four promising use cases of IoE were selected: Continuous Glucose Monitoring, Brain-Controlled Artificial Limbs, Brain-to-Vehicle, and Brain/Cloud Interface. They cover a variety of IoE usages such as health monitory, supporting disabilities, transport, and BCI. The use cases Continuous Glucose Monitoring [7] and Brain-Controlled

Artificial Limbs [8] have already working prototypes while others like the Brain-to-Vehicle [9; 10] and Brain/Cloud Interface [11] are still theoretical models. In this work, each of them is described, their IoE-related characteristics are reviewed, and arguments of why they belong to IoE are presented. These four use cases in the branch of IoE are then analyzed on different malware. Finally, for each malware, it will be discussed if it has relevance for the use case and, if yes, whether there is a risk of physical or economic impact.
The remainder of this document is structured as follows. In the beginning, Section 4.2 discusses the definition of IoT and IoE, as well as the definition of the related malware. After that, Section 4.3 looks into related work on IoE and already known cybersecurity issues. Next, Section 4.4 describes the four use cases Continuous Glucose Monitoring, Brain-Controlled Artificial Limbs, Brain-to-Vehicle, and Brain/Cloud Interface and analyzes each use case individually. Subsequently, Section 4.5 compares all the use cases to formulate a summary. At last, Section 4.6 presents an outlook on how future work could improve on this work.

4.2 Background

This section contains two subsections. The first one elaborates on the evolution of the IoE, the main differences of IoE to the IoT, and captures some different IoE contexts to consider for the analysis later on. The second subsection investigates software faults and the different types of malicious software, as well as possible physical and economic impacts the malware can have on an infected device, its owner, manufacturer, or any other involved stakeholder.

4.2.1 The Internet of Everything

Before properly explaining the upcoming IoE, we must first look at its ancestor: the IoT. Unfortunately, no universal definition could clearly state what IoT is. Instead, it is "an umbrella term which has a broad range of technologies, applications and variety of interconnected electrical or electronic devices" [12]. IoT represents the latest global standard of device interconnection involving machine-to-machine (M2M) communication, enabling the integration of input from the most miniature sensors and detectors into working industry standards and production processes. The processes are not changed when integrating with IoT devices, though. Instead, the goal is to replace manual verification with automated and frequently updated sensor input for dynamic or rule-based decision-making. These core concepts are depicted in figure 4.1.
In IoT, every device is connected to a network and receives its own unique IP address for targeted communication. With this, it can transmit its results to — and receive and process incoming data from — other nodes within the designated network. Integrated into this continuous flow of interchanged data, the individual IoT device contributes to achieving its particular goal and any overarching goals of the entire system alike [12].
Examples of IoT include specific applications (relying on sensor input of IoT devices collected over the network), advanced concepts combining multiple devices and software to an intelligent system (e.g., smart security), or more end-user-friendly and straightforward IoT wearables (like fitness monitors containing sensors and some rudimentary software).
In contrast, the IoE is the next stage in the evolution of the Internet. Figure 4.2 visualizes the IoE as an extension of the IoT: The IoE includes not only "things," unlike the IoT, but also people, processes, and data, which overall creates new opportunities and possibilities. Accordingly, one could argue that the IoT is a proper subset of the IoE. The intelligent interconnection between things, data, people, and processes aggregates simple information and transforms it into actions. In doing so, apart from machine-to-machine (M2M), IoE

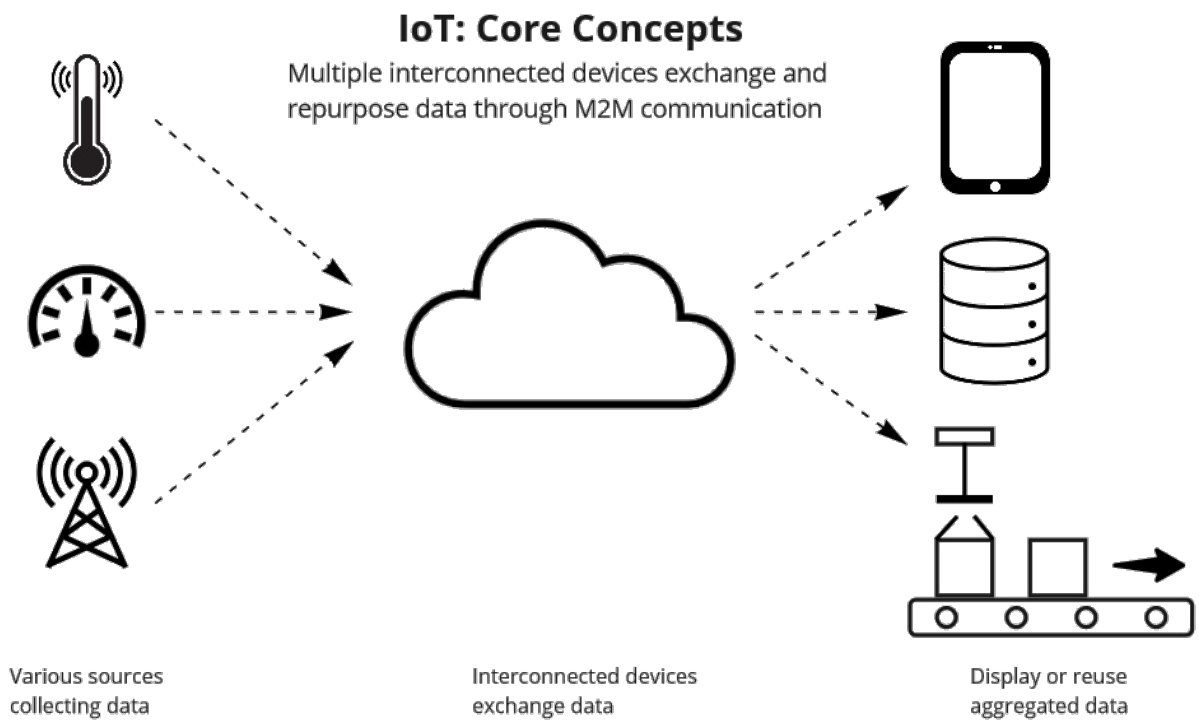

Figure 4.1: Core concepts of IoT.

also entails machine-to-people (M2P), people-to-machine (P2M), and people-to-people (P2P) communication.
Examples of IoE include more advanced, interconnected concepts and systems like connected smart homes (where multiple IoT devices contribute to process integration based on human involvement or daily routine) or smart cities (for which processes are optimized with the help of current and historical data combined, to connect roads with hospitals to achieve the most effective routing based on traffic data for example), to only name two of the most well-known and easily visualized examples. Nevertheless, it is also important to mention that not every use case is black or white: taking a FitBit as an example, it has IoT wearable characteristics (i.e., at most M2M communication). It can, however, be argued that the FitBit also contains M2P communication and different kinds of data, especially when expanding the scope to any connected mobile device like a smartphone. For example, such a mobile device would provide applications to visually present a history of collected fitness data to the user, make propositions on lifestyle adjustments based on data trends, etc.
When focusing on the characteristics that separate IoE from IoT, we must further specify what is meant by mentioning people, data, and processes. In the case of people, human beings can be viewed as interconnected devices. Related communications can manifest as P2M, M2P, or P2P, depending on the specific use case and involved components. Most importantly, people are the central point of IoE because, after the evolution from IoT to IoE, not only devices but also humans can be connected. One possible example would be a device connecting to a human brain and taking its activities as input to adjust the resulting behavior during operation.
IoE further integrates data as one major, evolving component. In these modern times, the number of available data sources grows continuously, maybe even exponentially. After analysis and classification, consolidating data in IoE allows deriving meaningful information that will enable new intelligent decisions and provide the basis for even more exciting studies. Following the previously mentioned example of linking devices to a human brain, collecting more (historical) data regarding end-user brain activities and usage will help optimize controlling devices with our brain alone. The struggle of designing an intuitive

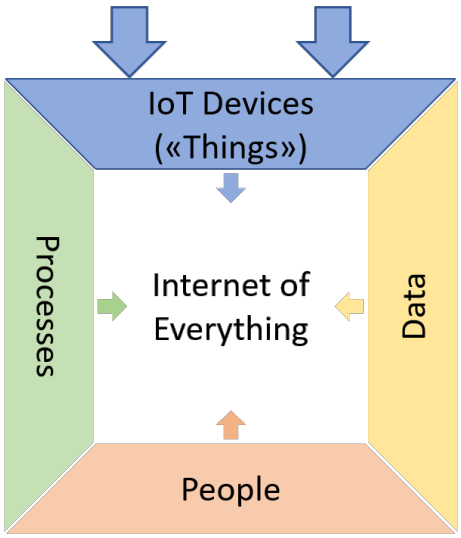

Figure 4.2: The four main aspects of IoE: things, data, processes, and people.

user interface would be eradicated since there would not be the need for a compelling (physical or digital) interface in the first place.
The last step would be to include (and hopefully optimize) existing processes into the IoE environment by combining collected data with any involved people. Processes are the second most crucial part of IoE, given that they can heavily influence the way and the efficiency with which data is collected, generated, or utilized. The idea is that processes maximize the value of any system in the digital world, which is why making them as efficient as possible is necessary to provide smart solutions to existing or upcoming problems. The example of a brain-operated device would correspond to a smart home environment wherein the different daily processes are optimized. An inhabitant could use the smart IoE technology to the fullest — by either simply thinking of a task to be completed or not having to think of it at all because the smart home already knows from previous days that said task must be fulfilled.

4.2.2 Software Malfunctions and Malware

In the following, the notion of software faults will be introduced and the understanding of the different types of malicious software will be defined, i.e., Trojan horse, worm, virus, ransomware, rootkit, bot, and APT. Subsequently, two forms of possible impacts the malware can have on an infected device, its owner, manufacturer, or any other involved stakeholder will be defined. The focus of this work is specifically on physical and economic impacts.
Malware: Malware is the short form of "malicious software." It is usually used as an umbrella term for any software intentionally designed to cause damage to a computing device [13] or secretly act against the computer user's interests [14]. Your device can become infected with malware through any connection, be it from the internet or simply from within a network containing another infected device. Other possible sources include offline physical device access, like a USB stick or a CD-ROM. Although there are several different forms of malware (e.g., Trojan horse, ransomware, spyware, scareware, adware [15]), this work will only focus on the few forms defined in this section.
Software Fault: Unlike malware, software that causes unintentional harm due to some deficiency is typically described as a software bug or fault [16]. It includes software that does not behave according to the behavior defined in its description. The malfunction must be unintentional; otherwise, it could be identified as malware. More specifically, the distinction between software errors, bugs, or faults and their respective time of identifica-

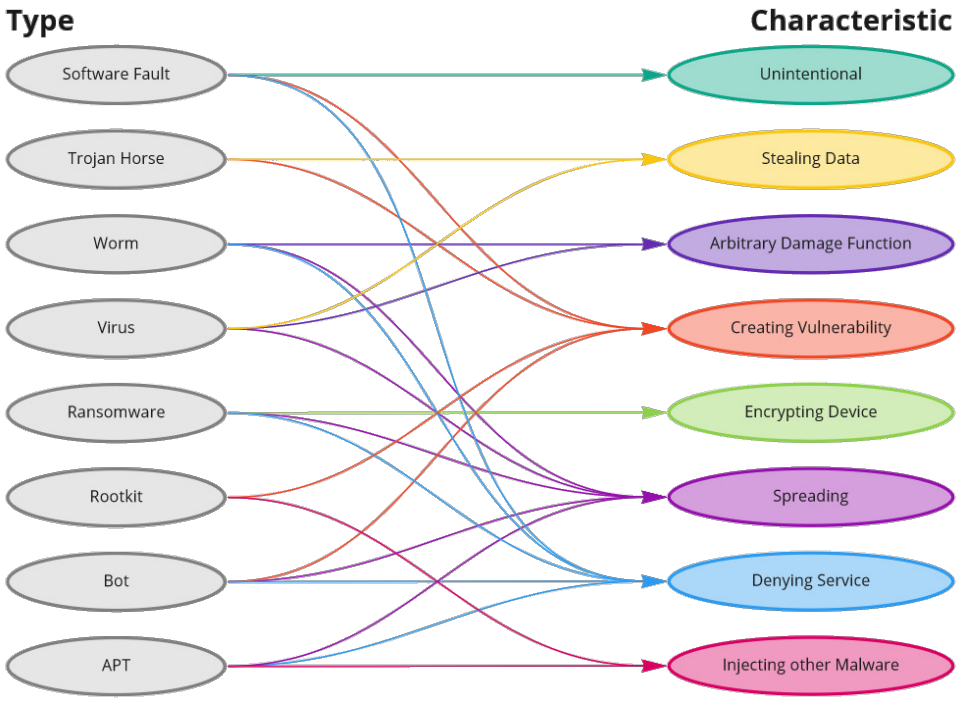

Figure 4.3: An overview of malware types and associated (malicious) characteristics.

tion in the software lifecycle will not be further discussed; instead, it will be simplified to use the three terms interchangeably as software issues, particularly termed software faults.
Trojan Horse: A type of software that includes hidden (often malicious) functions neither specified nor documented. The goal of a Trojan horse is to compromise a system or install a backdoor [17]. Usually, a Trojan horse cannot replicate itself. However, it is essential to mention that a machine first gets infected and then downloads the malicious software ("from the inside").
Worm: This kind of malware is designed to replicate itself in any network or distributed system [18] and execute its payload without the need for interaction. A worm can contain an arbitrary damage function, but this is not required. Accordingly, a worm is defined as "dual-use" because of its ability to be used for good (e.g., a network administrator automatically distributing software updates to all of the nodes contained in the network) or bad (i.e., the malware).
Virus: Contrary to a worm, a computer virus cannot be executed by itself [19]. Instead, it requires a victim program for execution and active human interaction for replication. The goal of a virus is to spread as far as possible and create new infections. Usually, a virus comprises three parts:

- the replication part (including scanning for already infected software to avoid multiple infections and increasing the risk of detection),
- an arbitrary damage function (including a trigger, e.g., time-based or action-/event-based),
- and the camouflage part with the primary purpose of staying undetected (e.g., a key logger).

Ransomware: A virus for which the damage function is an encryption tool that encrypts the storage medium of the infected device [20]. According to Wyke and Ajjan [21], there are two main ransomware types: the first one is crypto ransomware, and the second one is locker ransomware. The crypto ransomware can encrypt the affected files and deny access. The victim needs to pay a ransom to obtain the decryption key and be able to

use the files again. The locker ransomware locks the device and denies access to it [22]. To be able to use the device again, the victim also needs to pay a ransom.
Bot: A bot is usually referred to as a piece of software that allows an infected computer system to be remotely controlled. The vital characteristic of a bot is that it must not hinder the original functionality of the device in any way, so the chances of detection are as low as possible. Multiple bots form a botnet, i.e., a network of synchronized bots. The botnet is typically controlled by a single node (the "botmaster") and is used to perform distributed denial-of-service (DDoS) attacks.
Rootkit: The rootkit is a set of software tools installed after successfully compromising a computer system to hide future accesses or processes and files. That is commonly referred to as "installing a backdoor." The aim is to hide the existence of malware against a user or antivirus software by providing legitimate access to someone who initially was not intended to be accessing the infected system. A rootkit is often injected after an infection with a Trojan horse [19].
Advanced Persistent Threat: An advanced persistent threat, or APT for short, does not refer to a new class of software anomalies and manipulations but rather to a new type of use. Traditional malware attacks would mostly have only one attack vector targeting the entire population without any specific targets. In contrast, an APT maintains many different attack vectors for organizing multiple attacks targeting a single target and achieving long-term utility [23].

4.2.3 Impacts of Software Malfunctions and Malware

A software malfunction and the malware discussed above always have some impact on the device or the user of the device. In order to analyze the effects, the meaning of "impact" must be defined. Two types of impacts will be considered in this work: a physical and an economic impact.
Physical Impact: The term physical impact refers to the effect on a single person or group that directly impacts their lives. A physical impact can also include secretly processing personal data.
Economic Impact: An economic impact is a financial effect that something has on a situation, person, or company. The financial magnitude of the impact is irrelevant for it to be counted as an economic impact.

4.3 Related Work

There already exists some related work, which discusses IoT and IoE and their security issues. Since these papers are necessary for the analysis done in this work, they will be reviewed and summarized below.
Miraz et al. [1] discuss three different stages of the Internet: the IoT, IoE, and the Internet of Nano Things (IoNT). They define IoNT as an extension of IoE by cooperating nano-sensors as well as nano-networks. IoNT will potentially enable new medical possibilities. Miraz et al. believe that we will experience "a period of transition for novel interactions, ubiquitous computing, mobile and ambient intelligent applications and so forth in the remainder of the 21st Century" [1]. They also think that the distinction between IoT, IoE, and IoNT will help to predict the near future.
Lu and Xu [3] are looking into the cybersecurity aspects of IoT. They split the IoT security framework into three subparts: 1) basic three-layer architecture, 2) derived four-layer architecture, and 3) detailed five-layer architecture. In their work, they define the main four layers as listed in figure 4.4: sensing layer, the network layer, the middleware

Layers	Description	Attack Types
Sensing	Sensing objects and data. Attack focus: confidentiality	Replay Attacks, Timing Attacks, Node Capture Attacks, Malicious Data Attacks, SCA (Side Channel Attack)
Networking	Networking and data transmission. Attack focus: confidentiality, privacy, and compatibility	Spoofed, altered or replayed routing information, Sybil, Wormholes
Middleware	Data delivery. Attack focus: authenticity, integrity and confidentiality	Malicious Insider, underlying infrastructure, third-party relationships, virtualization threat
Application	Requested service provision. Attack focus: data privacy and identity authentication	Phishing Attack, Virus, Worms, Trojan Horse and Spyware, Malicious Scripts, Unauthorized Access

Figure 4.4: Four layers of cybersecurity by Lu and Xu [3, Table 2].

layer, and the application service) layer [3]. They discuss relevant cybersecurity issues for each of the layers.

Apart from Miraz et al. [1], Lu and Xu [3] also define standardization as a necessity for IoT. The architecture of the interfaces, data models, and protocols is a significant step to ensure security in IoT. Data confidentiality and configuration issues were stated as the leading security issues of IoT.

Chinchawade and Lamba [5] are looking into security issues in the IoE. They claim that all the security threads lie on the Internet, as it is the "heart of IoE" [5]. It will become more and more challenging to provide security because of the large scale of IoE, device heterogeneity, and preference of functions over security. They speak about different types of attacks: the Perception layer, Network layer, Middleware layer, and Application layer attacks. To ensure security, they are looking into the architecture of IoE. With the help of protocols as Quick UDP Internet Connections (QUIC) and Advanced Message Queuing Protocol (AMQP), or authentication schemes as Device Authentication scheme (AK) or Peer-to-Peer Authentication, IoE should get more secure to use in the future.

Masoud et al. [24] take a slightly different perspective on the matter. They define IoE as a combination of sensing, computation, information extraction, and communication in one device. Examples would be smartphones, smart fridges, or cars, as they have nodes that help them sense the environment and process data to regain and communicate information. Masoud et al. claim "IoE is a complex approach with massive applications, dreams, and myths" [24]. In their work, they show that smart device sensors could leverage practical applications, but also in hacking and attacking issues. They also discuss how these threats can easily be implemented and deployed without any programming skills. An important point mentioned is awareness which can prevent hidden data issues. Additionally, the user should get more control over his device, enhancing more warning messages and the possibility to choose different sensors. However, their work mainly focuses on a smartphone and not IoE in general.

Ajrawi et al. [25] write about a BCI, which converts brain signals into machine language. This communication can be bidirectional or unidirectional, and therefore the cybersecurity has high importance. They propose "the design of a security system that is based on RFID technology which utilizes EPCglobal Network that improve the communication between the implanted RFID sensors and the external readers to secure patient brain

activities" [25]. Additionally, they mentioned that the BCI's main requirement should be an identification of the device controller.
Bernal et al. [6] discuss what cyber security issues are faced in BCI. They look into a wide range of different types of attacks as cryptographic but also Battery or Social engineering attacks. BCI is at the mercy of all these attacks. However, according to Bernal et al., Malware attacks have the most significant impacts on BCI. Bernal et al. [6] not only look into different kinds of attacks, but they also discuss the countermeasures. An important point is the initiation of the awareness of the user. In order to reach this goal, training sessions and demos are necessary. The user's awareness makes it more difficult for attacks as social engineering or affection of the BCI devices through malicious emails or websites. Furthermore, accesses control mechanisms and authentication verification should block unauthorized attackers from entering the device. They also mention proactive mechanisms such as periodic system updates, antivirus, and monitoring systems to prevent the devices from attacks and react as fast as possible if they are affected anyway.
All of the papers mentioned above look at the structure of IoE and how to ensure cybersecurity. They believe that it gets more complicated to provide security with the increasing number of devices. Therefore, the probability is high that there will be no 100 percent secure IoE environment. Nevertheless, non of them is looking at the impact it can have on the user physically and economically if the system gets affected by malware. To fill this gap in the discussion of IoE, this work will discuss the main malware types and their physical and economic impact on the user, using the four promising use cases mentioned above.

4.4 Analysis of Use Cases

Given the vastness of the Internet with all its variations of different connections, IoE includes too many aspects to analyze in just one paper. Accordingly, the research and the analysis focus on four promising applications of the IoE: continuous glucose monitoring in diabetic patients, brain-controlled artificial limbs, self-driving cars with a brain-to-vehicle interface, and a universally usable brain-to-cloud interface.
This work will present an analysis of underlying cybersecurity issues concerning different types of malware and the resulting physical and economic impacts on involved stakeholders such as users, manufacturers, or attackers for each of the four use cases.

4.4.1 Continuous Glucose Monitoring

This use case assesses a new approach for Continuous Glucose Monitoring (CGM) in diabetic patients with type 1 diabetes. Following the solution proposed by Ismail [7], a given diabetic patient would receive a tiny chip implanted within the patient's body, e.g., in his upper arm. The chip is a glucose meter that continuously senses the glucose levels in the blood, processes the measurements, and transmits the results to a connected mobile device.
According to the scenario presented, the mobile device is a smartwatch that measures the patient's fitness level and respiration rate. The smartwatch is connected to the Internet and thus capable of forwarding any measurements (including their calculated severity) to a designated health monitoring company (HMC) via current mobile communication networks infrastructure, e.g., 3G to 5G mobile networks [7]. For the mobile device to process and forward incoming data from the glucose meter — or health data it collected on its own — the patient would most likely have to install a dedicated diabetes monitoring application on the device. When properly implemented, this application could also collect historical medical data of the patient and nicely present the bio values in the form of

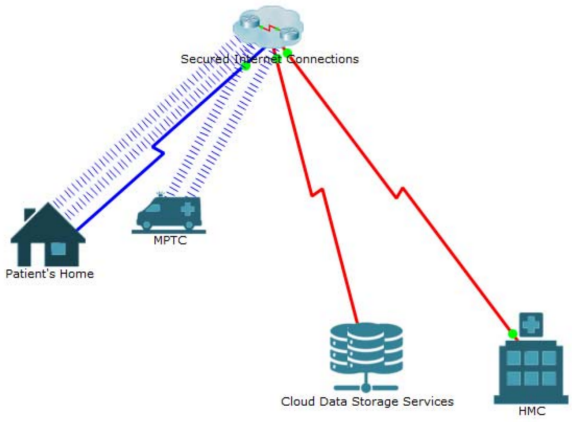

Figure 4.5: Diabetic patient monitoring IoE simulation model by Ismail [7, Table 3].

medical history. Like this, the patient could perform a preliminary analysis of his lifestyle and apply minor adjustments by himself without consulting a doctor.
The most important feature of this connected approach is the ability of the CGM implant and mobile device to automatically trigger a medical emergency if any of the bio values fall outside of a predefined range, posing a potential risk to the patient's health. The assigned HMC would receive the emergency and initiate a communication attempt (i.e., a call) with the patient directly on his mobile device requiring personal interaction. If the patient fails to acknowledge the communication request, there is a high possibility that he may be incapacitated and already unable to respond. Accordingly, the HMC would immediately dispatch a mobile patient treatment center (MPTC; in essence, a medical emergency response team) [7], as shown in figure 4.5.
When considering a restricted area in an office building or private apartments, the MPTC may stand in front of locked doors. Ismail foresaw this problem and proposed a possible solution in his work: a "smart door" with an electronic lock supporting access by a one-time door unlock key. The designated HMC would be allowed to send such a one-time key to the MPTC for them to be able to open any smart door in their path [7].
Moreover, it would be possible to integrate the system with an external "smart city" environment allowing for quick and optimized routing of the MPTC. This integration would make it possible to reach the patient's current location even faster in the case of an emergency when immediate assistance is required [7]. However, every API newly introduced to such an external system must be considered carefully and permitted access rights clearly defined.

4.4.1.1 Motivation for IoE

The idea of continuously monitoring the glucose levels in the blood of diabetic patients is not new. There exist solutions for continuous monitoring involving implanted sensors and connected mobile devices, collecting data that can then be read and evaluated by designated applications [26] or doctors during the next patient visit. It may even be argued that such solutions are (at least partly) belonging to the IoT since the monitoring and evaluation mainly include M2M communications without any human involvement. Still, the CGM system proposed by Ismail represents an evolution towards IoE.
Communication. The system contains direct P2P communications in addition to communications with or between machines. For example, there are interactions between doctors, patients, the medical staff, and the medical emergency response team.
Following a triggered medical emergency, the alarm acknowledgment includes P2M and M2P communication directly involving the patient. Additionally, the one-time door unlock mechanism for "smart doors" includes P2M communication between the medical

emergency response team and the tablet or smart door (e.g., typing a passcode in a code panel or using the tablet to unlock the door).
Of course, any (potentially automated) interconnections between the different devices use M2M communication. On top of that, the following two interactions count towards M2M communication: a) any recorded data on a mobile device regarding a patient's medical history being directly transmitted to some central storage for future analysis and b) the following access of said data for the actual analysis and predictions based on machine learning.
Processes. The entire system combines multiple aspects into a single connected process: many different devices and connections are involved (chip implant, mobile devices, smart city environment, etc.), with each connection designed to fulfill a specific purpose. Thus, all of these connections are now combined into a single process aiming to save a human life potentially.
The resulting process can also be parameterized: the chip implant is connected to a mobile device (i.e., a smartwatch) which can automatically trigger a medical emergency when the measured bio values fall outside a specific range according to predefined process parameters.
Data. The interconnection of all those devices allows data collection across the entire process. This makes process optimization easier by evaluating usage data or analyzing connection quality-of-service (QoS) measurements. Furthermore, it is possible to store medical data on a cloud server (preferably in anonymized form) for analysis, data mining, prediction or early detection, and treatment improvements.

4.4.1.2 Malware affecting the Continuous Glucose Monitoring use case

A **software fault** in either the chip implant or the designated diabetes application is highly relevant. On the side of physical impacts, there are safety or health issues as possible if the device would improperly measure the bio values of the patient due to a fault in the software.
In case of a false negative, failure to trigger an alarm could be a life-threatening mistake. But a software fault could also have direct or indirect economic impacts. For example, in the case of a false positive when evaluating the bio values of the patient, an unnecessary alarm could be triggered. Given that the patient would not notice or fail to acknowledge the alarm, a medical emergency team would be dispatched, which causes an economic impact of potentially large scale (depending on the exact country and the local health-care system).
Additionally, when the improper measurements only differ from the actual measurements but should not cause a medical emergency, the doctor may apply the wrong medical treatment based on the erroneous measurements. The economic impact here is rather difficult to assess, for the doctor could discover the mistake during the first visit (small impact after correction), or only after several weeks (high economic impact, contains costs of incorrect treatment but also requires starting over with correct treatment).
The physical impacts may be severe, while the corresponding economic impacts are unnecessarily high. Any of these impacts could potentially have been prevented by adequately testing the software before deployment in production.
A **Trojan horse** has high relevance for this use case, too. A Trojan horse aims to remain undetected while gathering data and potentially forwarding it to a third party or manipulating the data on the device itself.
There would be some physical impact in both cases: either your data is being stolen, or your medical history is being manipulated, resulting in false evaluation during the next routine doctor's appointment. Most likely, there would not be a direct economic impact when your data is being stolen. Nonetheless, it could be possible that there would be an

economic impact if that data reaches the wrong parties. For example, this could include a potential threat of increased health insurance costs or apply to any other party with malicious intents.
A third possible consequence would be the attacker gaining control over the mobile device, allowing him to change configurations and settings and cause the dedicated diabetes application to malfunction. A malfunction could result in the same impacts as previously described, only that it would be possible for data that will be recorded in the future. On the other hand, if your doctor was to perform a false evaluation of the manipulated medical history, the resulting improper treatment could also cause an economic impact on the patient.
A **worm** has the unfortunate ability to replicate and distribute itself to any connected device. Accordingly, it would be possible for the worm to reach the CGM implant from the infected mobile device, and, therefore, a worm has relevance for this use case. Subsequently, depending on the exact damage function of the worm, the CGM implant could cease proper functionality or stop working at all.
Depending on the exact damage function of the worm, similar economical impacts as in the case of a Trojan horse would apply. Should the worm infect the CGM implant and cause it to stop working, requiring a replacement would possibly be quite expensive and cause additional economical impact.
Due to its infectious nature, a **virus** is highly relevant as well. However, given that a device infected with a virus may be subject to data theft or manipulation, unintentional remote access, or any other arbitrary damage function implemented in the virus, it is generally assumed that the same impacts as for Trojan horses and worms apply.
Ransomware is of high relevance, too: when a computer virus has an encryption tool as its damage function, it encrypts the storage medium and renders affected applications unable to access any data.
As a result, the dedicated diabetes application would no longer be able to access relevant information, send periodic status reports to the health facility, trigger a medical emergency in case the values fall outside of the predefined threshold, or respond to incoming health alarms due to lack of data or proper authentication. It could also be possible that due to the encrypted storage medium, the application would be unable to store and process incoming bio measurements of the CGM implant.
The main economic impact specific to ransomware attacks is the demanded ransom. But, apart from that, any other, previously described impact resulting from either application or sensor malfunction applies as well. These include any impacts regarding medical data, medical emergencies, and improper treatment, but also goes as far as requiring a device replacement.
Relevance for this use case is also a **rootkit**. Although the goals of a Trojan horse are very similar to those of implementing a rootkit, the distinction between the effects of Trojan horses and rootkits is somewhat fuzzy. As a result, the same impacts apply to an infected device containing a rootkit which also apply to a device infected with a Trojan horse.
A **bot** can be considered to be of high relevance. Since a bot is part of a much larger botnet in most cases, there are also two perspectives, which need to be considered. One where the device is infected by a bot and thus acts as part of a botnet controlled by the botmaster in a remote-controlled manner. And another, where the device in question is the target of a distributed attack performed by such a botnet.
When the device itself has been infected and is now acting as a bot, a small yet good thing is that the computational load resulting from this unintentional activity would usually not be high enough to cause actual impacts on the mobile device. That is because the ultimate goal of a botmaster is for his bots to keep their activity below the level of exposure to persist the nodes in his botnet. However, on the economic side, the additional

Malfunction	Physical Impacts	Economic Impacts
Software Fault	Health	Direct costs (unnecessary, possibly high)
Trojan Horse	Data	Direct costs (improper treatment);
	Hardware	Indirect costs (e.g., health insurance)
Worm	Hardware	Direct costs (device replacement);
		Possible additional costs
Virus	*Same impacts as for Trojan horses and worms*	
Ransomware	Data	Direct costs (demanded ransom);
	Hardware	Direct costs (device replacement);
	Health	Direct costs (improper treatment);
		Direct costs (unnecessary emergency);
		Indirect costs (e.g., health insurance)
Rootkit	*Distinction between the effects of Trojan horses and rootkits is fuzzy; Same impacts as for Trojan horses apply*	
Bot	Device is infected:	Direct costs (power bill, low);
	- Hardware	Direct costs (unnecessary medical emergency);
	- Health	Indirect costs (sophisticated burglary, unlikely)
	Device is target of botnet:	
	- Hardware	
APT	*Presumably, a diabetes monitoring system would not be an attractive target while requiring tremendous efforts on the attacker's side*	

Table 4.1: Impact analysis of the CGM use case.

computational load and thus increased power consumption of the device will lead to an increased power bill.

Depending on the exact definition of a remotely controlled bot, its attack scenario resembles the one of a computer virus. More specifically, referring to the diabetes use case: in combination with sniffing and a command replay attack, an attacker could use the diabetes application to trigger a medical emergency willingly, for example. This example will cause an unnecessary medical emergency's already well-explained economic impacts (with potentially high costs). In addition, there is a low possibility of "sophisticated burglary", where the combination of different network-based attacks (sniffing, jamming, replay) could allow an attacker to retrieve the one-time unlock key for the patient's apartment door. The resulting economic impact is impossible to estimate since it heavily depends on the financial and living situation of the victim as well as the underlying goals of the attacker. On the other hand, when the device is the target of a botnet, there are most likely incoming distributed denial of service (DDoS) attacks. These DDoS attacks are at least slowing down the device and any connected networks it is contained in or even entirely blocking the device from working correctly (or at all). Regarding the use case at hand, this will probably not cause any economic impacts since a) the target is a single user and not a company losing their earnings for the duration of the attack, and b) the targeted device would not have to be replaced. Instead, the remedy would be to adjust the network configurations by contacting either the local network administrator or responsible Internet Service Provider (ISP). Alternatively, it is often possible to wait out the attack until the network load has reached normal levels again.

In contrast to the malware above, an **APT** is not considered relevant. An APT will maintain multiple (of the previously described) attack vectors targeting a single device/target. Presumably, a diabetes monitoring system would not be an attractive target while requiring tremendous efforts on the attacker's side. Therefore, such an advanced attack would not be launched in the first place, and thus no physical or economic impacts apply.

4.4.1.3 General Security Issues

Ismail did not specify how the key generation and unlock mechanism would work in detail regarding the security aspects of such an access-key approach. Thus, it remains to be defined who would generate an access key and what authentication mechanism would be used for secure authentication at the door. One possible approach could be that the door and the HMC (i.e., the MPTC tablets) would agree on a "proof-of-knowledge"-based interactive proof system and a specific computation to be performed. This computation would ideally require some dynamic parameter (i.e., the access key) to be included for it to perform. In the case of an emergency, the smart door would have to generate such a key for the computational problem to be solved and send the key to the HMC, which would then forward it to the MPTC device. Knowing the required computation in advance, the MPTC device would be able to perform said computation with the key and produce the correct solution to authenticate to the door. Like this, even when an attacker would get a hold of the access key being sent over the network, he could not properly authenticate to the door because the access key itself is not the solution. Also, the (uninformed) attacker would neither know about the existence of such a computation nor which computation would have to be performed.

4.4.2 Brain-Controlled Artificial Limbs

In this chapter, the mind-controlled smart arm prosthesis will be reviewed and analyzed. With the help of a skull cap with electroencephalogram (EEG) sensors and actuators, the system can measure brain activities used to move the arm prosthesis, as shown in figure 4.6. This innovative arm can receive information, but it can also give back intelligent feedback regarding the surroundings and the objects in contact with the arm [8]. The arm prosthesis itself could be defined as IoT since the cooperation between machines and humans is limited.
To ensure cooperation, the use case is extended to a one-sided and a two-sided communication between the machine and the human [8]. On the one hand, brain activities will be transformed into machine language to move the prosthesis. On the other hand, the prosthesis can give back information to the brain. Possible feedback could be the temperature or pressure. With this feature included, there exists unconscious interconnection between humans and smart machines.
Additionally, it will be assumed that the prosthesis is connected to the Internet for regular software updates in this use case. It can also be assumed that a smartphone application exists which provides the user with a status protocol of the artificial limb. These conditions are necessary for the malware analysis since many types of malware are spread via the network or Internet.

4.4.2.1 Motivation for IoE

Below it will be argued why the brain-controlled artificial limb is part of the IoE and not only the IoT.
Communication. There exists a bi-directed connection between the smart arm and the client's brain. A client can control the arm via brain activities sensed by the EEG sensors, but there is also a connection vise versa. For example, the prosthesis can give back some information regarding temperature or pressure. This information can therefore be sent back into the brain directly. Another critical point is that the connection is unconscious and does not need additional interfaces between the human and smart arm prosthesis.
Processes. Different devices need to be matched up perfectly in one process to have a working arm prosthesis system. For example, the EEG skull cap needs to interact with

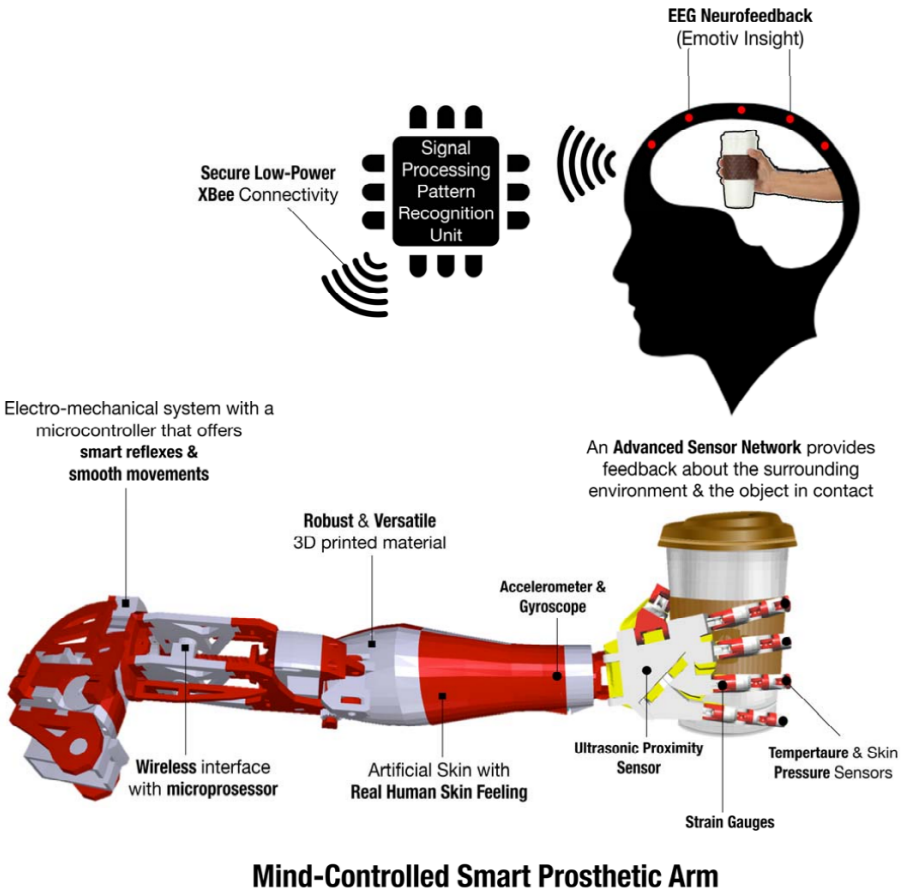

Figure 4.6: Brain-controlled artificial arm architecture by Beyrouthy et al. [8, Table 1].

the human brain and the arm prosthesis. Moreover, the smart arm needs to be connected to the Internet for plugins and system updates. The prosthesis is not a standalone mechanism; it is part of a technology ecosystem that fulfills the definition of IoE.

Data. An intelligent, brain-controlled arm prosthesis requires a lot of information to function. This information is transmitted from the brain to the EEG, which collects and processes it to forward the results to the sensors in the arm itself. This mass of data needs to be handled and merged in an integrated process to form a constant flow of information.

4.4.2.2 Malware affecting the Brain-Controlled Artificial Limbs use case

According to the definition stated above, a fault does not classify as malware. Nevertheless, it can have adverse effects on a system. Since a **software fault** can always happen, it also has relevance to this use case. A wrong configuration or implementation could lead to miscommunication between the arm and the patient. In certain circumstances, malfunctioning communication can lead to an accident, where people can get hurt.

An example could be if the arm starts to react incorrectly while driving. Therefore there exists a potential risk of a physical impact. Every physical impact can also lead to an economic impact. The patient can potentially sue the seller, or the company's reputation can be impacted when there is a system fault, leading to an economic impact. These two impacts need to be always considered when there is a physical impact.

Since the arm prosthesis is connected to the Internet, it is plausible that it could be affected by a **Trojan horse**. Another possibility to get a Trojan horse could be via plugins or updates [17]. If the system is infected, there are different outcomes. One of them could be that it shuts down the whole system. Depending on the situation, it can have a physical impact on the client and his surrounding. For example, when the arm stops

working while driving the car, the client loses control and creates an accident. The Trojan horse can not only shut down the device; it can also control it and adjust the settings. It could therefore change the configuration. There are two cases: In the first case, the Trojan horse generates by accident a malfunction. In this case, the system reacts in the same way as with a system fault. In the second case, it changes something on purpose. An example here could be that wrong information will be sent to the client. Since there is no interface between the system and the client's brain, the false information will be directly sent to the brain. For example, if the brain gets wrong information about a temperature, it may hurt the client and lead to physical damage. A Trojan horse can also enter your data and steal it, which would have a negative physical impact on the client and a positive economic impact on the attacker.

A **worm** can be considered relevant as the arm prosthesis could be affected when using the Internet to download new updates. It can also be affected physically via a USB stick [18]. Once the system is affected, the worm can remote control the whole arm. For example, it could shut down the arm or let it react in a certain way, leading to a physical impact. A worm can also interfere in the communication between the arm and the client and send wrong information to both of them. A communication issue can lead to accidents and, therefore, physical impact. A worm can also collect data and sell it [18]. In this use case, the artificial limb is connected to the Internet. So if the worm could spread via the company's network, it can quickly affect other devices. The affection of a group of devices leads to a significant economic impact for the company since they would take responsibility for any physical damage.

A virus has relevance for this example as the artificial limb can be affected via network or Internet [19]. The smart limb could also be affected physically by a USB stick. However, some types of a virus are not relevant for this use case, as an e-Mail virus, since it is not connected to a mail server. Suppose the artificial limb is connected to a smartphone affected by a virus. In that case, the virus can also spread and affect the prosthesis, as described by Ghallali and Ouahidi [27] in their analysis of smartphones spreading malware. Since the goal of a virus is to spread, it would like to remain undetected for as long as possible. In the example of an encrypted virus, the malware is primarily hidden and can listen and collect data while the user does not know about it. On the other hand, in the case of an artificial limb, there is a lot of personal data that is potentially interesting for different organizations. On the other hand, it can happen that a malfunctioning will be generated, which has the same physical and economic impact as a system fault. There are many different types of viruses [19], but most of them have similar outcomes. If it generates a malfunction, there is a possibility for a physical impact. The same goes for when data is collected or screened. If the data gets sold, it will generate an economic impact.

Also, **ransomware** can be relevant. Locker ransomware may stop the artificial limb from working and ask the user to pay a ransom to use the arm again. On the other hand, encrypting ransomware could deny access to different functionalities, such as feeling temperature or pressure. It can be discussed whether restricting access to a physical ability may be labeled as a physical impact, especially since said ability would not be possible without adding the prosthesis anyway. In this review, it will be considered a physical impact. On the flip side, crypto ransomware can collect data in the background and make the data unusable for the client. This could be done by redirecting the traffic to the attacker's server. The crypto ransomware could deny the artificial limb access to the EEG data, ultimately paralyzing the arm [20].

It could also restrict access to the sensors, which will result in the same outcome. In any case, the client needs to pay a ransom to gain back control over his artificial arm. The ransom payment will trigger a negative economic impact for the user and a positive impact for the attacker. Because being physically unable to use a prosthesis is restricting, the victim will likely pay the ransom to use his arm again. As the attacker can gain a

Malfunction	Physical Impacts	Economic Impacts
Software Fault	Health	Indirect costs (client can sue the manufacturer); Indirect costs (manufacturer reputation damage)
Trojan Horse	Data Health	Direct benefit (attacker sells data)
Worm	Data health	Indirect costs (manufacturer reputation damage); Direct benefit (attacker sells data)
Virus	Data Health	Direct benefit (attacker sells data)
Ransomware	Hardware (with possible impact on health)	Direct costs (ransom)
Rootkit	Data Hardware Health	Direct benefit (attacker sells data)
Bot	Health	Direct costs (power consumption)
APT	*Irrelevant (lacking public interest)*	

Table 4.2: Impact analysis of the artificial limbs use case.

lot of money, the client will have to pay at least once. Furthermore, combined with the idea that the malware could send wrong information to the brain (for example, too hot temperature), the blackmailer could create physical pain until the victim has paid.
A **rootkit** can be considered as relevant in this use case too. The rootkit would like to stay undetected as long as possible to get root access [28]. At first, it will not intervene in the flow and will not generate any other physical impact besides collecting and scanning personal data. Later, when root access is provided, the rootkit can do anything. Hence, the physical and economic aspects of all the malware above can be combined in this part.
A **bot** can be relevant since the artificial limb is connected to the Internet. As in the first use case, the difference between being a single infected bot and being the target of an entire botnet needs to be made. If the smart arm is affected, it won't generate a problem, as the bot wants to stay undetected. However, the prosthesis could react slightly slower, and the power consumption might be marginally higher. Therefore, a minor economic impact could be generated, but it can be assumed that there would not be a physical impact. If the brain-controlled arm is the target of a DoS attack, it might hinder the arm or even entirely shut it down. Under some circumstances, this could lead to physical impact.
An **APT** has relevant for this use case. For a brain-controlled artificial limb to be a target of an APT, there needs to be enough public interest as the goal of an APT is a long-term utility [23]. As in the diabetes use case, an artificial arm is a niche product, which is not attractive enough to invest this much time and effort to plan an APT.

4.4.2.3 General Security Issues

Apart from the security issues mentioned above, there are also general security aspects to consider in this use case. One part of the artificial limb is the EEG system which connects the limb with the user's brain. A BCI is used to provide this interconnection. Bernal et al. [6] discuss multiple security issues when speaking about BCI. Instead of looking at an attack's physical and economic impacts, they look into security concepts regarding integrity, confidentiality, availability, and safety. These concepts show another security analysis approach and help evaluate other impacts of attacks against BCI systems. However, when looking at these concepts and analyzing the use cases discussed in this work in different phases (e.g., brain signal generation, data acquisition, etc.), it reveals new security issues, such as jamming or spoofing attacks, which compromise data integrity and availability, as well as safety [6].

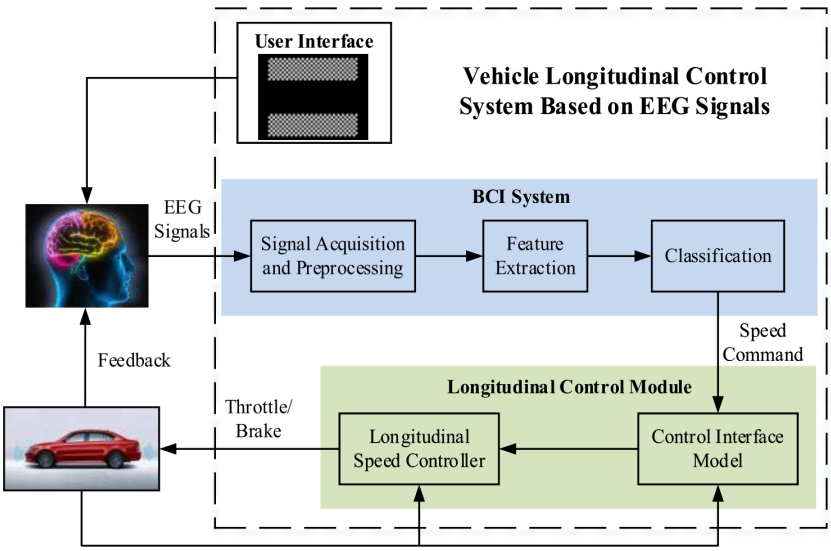

Figure 4.7: Structure of the vehicle control system by Lu and Bi [29, Table 1].

4.4.3 Brain-to-Vehicle

This section analyses a combination of the already existing autopilot of Tesla and the new project "Brain-to-Vehicle" from Nissan. The car's primary driver will be the autopilot, which is connected to the brain activities of the driver. While measuring the brain activities, the autopilot adjusts his driving style according to the preferences of the human client. Therefore, if the autopilot detects that the client feels uncomfortable, it will adapt its driving style, sound and volume, or heating accordingly.
According to Yan and Jia [9], the BCI uses brain electrical acquisition. It also uses electroencephalography (EEG) signals with surrounding information, as introduced by Bi et al. [10] and shown in figure 4.7. This new technology provides new possibilities for the disabled [9]. Like the already existing Tesla model, the smart car in this example also has an Internet connection, can browse and download applications. Furthermore, the smart car is also connected with the user's smartphone to allow unlocking the doors when entering the vehicle.

4.4.3.1 Motivation for IoE

This section will show that, with the underlying interconnection of "things," all four pillars of IoE are used and necessary to make a brain-controlled smart car work, and therefore, it is part of IoE.
Communication. The smart car is connected to the human brain through the skull cap with EGG. The human unconsciously sends information about his well-being to the vehicle, and it responds to this information by acting in a certain way. This communication is a closed circle. The human receives the new information, and therefore his brain activities change again. This loop shows the interconnection between the driver and the car. Additionally, there is no extra interface involved which is usually the case when dealing with IoT.
Processes. The underlying system fulfilling the purpose of this use case is a complex combination of different parties. All involved parts need to be perfectly synchronized in one process to make a brain-controlled autopilot work. The human with the smart car needs to be able to interact smoothly. On the other hand, all the EEG sensors, surrounding information, and smartphones need to interact appropriately to achieve their common goal.

Data. Much data needs to be collected and interpreted to provide interconnection and communication within the respective process. Different technologies will collect data. For example, the EEG skull cap is responsible for managing neuronal activities [9]. Thus, sensors in the car are collecting surrounding information, which allows detecting whether an upcoming obstacle represents an impending danger [10]. This data, in combination with the neuronal activities of the driver, provides a more secure driving experience.

4.4.3.2 Malware affecting the Brain-to-Vehicle use case

This chapter will discuss the main cybersecurity issues for a brain-controlled car, analyzing the main types of malware.
This evaluation starts with the subject **software fault**. Bugs which are unknown to the developer are always risky. A fault can lead to unexpected behavior of any kind. That the traditional Tesla autopilot has its flaws is shown by Dikmen and Burns [30]. Therefore, a fault has high relevance in this use case, too. A fault in a self-driving car's software can lead to severe physical and economic impacts. Assuming the autopilot has a fault, it may not detect an obstacle, or in some cases, drive too fast because of a wrong configuration. Any of these cases could lead to accidents involving the human driver as well as other people. The risk of a physical impact is therefore high. Another scenario could be that there is a fault in the car's lock. The result would be that a person without a key could enter the vehicle and steal the car or something in it. This theft represents an economic impact for the owner as well as for the manufacturer, as they would need to take responsibility for the fault in their lock system, for which the users can sue them. Both cases could be prevented with an extended testing phase.
One of the first malware to be discussed is the **Trojan horse**. The goal of a Trojan horse is to compromise a system or install a trapdoor [17]. The machine first gets the Trojan horse which then downloads the malicious software. The main entry point is via the Internet. It is, therefore, plausible that the smart car could be affected by a Trojan horse. Is the malicious software in the system, it can "steal personal information, password, tamper data, or destroy files" [17]. The destruction of an autopilot-relevant system file could lead to a fault, whose consequences were discussed in the section above. If the malicious software steals personal information or passwords, it can lead to physical and economic impacts. In addition, personal data, including brain activities, could be analyzed and sold since this information could be of high interest or value to some companies. Depending on the password and its repetition for other utilities, the thief could steal the car, enter your e-banking, or buy things with your credit card. A smartphone is considered a lock-in device to the vehicle. If the software has access to your phone and is connected with your smart home or your e-banking, there is a potential economic impact. Tampered data represents another interesting aspect. The malicious software could send wrong information to the autopilot. For example, it could change the destination location, increase or decrease the car's speed, or close the doors such that the driver is trapped in the vehicle. All of this could impact the physical or mental health of the driver. To prevent the scenarios described above, there exist different actions as installing antivirus chips on the network interface card, installing patch programs and repairing system loopholes, and using a network firewall [17].
The second malware is the **worm**. "A computer worm is a program that self-propagates across a network exploiting security or policy flaws in widely-used services" [18]. The difference between a worm and a virus is that a worm does not need a host to propagate itself. A worm finds its new target while the process of scanning through external target lists, pre-generated target lists, internal target lists, or passive monitoring. In this case, it will be assumed the smart car could be found on an internal target list, which means someone explicitly wants to target smart cars. This attack can happen physically, as the

attacker physically enters a USB stick with a worm, or the worm enters via the Internet while browsing. A worm has, therefore, relevance for this malware analysis. Is the worm part of the system it starts to spread. According to Weaver [18], a worm can have different payloads. One is the Internet Remote Control, where the worm opens a backdoor that allows the attacker to enter via the Internet and execute their code. In this case, the attacker could change specific parts of the autopilot or security system to his use (e.g., change the passcode to enter the car or change destination location). One of the most critical points of a worm is data collection. In the example of a brain-controlled car, much sensitive and personal data is saved in the vehicle and on the company's servers. If a worm could enter and collect this data, they know a lot about the car's owner, which they can use to their advantage. The collection of personal data represents physical damage. They could create economic harm if they use the knowledge to steal or ransom money or sell the information.

Apart from the malware mentioned above, also a **virus** can affect a brain-controlled car. Unlike a worm, a virus can only spread from one to another when its host is on the target computer [19]. A virus can distribute itself via the network, Internet, or removable hardware like a USB stick, CD, or similar. In the case of a brain-controlled car, all three options are plausible. Since the car is often in a public space, everybody can access it. If the thief finds a way to enter the vehicle, he can also use a USB stick to inject the virus. In their work, Khan [19] shows a lot of different kinds of viruses. One of the most difficult to detect is an encrypted virus. Many state-of-the-art antivirus products usually miss encrypted viruses because they use varying encryption mechanisms each time. It is thus plausible that the EEG system is affected by a virus before we know it. The virus has time to listen to and interfere in the communication. Wrong information could be sent to the autopilot, and the vehicle could get out of control. Another dangerous virus is a FAT virus that prevents access to specific hard drive sections [19]. This can lead to malfunctioning software since some files are damaged. As discussed in the section above, malfunctioning software can always cause an accident and, therefore, a physical impact.

There is a long list of different other viruses. They have different ways to enter the system, but most have the same outcome. One is physical damage to the human driver or other people in his surroundings. The other one is economic: in some way, the hacker wants to generate money by selling information, demanding a ransom, or stealing data or physical objects.

There remains the question of whether **ransomware** could affect a brain-controlled vehicle. A smart car could get ransomware via traffic redirection, botnets, or social engineering [22]. According to Zakaria [22], there are two main ransomware types: the first one is crypto ransomware, and the second one is locker ransomware. If the car were affected by crypto ransomware, there would be no sign for it at first. However, at some point, the main files are going to be encrypted such that the victim needs to pay the ransom to obtain the decryption key and use the files again. In the case of locker ransomware, the ransomware could deny access to the car or other parts of the system such as navigation, sound, or similar. A client being locked into his vehicle certainly is a physical impact. To get back access, the victim needs to pay the specified ransom. The main goal of such malware is to get money from the owner and generate an economic impact.

The **rootkit** is another malicious software that could lead to problems in the brain-to-vehicle process. Undetected software can give access to another person and provide control over the system. A rootkit can be relevant as it is often a result of a Trojan horse [19]. Once the system is affected, the rootkit can give access to an external person to take control. As the name suggests, this person — the attacker — then has root rights and access to all functionality. The attacker can change parts on purpose and get all the information he wants, leading to a physical and economic impact. If a hacker would like to take over the control of the car while driving, he could easily do so. Especially when

Malfunction	Physical Impacts	Economic Impacts
Software Fault	Hardware Health	Direct costs (car crash, medical treatment); Direct costs (car stolen); Indirect costs (client can sue the manufacturer); Indirect costs (manufacturer reputation damage)
Trojan Horse	Data (with possible impact on hardware and health)	Direct benefit (attacker sells data)
Worm	Data Health	Direct costs (demanded ransom); Direct benefit (attacker sells data)
Virus	Data Hardware (with possible impact on health)	Direct costs (power bill, low); Direct benefit (attacker sells data)
Ransomware	Data Hardware	Direct costs (demanded ransom)
Rootkit	Data Hardware (with possible impact on health)	Direct benefit (attacker sells data)
Bot	Hardware (with possible impact on health)	Direct costs (power bill, low); Direct costs (car crash, medical treatment)
APT	Hardware (with possible impact on health)	Indirect costs (client can sue the manufacturer); Indirect costs (manufacturer reputation damage)

Table 4.3: Impact analysis of the Brain-to-Vehicle use case.

people want to harm each other, one could use this to generate public accidents in which many people could suffer severe physical impacts.
In the relevance analysis of a bot, one needs to differentiate between two cases. A device can be effective but not be the target of a Denial of Service (DoS), and it can be the main target of a DoS. If the device is not the target, it acts as a "zombie" and helps attack the target. Since the malicious software needs to be undetected on the smart car, it is hard to realize if it is affected. The affection by itself generates most likely no physical impact, but it can higher the power pill and cause, therefore, an economic impact. On the other hand, if the smart car is the target of a DDoS attack, it can slow the response or shut down whole system parts. In a different situation, that can lead to accidents which will result in a physical impact.
As the new brain-controlled car is most likely a new technique with new exciting data, it could be a potential victim of an **APT**. An attacker could try to enter the system, gain access to information, and control the autopilots of the car to use the vehicle remotely. Most of the time, an APT has an economic or ideological background. However, an exciting and frightening idea is that criminal organizations could use crewless cars to generate physical damage on a public scale.
All of the malware functions above are already known by companies like Tesla and Nissan [31]. However, the new EGG brain-activity scanned data feature will be a more exciting target for malware since more personal data is involved. Furthermore, there are more interfaces malware can affect since there is now a new entry point for malware.

4.4.3.3 General Security Issues

Besides the issues already mentioned, there are more general security issues. Like the brain-controlled artificial limb, the brain-controlled car is also a BCI system. Therefore, it has various other security impacts such as integrity, confidentiality, availability, and safety [6]. Moreover, when looking at the data acquisition and stimulation of the BCI, there are other attacks, besides malware, which can compromise the security as jamming and spoofing attacks. Again, this can affect the integrity, availability, and safety of the

system [6]. Since the car is connected to the smartphone, all security challenges for the smartphone can be added to this use case, too. Another security aspect is when to react to the client's request and when not to. For example, the smart car would not increase the speed when it has already reached the speed limit, even if the driver still wants to drive faster. In the case of a potential car crash, it is not always so clear to decide which option is the best. Individual human drivers would react differently in such cases, leading to various physical impacts. It could be decided that the autopilot will fully take over in such cases, ignoring what the human driver would do.
Consequently, the autopilot needs to assess the situation, which is error-prone, especially if there is a system fault or malware included. However, this raises ethical questions like which life has more value for the autopilot, the driver's life, or the life of the passer-by. In which cases to listen to the human driver and when not to, opens up new security issues for the driver and other people involved in accidents.

4.4.4 Brain/Cloud Interface

The idea of the Brain/Cloud Interface (B/CI, not to be confused with a Brain-Computer Interface BCI) is to connect the human brain to the Internet. The B/CI will connect any reachable cloud-computing network to the neurons and synapses in our brain without any considerable delay, providing instant and direct access to all knowledge humanity has gathered and put online — by only thinking about it. Moreover, direct access to the Internet could enable the B/CI to power Brain-to-Brain communication systems as envisioned by Grau et al. [32], shown in figure 4.8. That would be made possible by implanting nanorobots into the brain, termed "neuralnanorobotics" [11] (NNR for short), which could autonomously position themselves at the correct location to reach relevant neurons. The NNR are designed to monitor the brain waves, translate them back and forth between computer-readable data streams, and wirelessly transmit the encoded information to and from a network of cloud-based supercomputers [11].
While connecting a new type of infrastructure to the Internet is nothing new in itself, the most astonishing thing is that there is no need for a device between the human and the Internet anymore: you can directly google from your brain. Previously, we had to rely on a smartphone, tablet, or desktop computer. Now, we can skip these devices and directly connect our brains to the Internet Service Provider's (ISP) infrastructure, e.g., a base station for mobile communication networks like 4G or 5G.
Naturally, as is required for the interconnection to work correctly, the B/CI would have a bi-directional connection between the brain and the Internet. This bi-directionality includes that a user can not only send requests and receive a response, but the inverse is also true: other people could directly communicate with his brain. That would not necessarily have to be non-verbally via brain activities but could technically also contain more traditional methods, i.e., any currently existing form of wireless communication compatible with the protocols or standards implemented by the NNR in the target brain. According to Martins, co-author of the original study [11], the "challenge includes not only finding the bandwidth for global data transmission, [...] but also how to enable data exchange with neurons via tiny devices embedded deep in the brain" [33]. Although the concept and main aspects should be feasible with the newly designed NNR alongside the current network infrastructure, the technical details remain subject to the latest research.

4.4.4.1 Motivation for IoE

In this work, it will be would strongly argue for the Brain/Cloud Interface to be an example of IoE technology. One aspect would undoubtedly be the direct involvement of the

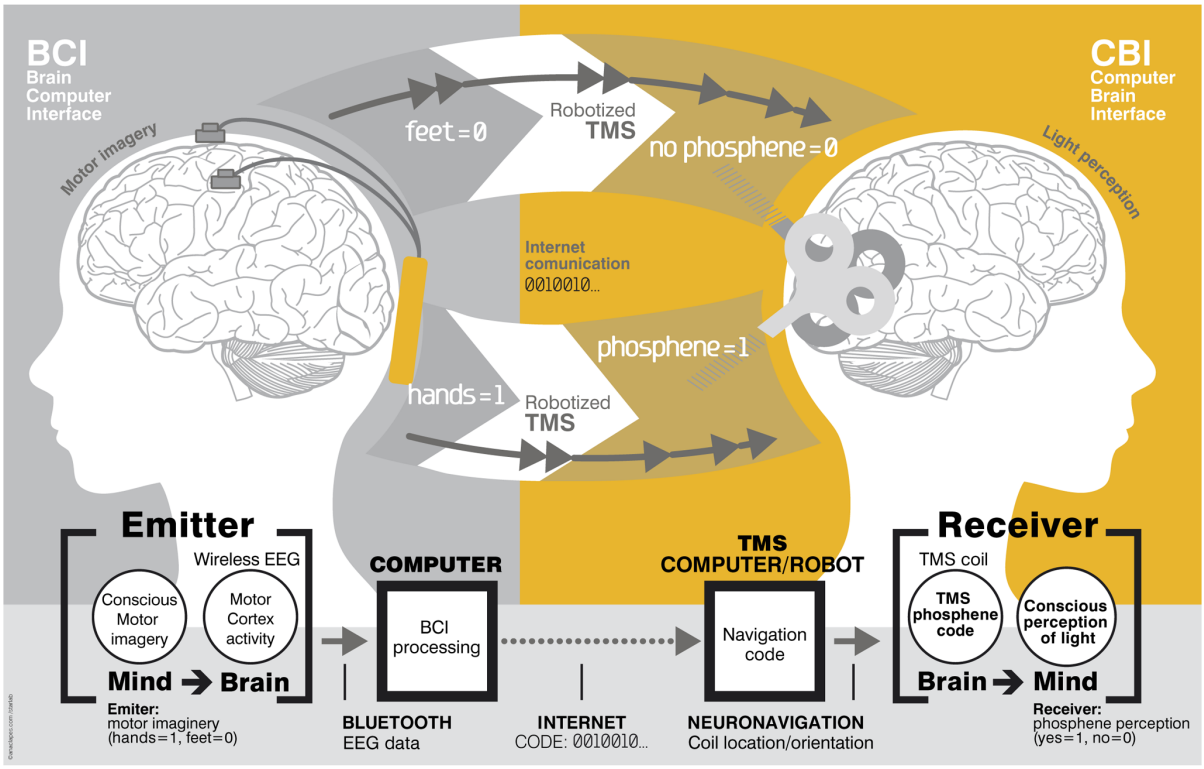

Figure 4.8: Brain-to-Brain communication systems by Grau et al. [32, Fig 1].

human being as a computing resource. This combines all three aspects of communication, processes, and data.

Communication. The B/CI allows for new forms of communication, like direct P2P communication, without intermediary or auxiliary devices except for the infrastructure required to maintain the communication itself. The resulting P2P communication is of the purest form, interconnecting multiple human beings independent of distance while not involving any machines the humans would have to interact or communicate with actively. Potentially, B/CI P2P communication could also be independent of the subjects' culture or language, depending on the exact usage of the neuralnanorobots.

Processes. Directly integrating the end-users as computing resources fundamentally alters accessing the Internet and simplifies existing processes by cutting out redundant intermediate steps and hardware.

Data. With the implanted neuralnanorobots measuring, processing, and translating the user's brain waves, the B/CI opens up a new category of processed data: brain activities and every associated thought or emotion. Although previous technologies have measured brain activities before, they have never been used in this context. Moreover, the encoded (and very much private) brain activity data has not been exposed to external networks to this extent, making for a formidable target of cyberattacks, blackmail, or other criminal aspirations.

4.4.4.2 Malware affecting the Brain/Cloud Interface use case

A **software fault** is considered to be highly relevant. Zero tolerance for faults since operating directly in or on the human brain. Given that the NNR must be able to alter synapses between neurons (and possibly also other brain tissue), any mistake or malfunction while doing so could have a substantial physical impact. The economic impact would range from the costs of surgically replacing the NNR to medically treating any brain damage resulting from the fault.

A **Trojan horse** can be highly relevant. It could most likely be disguised among the various things a human mind — with direct access to the Internet — would think about throughout a single day, especially when considering the many random context switches that occur when our mind drifts off. Considering the sometimes careless manner in which people treat their computer and private data already today, providing them with the possibility to put their brain at stake would be upright dangerous.
On the physical side, a Trojan horse usually tries to lay low and would likely hide a listener of some sort, e.g., to keep track of emotional state, synapses firing patterns, memory accesses, or every outgoing request (content, target, etc.). A following physical impact would be the exposure of our deepest and most private thoughts — converted into digital data and thus easily distributed. Upon the user discovering a compromise, psychological consequences could heavily vary, including mental traumas or developing mental disorders.
Economically speaking, a Trojan horse extracting your most private data would not cause any direct impacts. However, we live in a world where personal data is becoming more and more valuable to corporations. Being able to accurately detail the advertisement you get to see or perfectly optimizing the customer experience at the lowest possible costs is among the things most interesting to big firms. For the users themselves, there could also arise indirect economic impacts, or rather disadvantages for that matter. Similar to how job recruiters consult social media and (partially) publicly accessible data about new applicants, suddenly being provided with private data recorded directly from the applicant's brain could be the cause for discrimination or other unfair treatment, ultimately leading to not getting a job or being denied service.
A **worm** is considered to have relevance for this use case. An infected human brain is comparable to a mobile device: not limited to being part of a single network, able to connect to any (reachable) device or network, easily reachable by any other device or brain over the B/CI. Accordingly, a worm that infected a human brain over its B/CI can spread very far and incredibly fast. Furthermore, with the human as host, an attacker could even physically cross barriers which he would have to tediously dig through on a software level, e.g., directly connecting to the company-wide intranet when entering the office building instead of finding an external exposed vulnerability.
From this perspective, a worm would either have similar physical and economic impacts as, for example, a Trojan horse or a virus would. Alternatively, if not targeting the user and instead simply using him as an intermediary means to infect other devices or networks, the physical or economic impacts would not apply to the user.
A **virus** can have high relevance, mainly because of the significant damage potential due to the wide range of possible damage functions directly impacting the brain.
Accordingly, even minor damages to the NNR would most likely have an immediate, high physical impact in the forms of brain damage, potential loss of functionality, and other medical issues or psychological conditions.
The economic impact resulting from treating such issues or conditions would be equally high given the sensitivity and complexity of the brain.
Ransomware, on the other hand, is irrelevant for this use case. It is generally impossible to encrypt the chemical reactions within the brain since they are of a rather physical than digital nature. Furthermore, the synapses and neurons do not form a "database" that could be encrypted.
Also, a **rootkit** is irrelevant. The B/CI does not support schemes for external access; there is precisely one user which does not require remote access. Thus it would not be implemented in the first place. Lacking the possibility of legitimizing external access, implementing a rootkit does not make sense in this use case.
When analyzing a **bot**, we must again consider the two different perspectives. The scenario of a bot infection is irrelevant for this use case, mainly because there is no computational

Malfunction	Physical Impacts	Economic Impacts
Software Fault	Data Health	Direct costs (replacements, surgery)
Trojan Horse	Data (with possible impact on health)	Indirect costs (failed job applications)
Worm	*If targeting the user: same impacts as for Trojan horses and viruses*	
Virus	Health Hardware (with impact on health)	Direct costs (surgery, treatment) Direct costs (NNR replacement)
Ransomware	*Irrelevant (no stored data to be encrypted)*	
Rootkit	*Irrelevant (external access not supported)*	
Bot	Device is target of botnet: - Health - Hardware (secondary)	Device is target of botnet: - Indirect costs (treatment, therapy)
APT	Data Hardware Health	Direct costs (blackmail, other payments); Direct costs (repair, replacements); Direct costs (surgery, treatment, therapy); Indirect costs (collaboration third parties); Direct benefit (attacker sells data)

Table 4.4: Impact analysis of the B/CI use case.

capacity to be misused in the neuralnanorobots. Additionally, trying to access the brain's resources would be detected by the human (if at all possible).
However, it is highly relevant when the brain is the target of a coordinated botnet attack. Such an attack could physically overload the brain and sensory system without any way of filtering, unlike when compared to our ordinary biological senses, for which most of the input is unconsciously being filtered out. Overloading the sensory system ultimately puts the victim's mental health at stake, especially when the attack is long-lasting. Additionally, it could diminish the capacity to think properly by forcing its focus on "defense" and staying sane.
Considering the assumed popularity and possible applications of the B/CI, an **APT** is highly relevant for this use case as well. Following the different consequences and impacts listed above, the brain of any user would be most vulnerable to sophisticated attacks, and the attacker could achieve the greatest possible effect due to the brain's sensitivity and the delicateness of gathered personal data. Therefore, the damage potential of coordinated attacks targeting single victims would be tremendous. Especially when the B/CI has established its position in the mass market, any person could theoretically fall victim to an APT. However, it would be more likely for attackers to target specific persons of interest than a more ordinary person.
Considering the physical aspects, an APT could have impacts regarding data, the implanted hardware, or the user's (mental) health. It would greatly depend on the chosen malware and implemented attack vectors. On the financial side, an attacker could cause economic impacts in the form of blackmail, hardware repair or replacement, medical treatment, or other indirect costs in collaboration with malicious third parties. Financial benefits for the attacker concerning gathered data can also not be excluded.

4.4.4.3 General Security Issues

Apart from the previously stated malware analysis, the B/CI opens up additional, more general security aspects. Under the prerequisite of global distribution, the introduction of the B/CI to the mass market would cause networking issues for the existing communi-

cation infrastructure due to high-cadence requests, in addition to the data traffic we have today already.
In a more isolated setting, however, the implanted NNR expose physical side-channels (like any other device) and are potentially vulnerable to side-channel attacks by leaking information that is not intended to leave the system. Examples could include brain wave patterns, the frequency of specific synapses firing, the heat emitted during computations, or similar aspects.
With the trend for neuro-technologies being on the rise, a new generation of IoE should soon be expected: the Internet of Minds (IoM), which connects the brain to the Internet and other computing devices for them to be neuro-controlled. No matter their usefulness, introducing such IoM technologies would certainly raise major ethical issues regarding privacy or autonomy. Furthermore, these technologies "open a potential side-channel for downloading information directly from the brain without either consent or awareness of the hack" [33]. The neuralnanorobots required for IoM technologies must alter the brain tissue of the user to ensure proper feedback and persistence of the query response. Yet, directly writing such information to the brain would spark even more ethical discussions since "that could fundamentally alter one's sense of agency, and perception of the world" [33]. In addition, allowing to modify the brain structure of any user directly drastically increases the importance of cybersecurity and the severity of any security holes or breaches. When exploited in the right way, it might even be possible for an attacker to take over the NNR for remote control. The extent to which they could be misused is unknown until further studies on a working prototype have been conducted. Could it go as far as to take over "control of the brain"?
We will have data plans for "brain surfing" like the mobile data plans currently in place for surfing on our mobile phones. They would be required to implement Quality-of-Service (QoS) monitoring and billing properly. Such data plans are not considered related malware, but they could easily be affected by it. For example, by compromising the administration of an Internet Service Provider (ISP) and manipulating parameters or settings of customer contracts, e.g., reducing their allowed data limit or administratively increasing the amount of data consumed so far. With possible effects in mind, what will happen when our data "runs out"?

4.5 Conclusions

This work has studied four relevant and promising use cases. Each use case has been defined, its relation with the IoE has been justified, and related cybersecurity issues analyzed.
After performing such an analysis, it can be concluded that:

1. The evolution from the IoT to the IoE comes at a certain cost since integrating people in the processes and collecting related data creates new vulnerabilities. Upon infection of a device holding or collecting data, the user's personal or medical data could be affected. An attacker could steal or sell data, manipulate data to accumulate damages, or deny access to the data and demand a ransom for releasing it. More specifically, systems using brain interfaces such as brain-to-vehicle or the B/CI have certain risks in common: they all process the user's brain activities and transform them into digital signals. By transmitting them to connected systems where they are processed or stored, they expose them to potential malware infection and exploitation. This issue is especially critical as these interfaces process different brain activities, from mechanical commands to the body to thoughts or emotions. They represent the most private kind of data, which should never be leaked to any third party. Nonetheless, this is not an issue for lower-tier brain interfaces such as the one

used in brain-controlled artificial limbs, mainly because such interfaces process mechanical commands and sensory information while ignoring other cognitive activity.

2. Interconnected systems specifically designed for a particular target group, e.g., the CGM of diabetic patients or the brain-controller artificial limbs, only serve a potentially small population niche. Accordingly, any more advanced malware like an APT would likely not be a problem because such malware requires tremendous efforts on the attacker's side. At the same time, the system does not make for an attractive target, i.e., there are other, more attractive targets out there that have a more extensive user base or yield a higher sum upon exploitation. However, if the underlying public interest is high enough, the chance of an attacker launching an APT is increasing rapidly. This becomes especially problematic for the manufacturer of the targeted system because an APT attack is difficult to prevent or stop. Once chosen as a target, the company and its system have a substantial physical and economic risk. For example, a brain-controlled smart vehicle would generate public interest and thus be an excellent target for such advanced malware: the system has a large user base, and lots of valuable data is involved. The benefits of exploiting said data or otherwise misusing the vehicles highly motivate such an attack. Nevertheless, most of the problems regarding malware are already faced today, as is demonstrated by other works such as those of Choi et al. [31] or Nie et al. [34] using the example of the Tesla autopilot.

3. Lastly, when comparing the CGM and B/CI, we can conclude that they both have a high potential of causing direct physical impacts on the user's health. Even though they are entirely different systems, their somewhat invasive nature and direct integration in the human body make them critical vulnerabilities. Relatively harmless and undesirably frequent malfunctions, e.g., a simple software fault, could already have massive impacts. When extending the scope to include more sophisticated threats like a worm or virus, the potential for physical damage remains at least as high. Still, it might additionally have economic impacts – especially when the malware is well-tailored to the targeted system.

4.6 Future Work

Given a large number of possible IoE applications, there is the need for more extensive research and analysis of different use cases. Alternatively, it could be possible to combine multiple use cases, focus on their common aspects, and investigate the influences or impacts malware could have on them with respect to the entire IoE context. In any case, the performed analysis should be deepened to provide more revealing and stronger evidence, i.e., in an experimental setting on a more technical level rather than performing a theoretical analysis.
A different approach would consist of trying to avoid the malware affection in the first place by following suggestions mentioned in previous works. One suggestion would be the four layers of cybersecurity discussed by Chinchawade and Lamba [5] and Lu and Xu [3]. Another proposal from Masoud [24] states that it would be beneficial to improve the users' awareness and give more possibilities to choose specific sensors. This would also include a rather empirical approach to verify the solution, successfully avoiding (at least some) malware infection.

Bibliography

[1] M. H. Miraz, M. Ali, P. S. Excell, and R. Picking, "A review on internet of things (iot), internet of everything (ioe) and internet of nano things (iont)," in *2015 Internet Technologies and Applications (ITA)*, 2015, pp. 219–224.

[2] J. Holler *et al.*, "From machine-to-machine to the internet of things: Introduction to a new age of intelligence," *Elsevier*, 2014.

[3] Y. Lu and L. D. Xu, "Internet of things (iot) cybersecurity research: A review of current research topics," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2103–2115, 2019.

[4] S. Duan, V. Shah-Mansouri, Z. Wang, and V. W. S. Wong, "D-ACB: Adaptive Congestion Control Algorithm for Bursty M2M Traffic in LTE Networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 9847–9861, 2016.

[5] A. J. Chinchawade and O. S. Lamba, "Authentication schemes and security issues in internet of everything (ioe) systems," in *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, 2020, pp. 342–345.

[6] S. L. Bernal, A. H. Celdrán, G. M. Pérez, M. T. Barros, and S. Balasubramaniam, "Security in brain-computer interfaces: State-of-the-art, opportunities, and future challenges," vol. 54, no. 1, 2021. [Online]. Available: https://doi.org/10.1145/3427376

[7] S. F. Ismail, "Ioe solution for a diabetic patient monitoring," in *2017 8th International Conference on Information Technology (ICIT)*, 2017, pp. 244–248.

[8] T. Beyrouthy, S. K. Al Kork, J. A. Korbane, and A. Abdulmonem, "EEG Mind controlled Smart Prosthetic Arm," in *2016 IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech)*, 2016, pp. 404–409.

[9] Z. Yan and X. Jian, "Research on brain-computer interface system for vehicle control based on motion imagination," in *Proceedings of the 2020 2nd International Conference on Big Data and Artificial Intelligence*, ser. ISBDAI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 516–520. [Online]. Available: https://doi.org/10.1145/3436286.3436495

[10] L. Bi, H. Wang, T. Teng, and C. Guan, "A novel method of emergency situation detection for a brain-controlled vehicle by combining eeg signals with surrounding information," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 10, pp. 1926–1934, 2018.

[11] N. R. B. Martins *et al.*, "Human brain/cloud interface," *Frontiers in Neuroscience*, vol. 13, p. 112, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2019.00112

[12] S. Purohit, S. Purohit, and A. Mathur, "An evolutionary development from iot (internet of things) to ioe (internet of everything)," *IRJET*, 2021.

[13] Microsoft, "Defining malware: Faq," 2009. [Online]. Available: https://docs.microsoft.com/en-us/previous-versions/tn-archive/dd632948(v=technet.10

[14] M. Russinovich, "Sony, rootkits and digital rights management gone too far," 2005. [Online]. Available: https://techcommunity.microsoft.com/t5/windows-blog-archive/sony-rootkits-and-digital-rights-management-gone-too-far/ba-p/723442

[15] Wikipedia, "Malware," 2021. [Online]. Available: https://en.wikipedia.org/wiki/Malware

[16] T. Klein, *A Bug Hunter's Diary: A Guided Tour Through the Wilds of Software Security*, ser. No Starch Press Series. No Starch Press, 2011. [Online]. Available: https://books.google.ch/books?id=NivmOf2J7qQC

[17] Z. Zhenfang, "Study on computer trojan horse virus and its prevention," *International Journal of Engineering and Applied Sciences*, vol. 2, no. 8, 8 2015.

[18] N. Weaver, V. Paxson, S. Staniford, and R. Cunningham, "A taxonomy of computer worms," in *Proceedings of the 2003 ACM Workshop on Rapid Malcode*, ser. WORM '03. New York, NY, USA: Association for Computing Machinery, 2003, p. 11–18. [Online]. Available: https://doi.org/10.1145/948187.948190

[19] I. Khan, "An introduction to computer viruses: problems and solutions," *Library Hi Tech News*, 2012.

[20] A. Bhardwaj, "Ransomware: A rising threat of new age digital extortion," in *Online banking security measures and data protection*. IGI Global, 2017, pp. 189–221.

[21] J. Wyke and A. Ajjan, "The current state of ransomware," *SOPHOS. A SophosLabs Technical Paper*, 2015.

[22] W. Z. A. Zakaria, M. F. Abdollah, O. Mohd, and A. F. M. Ariffin, "The rise of ransomware," in *Proceedings of the 2017 International Conference on Software and E-Business*. Association for Computing Machinery, 2017, p. 66–70.

[23] P. Hu, H. Li, H. Fu, D. Cansever, and P. Mohapatra, "Dynamic defense strategy against advanced persistent threat with insiders," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, 2015, pp. 747–755.

[24] M. Masoud, Y. Jaradat, A. Manasrah, and I. Jannoud, "Sensors of smart devices in the internet of everything (ioe) era: big opportunities and massive doubts," *Journal of Sensors*, vol. 2019, 2019.

[25] S. Ajrawi, R. Rao, and M. Sarkar, "Cybersecurity in brain-computer interfaces: Rfid-based design-theoretical framework," *Informatics in Medicine Unlocked*, vol. 22, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352914820306407

[26] U.S. Food and Drug Administration, "Eversense continuous glucose monitoring system - p160048/s006," 2019. [Online]. Available: https://www.fda.gov/medical-devices/recently-approved-devices/eversense-continuous-glucose-monitoring-system-p160048s006

[27] M. Ghallali and B. E. Ouahidi, "Security of mobile phones: Prevention methods for the spread of malware," in *2012 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, 2012, pp. 648–651.

[28] X. Wang, J. Zhang, A. Zhang, and J. Ren, "Tkrd: trusted kernel rootkit detection for cybersecurity of vms based on machine learning and memory forensic analysis," *Mathematical Biosciences and Engineering*, vol. 16, no. 4, pp. 2650–2667, March 2019. [Online]. Available: https://doi.org/10.3934/mbe.2019132

[29] Y. Lu and L. Bi, "Eeg signals-based longitudinal control system for a brain-controlled vehicle," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 2, pp. 323–332, 2019.

[30] M. Dikmen and C. Burns, "Trust in autonomous vehicles: The case of tesla autopilot and summon," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017, pp. 1093–1098.

[31] J. W. Choi, T. Choi, S. Kim, and S. Jo, "Towards utilization of error-related potentials for brain-to-vehicle communication," in *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, 2019, pp. 1–6.

[32] C. Grau *et al.*, "Conscious brain-to-brain communication in humans using non-invasive technologies," *PLOS ONE*, vol. 9, 08 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0105225

[33] J. Delaney, "Head in the clouds," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 6, no. 3, pp. 9–11, 2020.

[34] S. Nie, L. Liu, and Y. Du, "Free-fall: Hacking tesla from wireless to can bus," *Briefing, Black Hat USA*, vol. 25, pp. 1–16, 2017.

Chapter 5

How to Design a Vault for Crypto Exchange Centers?

Juan Cabanas Illescas, Manuel Bolz

The developments in the crypto space since the introduction of Bitcoin in 2009 have made it clear that the hype around cryptocurrencies is not going to stop any time soon. With the recent surge in the price of Bitcoin and Co. cryptocurrencies were one of the most lucrative investments over the last years, and because of that, there is also an increase in demand for secure storage for private and institutional investors to hold their assets long term. The goal of this report is to explore secure storage solutions for owners of cryptocurrencies to protect their assets from theft and other risks. In this report, we therefore introduce a new type of storage system: the crypto-vault. To justify the need for a crypto-vault, it has been looked at previous storage implementations such as different types of wallets, and it has been shown different ways in which a vault adds more layers of protection in comparison to simple wallets. For that, it is given an introduction on what a vault is, how it can be technically implemented, and which security measures crypto exchange platforms take to offer the highest level of security to customers.

Contents

5.1 Introduction

In this report, a high-level design of a crypto-vault, a secure storage solution for cryptocurrencies, is introduced and the technical aspects of the key generation and key management are covered as well as the security infrastructure and concept of crypto exchange centers that provide a crypto-vault service.
In the following sections, a short introduction and overview on blockchain technology and cryptocurrencies and some basic terminology on exchange centers is given. And after the general introduction, the main body of this report addresses the key components of a crypto-vault.

5.1.1 Blockchain and Cryptocurrencies

According to Euromoney Learning [36], blockchain technology can be understood as a shared, digital ledger used for recording transactions and tracking assets. This technology consists of a list of records (blocks) that are duplicated and distributed across an entire network of computer systems. Each record contains a cryptographic hash with the information of the previous block and transaction data. They are decentralized and managed with peer-to-peer networks as a public distributed ledger. This technology is built within a high-level security system, making the blocks almost unalterable with a high level of tolerance to errors. This blockchain technology is commonly used to trade digital assets named cryptocurrencies [37].
Since 2016, cryptocurrencies have been in the eye of individual investors, companies, and governments [35]. But according to a survey conducted by Bloomberg in the US 61% of people who had heard of cryptocurrencies said they had little or no understanding of how they work [34]. Cryptocurrencies can be understood as a medium of exchange, such as the Swiss Franc. They are the digital assets that use blockchain technology to validate the different transactions on the network. The difference between fiat money and cryptocurrencies is that every monetary unit is fully digital, controlled by the distributed network, and securely encrypted. The most known cryptocurrencies are Bitcoin, Ethereum.

5.1.2 Exchange Centers

A cryptocurrency exchange center is an intermediary that allows customers to buy, sell and trade cryptocurrency coins and tokens. Exchange platforms can be classified into two groups:

1. `Centralized Exchange Centers(CEX)` are online trading platforms that match buyers and sellers via an order book this means that the transactions are managed and centralized by a company [33]. They are also used to conduct trades from fiat money to cryptocurrencies (and vice versa). CEX are often easier to use, typically highly liquid, resulting in faster transactions and normally offer more functionalities than other types of crypto exchange centers. An example of this CEX would be the platform Coinbase.

2. On the other hand, `Decentralized Exchange Centers (DEX)` are an alternative to exchange cryptocurrencies [33]. They substitute the intermediary that generates a trust peer-to-peer exchange. The most important features are the security they provide in terms of custody, and the elimination of the identity verification process. The downside of DEX are limited efficiency and slower transactions. An example of a Decentralized Exchange Center would be the platform UniWasp.

The design of a vault which is presented in this report is from the perspective of a CEX because, in terms of trading volume, centralized exchange centers are much more relevant,

having reached volumes higher than 60 Billion Dollars in 7 Days; more than 7 times the maximum 7-day-trade volume of DEX [2]. There are currently two large CEX that offer a crypto-vault service in the sense of the presented design in this report: Coinbase and Bitcoin Suisse. This report therefore bases some of the findings on these two companies. To give the reader an overview on the two companies, a short portrait of the two is given in the following sections:

Figure 5.1: Logo of Coinbase [24].

1. `Coinbase` is an American crypto exchange center that was founded on January 1. 2012 [24]. They are one of the biggest crypto exchanges. According to their website, they have a quarterly trade volume of over 327$ billion, operate in over 100 countries and have over 73 million users currently. They provide an easy solution for customers to buy, sell and trade cryptocurrencies and manage their portfolios. To store the crypto assets, they provide two solutions: a wallet and a vault.

Figure 5.2: Logo of Bitcoin Suisse [11].

2. `Bitcoin Suisse` is a Swiss-based crypto asset management company and founded in 2013 [11]. They specialize in the trading of cryptocurrencies, help customers with staking (a way in which owners of proof-of-stake blockchain-based cryptocurrencies can earn rewards), and they also provide services when it comes to large crypto trades in their Prime Brokerage section. According to their website, they currently manage over CHF 5 billion in cryptocurrencies. As Coinbase, they also added a crypto-vault service to their provided services.

5.1.3 Types of Storage Services

In this section, the different types of storage solutions the crypto market offers at the time of writing are introduced. The summarized types of storage can be seen in figure 5.3:

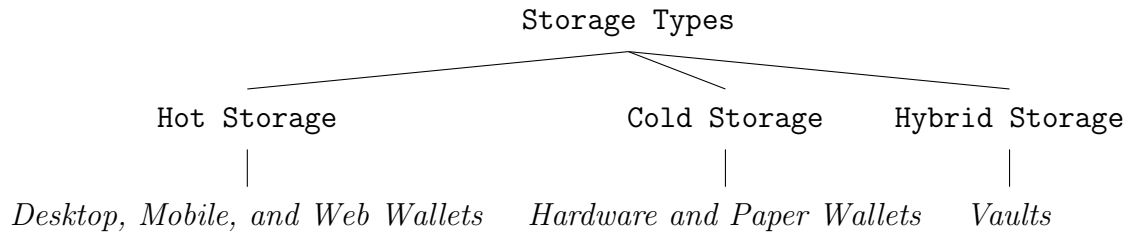

Figure 5.3: Different Types of Key Storage [3].

There are two types of storage: hot and cold storage. `Hot storage` is when a wallet is in some way connected to the internet. The funds are accessed via a digital private key (*cf.* section 5.2.3). Hot storage has the properties that it is very user-friendly in that the assets stored in a hot wallet are easily accessible and transferable, but the downside is the lack of security in comparison with cold storage options [1]. In the current market, there are 3 types of hot storage solutions:

1. `Desktop wallets` are wallets where a user has to download software to store their keys in. This software is then locally run on a personal computer [3].

2. `Mobile wallets` are quite similar to desktop wallets but the software is optimized for smartphones [3]. Mobile wallets allow a user to make transactions much easier by scanning a QR code to send funds to other wallet addresses [13]. One example of a desktop and mobile wallet would be the Exodus Bitcoin and Crypto Wallet [6]. It is an application that can be downloaded either to a personal computer or as an app on a smartphone or both. A user can then either use their wallet on their PC or on their phone to make transactions. The two applications are synchronized.

3. `Web wallets` are a type of storage solution, where no software has to be downloaded. Web wallets can simply be accessed through a browser. When a user initially buys cryptocurrencies, the keys are typically first stored in a web wallet offered by the exchange [3]. An example of a Web-based wallet would be Metamask [7], which is used for many applications in decentralized finance (DeFi). This wallet can be installed as a browser add-on, which makes it very easy to access for a user but also a target for hackers. Another technical implementation of web wallets are wallets that are provided on exchanges such as Binance or Coinbase. Here the keys are stored and managed by the exchange platform as a custodian.

When compared in terms of security, web wallets are the solution that poses the most risk. This is because a web wallet that is provided by a centralized exchange platform is an extremely valuable target for an attacker, as those exchanges store millions of dollars in cryptocurrencies. This is why an exchange platform must have a complex security system in place to prevent such attacks.
To combat some of the risks stemming from the accessibility of hot storage, there is another type of storage, which is called `Cold Storage`. Its characteristics are that the wallet is, for the most part, not connected to the internet. In contrast to hot storage, cold storage has the upside of more security but the downside of a lack of accessibility of the assets stored in cold storage [1]. There are two types of cold storage solutions.

1. `Hardware wallets`, are physical electronic devices to store private keys on. They are only connected to the internet during the process of a transaction - when sending or receiving funds. After the transaction, these devices are stored offline [3]. These types of wallets often support multiple cryptocurrencies. An example for a hardware wallet, or more specifically for a USB storage wallet, is the Ledger Nano X [14] which was introduced in 2019 and allows a user to manage and store more than 1800 coins and tokens.

2. `Paper wallets` are the last type of wallet. When using a paper wallet a user prints the private key onto a piece of paper [8]. This type of wallet was primarily used in the early days of cryptocurrencies but is now replaced by other solutions such as the ones mentioned above. The reason is that while they are very secure when a user takes the necessary precautions and stores the paper in a place that cannot be accessed by any unauthorized person, paper wallets have some major drawbacks when it comes to usability. For example, when the paper is destroyed or lost, a user

will not be able to recover his or her wallet. Another point of failure is the printer. An attacker might try to infiltrate the printer that was used for the printing of the private key. Another problem is, when a user uses a low-quality printer or ink, so that the color of the print fades over time and becomes illegible, which also means that the assets are lost.

Recently, the line between hot and cold storage has become more blurry due to the development of `hybrid storage solutions`. The primary idea behind `vaults` is to combine the security of cold storage with the accessibility and user-friendliness of hot storage [5]. In the next section, a high-level design for a vault for crypto exchange centers is introduced.

5.2 Vaults

As the trade volume of cryptocurrencies increases, the need for additional security is a major issue. When cryptocurrencies bought on an exchange platform are ”transferred” into a wallet for safekeeping, the coins and tokens are not truly stored in the wallet, they remain on the blockchain. A wallet simply stores keys that point to the coins on the blockchain. So, through these keys, one can claim ownership of one's coins and token. Wallets are therefore just simple databases to store digital keys in. Those keys are completely independent of the corresponding network. The keys are generated and managed by the wallet software but these methods are sometimes vulnerable to attacks. In the following sections, a design for a more secure way of storing cryptocurrencies is introduced: A digital crypto-vault [3].
A crypto-vault is a cryptocurrency storage solution that provides additional layers of security compared to digital wallets, ensuring assets such as Bitcoin are securely stored and protected from a multitude of risks [38]. Looking at it through a more traditional lens, a crypto-vault can be seen as the crypto version of a bank saving account. A crypto-vault is designed to be more secure, reliable, and managed by more than one user. In the next section, the risks that users of wallets face will be introduced and used to justify the need for a more secure storing system.

5.2.1 The Need For Vaults: Security Risks of Wallets

Security is one of the main reasons why vaults exist. So, it is important to justify the need for a more secure solution to protect crypto assets from theft. In their work, Froehlich, Gutjahr, and Alt have identified 3 main sources of risks users of cryptocurrencies face: Risk of human error, risk of betrayal, and the risk of malicious attacks [16]. These findings can also be applied to the engagement with wallets.
Looking at the `risk of human error` [16], what is meant is the risk that a cryptocurrency owner loses his assets by mistake. When users manage their funds in a wallet, there is for one the risk of losing access to the wallet and in the worst case also the access to the recovery seed phrase. This in turn means that there is no possible way to restore the keys in the wallet. Vaults can solve this problem in that the users give a vault provider custody over their assets, so the risk of losing access to the vault is mitigated under the assumption, that there are sufficient security and recovery measures taken by the custodian to prevent such a worst-case scenario. Those precautions would be discussed later on in this paper.
The second possibility for a user to lose his or her assets is to make a wrong transaction. A faulty transaction can happen, when the keys are transferred to a wrong or non-existing address. As soon as a transaction is validated on the blockchain, it is irreversible due to the principles of the blockchain. A vault system can also give a user a second layer of protection in this case by introducing a time delay during which the transaction is not

yet being sent to the blockchain and in which a user can abort the transaction. This implementation is discussed in section 5.2.3.3.
The `risk of betrayal` [16] might not be as relevant to wallet users in a technical sense but is relevant when it comes to the decision of which wallet or vault provider to trust. There were cases such as Bitconnect in the past where an exchange was later exposed as being a Ponzi scheme which left users with massive financial losses [15]. When a user decides to give custody over his or her assets to a third party, he must be sure that there is a high level of trust. In general, users should only consider trustworthy, well established and audited firms as custodians. To provide such a layer of trust the security system of Bitcoin Suisse, for example, is audited by PwC, Grant Thornton AG, and Zuehlke and is, therefore, ISAE 3402 certified [10].
The last risk users of wallets face is the risk of `malicious attacks` [16]. In section 5.1.3 it is mentioned that hot storage is generally less safe than cold storage due to its easier accessibility through the internet. A wallet is also a single point of failure, and therefore an "easier" target for an attacker because there are fewer layers of protection compared to vaults. In section 5.2.3.2 this matter is discussed further and it is shown how a vault manages to protect the assets against malicious attacks.

5.2.2 Multiple Asset Support and Governance

To offer a high-value service for customers, a vault should `support multiple assets`. Here, taking Coinbase as an example, in their Usage and Trading Statistic of 2021 they reported a total sum of $223 billion in assets on their platform during their first quarter of 2021, with Bitcoin and Ethereum representing 83% of those assets (which is summarized in table 5.1) [9]. Looking at the trading numbers in table 5.2 one can see that other crypto assets accounted for 44% of the trading volume of cryptocurrencies in 2020. This is an increase of 26 percentage points from 2019.

Table 5.1: Share of held cryptocurrencies on Coinbase according to their 2021 report [9].

Asset	Share
Bitcoin	70%
Ethereum	13%
Other Crypto Assets	13%
Fiat	4%

For a vault to offer the most compelling service, it is important that users can store many different cryptocurrencies in their vault. Coinbase, for example, currently supports 108 crypto assets [9], and Bitcoin Suisse states on its website, that they support next to Bitcoin, Ethereum, Polkadot, and Cardano also all ERC-20-Token and Tezos FA1.2/2-Token [10]. In an ideal setup, a vault system should support at least as many coins and tokens as a hardware wallet like the Ledger Nano X which currently supports more than 1800 at the moment (as mentioned in section 5.3) [14].

5.2.3 Key Management

This section introduces a concept for the `key management` in a vault and makes a comparison to the key management of a wallet. For that, the concept of distributed key generation, transaction management in a vault, and where companies store the hardware devices that contain the keys are introduced.

Table 5.2: Share of Traded cryptocurrencies on Coinbase according to their 2021 report [9].

Asset	Share (2020)	Share (2019)
Bitcoin	41%	58%
Ethereum	15%	14%
Other Crypto Assets	44%	18%
Litecoin	No data	10%

5.2.3.1 Key Generation in General

In a first step, the general concept of the key generation that is happening in most wallets is examined. So, when looking at how cryptocurrency coins and tokens bought on an exchange platform are stored, it has to be mentioned that the coins or token themselves remain on the blockchain and are not directly stored in a wallet. Wallets are simple databases that store pairs of keys: a private key and a public key. Those key pairs are used to establish ownership of the assets on the blockchain. To explain the key generation process this report uses the Bitcoin Network as an introductory example [20]. For that, it has to be mentioned that the implementation of the key generation might vary on different blockchains but the underlining concepts stay the same.
When a user initially creates a wallet there is usually a seed created. A seed is typically a series of 12 or 24 words, also called a mnemonic phrase [17]. This seed is used for two reasons, first to derive the individual keys used for transactions, and secondly to restore the wallet in the case of access to the original wallet being lost. For example when the device where the wallet was saved was destroyed.
The second step in the storing process of bitcoin is the generation of a `private key`. In case of the bitcoin network, the private key is a random number that is generated by the underlying operating system's random number generators that create a 256-bit number. Through a private key, the information can be encrypted and decrypted, so anyone with access to the private key has access to the assets. The user can think of the private key as the pin to a bank account. The private key must be stored securely and protected from unauthorized access [20].
Because the private key can't be shared with others, there has to be another key: the `public key`. The public key is calculated from the private key and enables a user to make transactions on the blockchain, in this case sending and receiving assets. As an analogy to the traditional banking system, the public key can be compared to a bank account number. Like a private key, a bank account number is used in transactions but doesn't allow for others to access the bank account itself. To calculate the public key a mathematical function called elliptic curve multiplication is used [20].
`Elliptic Curve Multiplication (ECC)` is a type of asymmetric encryption meaning it is a one-way function that can easily calculate the public key from the private key but not vice versa [19]. ECC is an alternative to the Rivest-Shamir-Adleman (RSA) algorithm that is also used to encrypt data.
To make a transaction, the Bitcoin Network requires `(Bitcoin) Addresses`. The Bitcoin Address is calculated from the public key by the use of a hashing function. This is done to turn the public key into a string of numbers and letters of a length between 26 and 35 characters. Bitcoin uses the Secure Hash Algorithm (SHA) and the RACE Integrity Primitives Evaluation Message Digest (RIPEMD) algorithms. First, the public key is put into SHA hashing function, and the SHA256 hash is calculated, and from that the RIPEMD160 hash, which gives a 160-bit number. To help with human readability and avoid ambiguity, this last hash is presented to the user in a "Base58Check" encoding.

This is the address that a user typically sees and uses inside a wallet. An overview of the individual steps in the key generation is given in figure 5.4 [20].

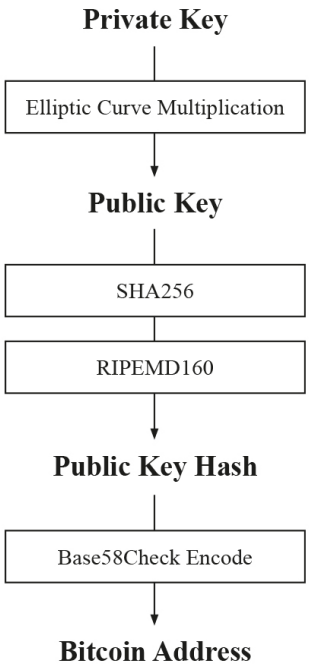

Figure 5.4: Key generation on the Bitcoin Network [20].

Usually, wallets create new public keys for each new transaction, so that it becomes very difficult to track the wallet users' payment history [17]. This guarantees a certain level of anonymity on the network although the Bitcoin Network is not completely anonymous. The problem still is that there is usually only one private key generated and stored in a single wallet which creates a single point of failure. In the next section, the differences in the key generation in wallets and vaults are discussed and the concept of Multiparty Key Generation (MPC) is introduced.

5.2.3.2 Multiparty Key Generation

In the case of a vault, a user gives custody of his or her keys to the company that offers this storage service. Such a custodian, therefore, has to take measures to protect those keys. To make the key generation process more secure, it can be proposed the use of `Multiparty Computation (MPC)` [21]. MPC is a protocol that allows multiple users or entities to be part of the key generation process. In the case of Coinbase the private key is split into multiple shares using `Shamir Secret Sharing (SSS)` and the original private key is deleted in the process.
SSS is a secret-sharing protocol that allows for a secret to be split into multiple parts (or shares) and to be distributed to multiple entities, but with the property that there is only a certain number of those entities needed (and therefore shares of the secret) to reconstruct the original secret [22].
By the distribution of the private key, a very secure system is provided, where there is no single point of failure anymore. The distribution makes it much more difficult for an attack because, in contrast to wallets, an attacker now would have to break into multiple distributed systems to gain access to the funds stored in the vault. In the next section, the discussion is going to focus on how the management of a transaction in a vault works [21].

5.2.3.3 Transaction Management: Threshold Signing Service

It has been seen that most wallets automatically create new public keys and addresses every time a transaction is made. This raises a problem, especially with the recent development of cryptocurrency staking which requires multiple usages of the same address. MPC solves this by using the distributed private key shares to create a valid signature to sign transactions on the blockchain. Each key share is used to create a partial signature and those partial signatures are then combined into a single valid signature. This means that there is no need to recombine the individual shares back into a single private key; the shares always stay split and distributed. This process is called a `Threshold Signing Service (TSS)` [21].
There are 5 steps in the MPC key generation process that create a TSS. (1) The first step is computing the so-called Party Keys. This is a set of private and public keys, that are generated in a trusted environment. They are then stored on Hardware Security Modules. This makes sure that no one can access the keys without having physical access to those modules. (2) From the Pary Keys, a set of Signing Keys are created that are later used for the Threshold Signing Service. The keys generated in step (1) are used to encrypt those Signing Keys. Those two steps are only done only once in the lifetime of a Signing Key [21].
The next steps are repeated every time a transaction needs to be signed to be validated on the blockchain. (3) In the first step of the Signing Protocol, the owners of the Signing Key Shares generate Nonce Values. A Cryptographic Nonce Value is an arbitrary number that can be used in cryptographic communication [26]. The generated values are then sent to all parties. (4) The Nonce Value is then used to generate the Partial Signatures in the next step. (5) In the last step those Partial Signatures are then combined to get a valid signature, that can then be used to sign the transaction [21].

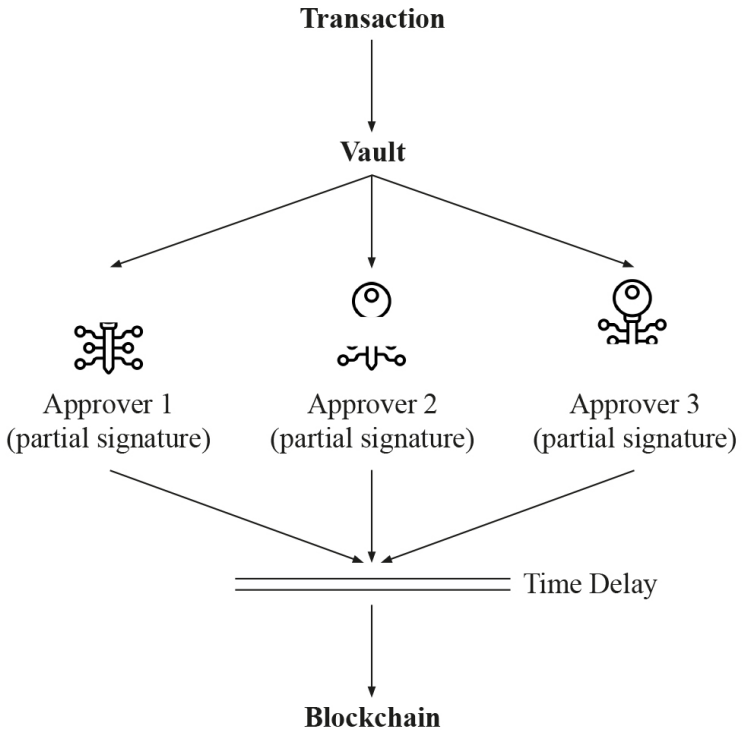

Figure 5.5: Simple representation of a transaction in a vault. When a user signs in to his or her account and initializes a transaction, the TSS protocol is activated and if m our of N approvers authorized the transaction and the waiting period (time delay) is over, the transaction is validated on the blockchain.

The `transaction management` in a vault refers to where the TSS is implemented with a time delay to add another layer of security (*cf.* Figure 5.5). In a wallet once a user initialized a transaction it is immediately signed and validated on the blockchain, so once authorized, a transaction from one wallet to another is irreversible [25]. However, in a vault, once a transaction is initialized, the validation on the blockchain is delayed by a specific time period, e.g. 48 hours [23]. During this period the holder of the Private Key Shares have to approve the transaction and only after the specified threshold of m out on N users is met and the delay period is over, the transaction is signed and validated on the blockchain [21]. Additionally, the transaction can also be aborted at each step during the delay period. This solution adds another layer of security to mitigate the risk of human error and malicious attacks mentioned in section 5.2.1 by allowing a user to abort a faulty transaction.
In the case of an institution, it makes sense that the entities holding the shares are different people. There could be one person in a company initializing the transaction, and one or multiple other instances to check if the transaction is indeed correct. In the case of a single user, it makes more sense to think of those entities as multiple email accounts or phone numbers, where one user has to approve the transaction. This still adds more security because an attacker would have to break into multiple accounts [23].

5.2.3.4 Physical Key Storage

Having now established how MPC and TSS can be used for a vault to provide a very secure protocol for the key generation and the management of transactions. But as seen in the previous section 5.2.3.3 that the keys are managed by a centralized exchange platform. Naturally, there are concerns about what happens when such an exchange is compromised [21]. In this section the concept and market solutions for the physical storage of sensitive data such as the private key (shares), and the Hardware Security Modules, are discussed. The first question to answer is, which characteristics a country or place should have to be considered a safe storing location. For that, we first looked at the US-News 2021 "Best country Rating" and especially the criteria for quality of life. The ranking is based on survey data and conducted by the global marketing communications company VMLY&R, and the Wharton School of the University of Pennsylvania [27]. That way it can be determined the best indicator of a safe data storage location is the categories economically stable, politically stable, and safe. These attributes have been chosen because they represent safety and stability, which are very important properties that a safe data storing location should have. The points scored in those attributes are summarized in table 5.3. All listed countries score very high in those attributes, with Switzerland and Canada scoring the highest in these particular attributes. When these results are compared with the results from the Data Danger Zones Report conducted by the data center Artmotion based on independent data from the United Nations, World Economic Forum, Transparency International, Global IntAKE and Control Risk, where they assess countries in terms of safety for data storage, Switzerland also scores highest there [28]. The results from the Data Danger Zones Report are summarized in table 5.4.
In conclusion, we argue that all countries on those lists can be used for safe data storage and it depends more on the locations of the vault provider which countries and location are the best options. The safest solution should be to use geographically distributed data storage centers.
Bearing that in mind, it is feasible to propose a market solution for a hyper-secure data storage location by an example of re-purposed military bunkers located in the Swiss Alps. The name of the owner of the bunker, which is located in Uri in Switzerland, is Xapo. Xapo is an online bank that also offers very secure storage for Bitcoins [30]. The idea behind this bunker is to prevent unauthorized access to the private keys that are stored

Table 5.3: Highest Scoring Countries for Safety and Stability [27].

country	economically stable	politically stable	safe	overall score
Switzerland	98.7	100.0	100.0	298.7
Canada	99.9	100.0	96.7	296.6
Norway	92.3	95.9	97.1	285.3
Denmark	89.0	96.9	97.5	283.4
Sweden	94.5	95.4	93.3	283.2
Australia	96.2	95.3	91.3	282.8
New Zealand	88.9	93.1	95.9	277.9
Netherlands	89.5	97.1	90.5	277.1
Finland	82.8	90.1	94.7	267.6
Germany	100.0	91.4	71.3	262.7

Table 5.4: Results from the Data Danger Zones Report [28].

country	data risk score
Switzerland	1.6%
Singapore	1.9%
Iceland	2.3%
Luxembourg	2.6%
Hong Kong	3.6%
Taiwan	3.9%
Austria	5.2%
New Zealand	5.2%
Portugal	6.9%
Denmark	7.6%
Finland	7.6%
Lithuania	7.6%
Norway	7.9%
Sweden	7.9%
South Korea	8.3%

offline. On one hand, the former military bunker provides a very high level of security against physical attacks, because it was initially built during the cold war to withstand even nuclear attacks. Although this is interesting for military purposes, the likelihood of such events is extremely small and it can be argued that such a high level of physical safety may not be necessary for a data center to be considered as a secure crypto storage location. What is more interesting is the access security system and the principles that can derive from it. In case of the bunker owned by Xapo, they take photographs, credentials, and fingerprints before anyone enters the actual bunker. The second step in the security concept is called a Mantrap [29].

A `Mantrap` is in its simplest implementation a small room with two doors. At the first door, an entrant presents his or her credentials as mentioned before. When the entrant steps inside, the first door closes and they additional credentials have to be given while waiting in this room. In the case of the bunker owned by Xapo, the vein pattern of the hands is required to enter the next section of the bunker [32].

Inside of the bunker are multiple armored steel doors that are closed every night. The last step in access management is a sealed door to the data center. It is sealed with tape

to guarantee that no one has entered the room. The only reason to enter the room is when a client wants to withdraw their assets [29].
In conclusion, we argue that this system only makes sense for institutional or ultra-high-net-worth individuals that want to store their assets long term, because the process of manually transferring the assets out of cold storage is extremely costly and time consuming, and considering an exchange center such as Coinbase or Binance with millions of customers it would be nearly impossible to to accomplish a transaction this way. There is, however, not much publicly available data on the access management and physical storage of the keys for these types of exchanges because this is most likely not in the interest of these companies to make such information public.

5.3 Security Measures by Exchange Centers

With the increasing demand in blockchain technology and cryptocurrencies, the crypto exchange centers have grown exponentially, and with it the demand for more advanced security systems, processes, and insurance policies. After understanding the main logic behind the security of the key generation process and the physical security of the key storage facility, it is important to also look at the more general security aspects that are not specifically implemented for a crypto-vault, but that surround the vault and the firm that provides this service.

5.3.1 Security Configuration and vulnerability Management

A first step in analyzing what security measures an exchange platform needs to take, is thinking about potential risks or vulnerabilities such a crypto exchange center experiences. In figure 5.6 we see a simplified representation of different aspects of such a security configuration. The firm, in this case the vault provider, is in the center and interacts with different entities. The employees are one of the possible vulnerabilities. Because the risks of human error and malicious attacks also extend to the employees and not only to the user itself. So, to protect the company, the security concept starts with the hiring process. To ensure a firm only hires trustworthy employees, the firm needs to check the background of each potential employee. After the hiring process, the firm also needs to keep its employees up to date when it comes to potential risks such as phishing emails and malicious software and there has to be a system that manages the access privileges of employees and determines clearly who has access to which data and location [31].
On the other hand, there is the interaction of the firm and its users. It is standard that the firms encrypt their web communication. For example, the website traffic of Coinbase flows over HTTPS Encrypted SSL. On top of that, many security configurations are required for both users and suppliers. These configurations, especially in crypto-vaults, are the key to success. An example of such a requirement would be the double factor authentication that is required for every account on the system by companies such as Coinbase [18]. As an added precaution, in Coinbase, the system configuration establishes that less than two percent of all customer funds and data are stored online in hot storage. This provides an extra layer of security because data stored in cold storage is much more difficult to access for hackers in comparison to data that is accessible through the internet. The sensitive data that Coinbase receives (e.g. passwords) is encrypted through a hash function before it is stored [31].
To prevent future attacks, it is common practice to use different processes and programs that help to find holes and vulnerabilities in the security system. In the case of Coinbase the ”Bug Bounty Program” is dedicated on this task. It is a service offered by

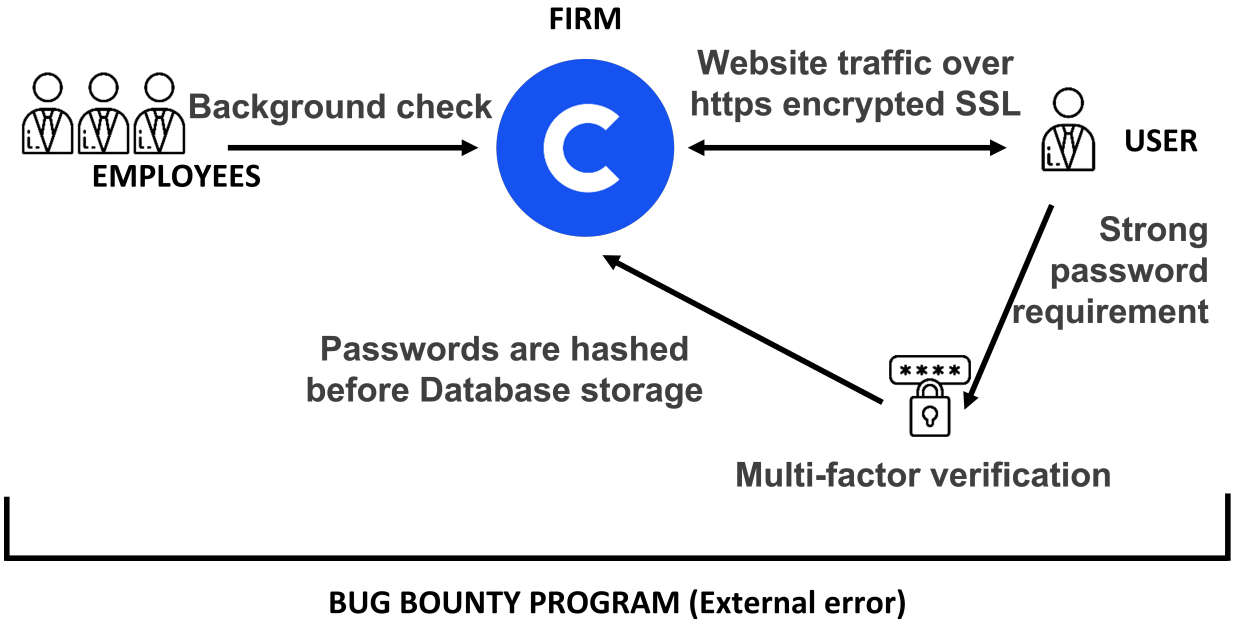

Figure 5.6: Security configuration based on Coinbase [31].

many websites, organizations, and software developers, that exchanges compensation and recognition for reporting bugs as a precautionary action against future attacks [12].

5.3.2 Client Data Management

As the General Data Protection Legislation (GDPR) gets more complex, attackers are also developing new ways to steal data. So for crypto exchange centers nee to have a well-developed system that manages the client data safely and accurately. Especially in more holding-oriented systems, that can potentially manage a great volume of delicate transactions.

The right selection of the storage type can substantially lower the probability of certain kinds of attacks. That is why according to Coinbase Inc. the storage of 98 percent of customer funds takes place offline, this means that the majority of its connection with the internet is just when an operation takes place [31]. This reduces the chances of client data access by entities with malicious intent.

But not only malicious entities are roadblocks that the company have to tackle when hosting a crypto-vault, but also the failures of the different nodes of the systems can be dangerous. A natural disaster, a pandemic, or any other issue could lead to a disconnection or isolation of the servers that can result in major failures on the company platform and also, make the data more vulnerable. To solve this, the client data is frequently backed up and stored distributedly with redundancy, in the case of Coinbase Inc, it is distributed along the different servers ensuring the continuity and the security of this data [31].

Apart from the client data that is stored on multiple, distributed servers with redundancy, to ensure it is completely safe, CEX have some solutions that include cold storage. To store data in cold storage, it can either be stored as physical paper copies or in backups on 140 USB drives that follow the Federal Information Processing Standards (FIPS) [31].

Lastly, when storing client data, in the case of big companies, the data is encrypted so in case of a breach, the data is not directly exposed. In the case of Coinbase, the stored data is encrypted according to the Advanced Encryption Standard established by the United States National Institute of Standards and Technology, which requires ”AES-256 ENCRYPTION” of data [31].

5.3.3 Disaster Recovery and Business Continuity Management

Every company should not only be prepared for smaller inconveniences, but also for major and potentially disastrous problems that could occur. This is why in order to provide a continuity of the offered services, companies should have a business continuity and disaster recovery plan. It enables the company to get the services provided running again as fast as possible and makes the process of getting back into full operational capacity easier and more efficient. This reduces the risk of harm to the companies reputation and reduces the time period in which sensitive data may be vulnerable, and therefore decreases the chance of a loss in value and trust for the company, and generally provides a guideline of how to react in certain situations.
Several different solutions can provide internal and external business continuity and data recovery services. One solution, for example, is an API plug-and-play platform, and another solution are government imposed safety standards. Sometimes the implementation is done by the crypto exchange centers themselves, but there are also external solutions. For example, in small Exchange centers an internal solution may be too expensive and an external solution such as the one offered by Coincover could potentially be more efficient [4].
In light of this, Coincover provides to crypto exchange centers safety standards and specific insurances built into a plug-and-play platform, which improves the security measures, adding additional protection against attacks and fraud, the ability to offer clients cover options, business continuity planning, and assistance and also helps facilitate security storage with government-grade standards [4].

5.3.4 Accessibility and creation of vaults

The theoretical concept of a crypto-vault can be quite overwhelming to someone who is only interested in the usage of such a service. So, in this section the creation of a crypto-vault from a users perspective is shortly discussed on the example of the steps necessary to set up a crypto-vault on Coinbase [23].
The vault can be set up and accessed by any user through 2 devices and a required minimum of two email accounts that approve the transactions. In the overview of the asset the user has to press the vault icon, then insert a name for the vault and would be redirected to a safer page where the configuration would take place, selecting who will approve the withdrawals and transactions (note that has to be 2 or more users) and the number of coins that are going to be transferred from the wallet to the vault. After the configuration, the users receive e-mails to the provided email addresses to confirm the creation of the vault.
After completing these steps the transaction takes 48 hours (in the case of Coinbase) until the digital currency is available in the vault.

5.4 Discussion

crypto-vaults for centralized crypto exchange centers offer a great level of security that builds on top of traditional crypto wallets. Security may not be the most exciting topic in the world of crypto but it still is one of the most important aspects (if not the most important) when it comes to the long-term usage of cryptocurrencies. We introduced a high-level design of a crypto-vault in the previous sections and explored the technical aspects as well as the security infrastructure that surrounds and makes up a crypto-vault. We now want to discuss the potential, but also the disadvantages of crypto-vaults.
First, we look at the use case of crypto-vaults. With their high level of security, crypto-vaults are a very interesting solution for long-term holders of cryptocurrencies. The delay

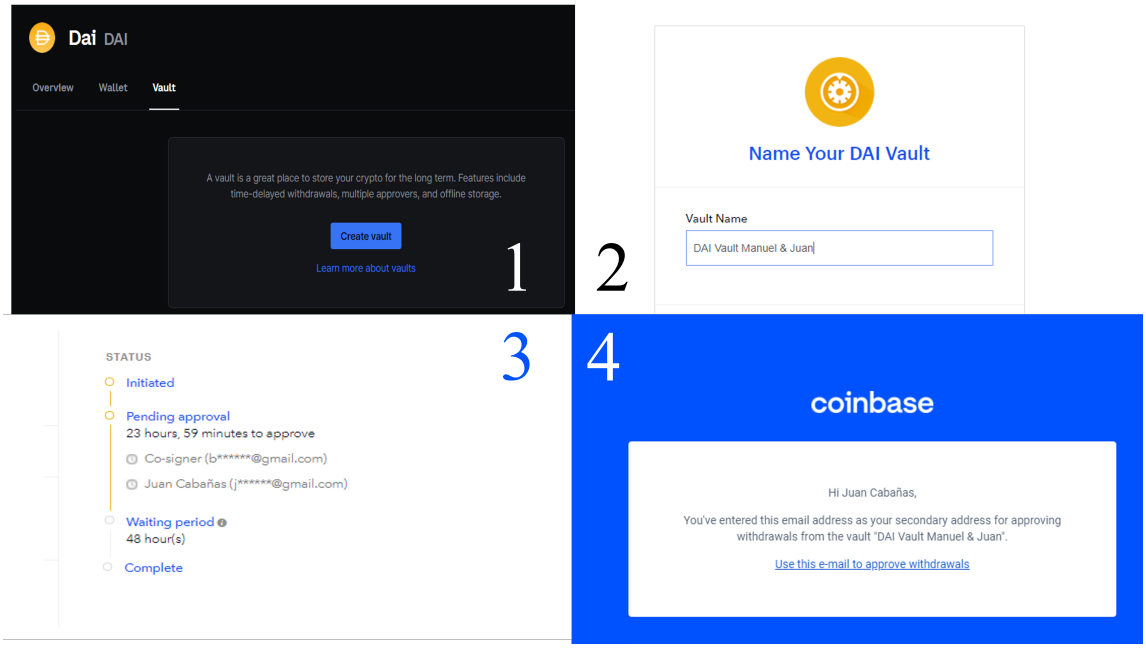

Figure 5.7: Step by Step process of creating a crypto-vault on Coinbase.

period that is incorporated into a vault and the threshold signing service protocol makes the vault a lot more secure than regular wallets. For long-term holders, this is an advantage, but at the same time, this makes vaults extremely unpractical for short-term holders and (day)traders. We believe that crypto-vaults have much potential for corporations that want to invest in cryptocurrencies and for high-net-worth individuals, both to hold the assets for a longer time. We have seen more and more institutional money flowing into the crypto space with firms such as Blackrock and Tesla. But as with many things in crypto, only time will tell if this trend is here to stay.

A major disadvantage of the delay time is also the volatility of the crypto market. Since the introduction of Bitcoin, there were multiple crashes in which coins and tokens can lose a lot of their value within hours or days. Not only could this prevent many institutional investors from investing in crypto assets but it also affects the usage of vaults. In the case of a crash, a vault user wouldn't be able to sell his or her assets immediately, and the assets transferred will lose a lot of their value during the delay period. On the other hand, this could also be seen as another layer of protection in the sense that it prevents a user from panic selling assets in a crash, and protecting the investor from amassing huge permanent losses.

Another open question is, how a design of a vault for decentralized exchange platforms would look like. The design shown in this report only works for centralized platforms because one counts on the fact that there is a central authority that provides the security system surrounding a vault. And a CEX also makes the auditing process a lot easier. And an audited, centralized provider is much more likely to gain trust from individual and institutional investors. It is impossible to predict if decentralized platforms can overcome such trust issues and gain a larger market share in the future at the cost of CEX. This would mean that the design explained in this report would need to change in order to keep its relevance. But apart from that, CEX have the issue, that there is a single point of attack. In the case of the Multiparty Key Computation and Threshold Signing Service, the single point of attack was resolved, but by giving custody of the keys to a CEX, this point only shifted from the wallet to the security system of the exchange platform. Nevertheless, we still think the advantages of a vault in terms of security and usability outweigh the potential risks of a CEX because as we mentioned in section 5.3.3 there are ways to insure assets even in the case of a data breach or other disasters.

In conclusion, crypto-vaults are not a revolutionary new technology, but they do combine interesting cryptographic concepts with more traditional security systems, and offer new possibilities for secure long-term storage of cryptocurrencies. Will they replace our regular bank account? Most likely not, but they can be seen as an extension of our bank account into the crypto space.

Bibliography

[1] P. Smith, N. Srivastava, *A Market Overview of Custody for Digital Assets*, Deloitte. Digital Custodian Whitepaper, 2020, doi:https://www2.deloitte.com/content/dam/Deloitte/xe/Documents/finance/me_Digital-Custodian-Whitepaper.pdf.

[2] *Decentralized Exchange Center Data*. theblockcrypto.com https://www.theblockcrypto.com/data/decentralized-finance/dex-non-custodial (accessed Nov. 29, 2021).

[3] *What are the Different Types of Cryptocurrency Wallets?*. europeanbusinessreview.com. https://www.europeanbusinessreview.com/what-are-the-different-types-of-cryptocurrency-wallets/ (accessed Nov. 15, 2021).

[4] *Coincover General*. coincover.com https://www.coincover.com/ (accessed Nov. 29, 2021).

[5] *crypto-vaults*. daytrading.com. https://www.daytrading.com/crypto-vaults#:~:text=Crypto (accessed Oct. 29, 2021).

[6] *Exodus Bitcoin & Crypto Wallet*. exodus.com. https://www.exodus.com/ (accessed Oct. 29, 2021).

[7] *A Crypto Wallet & Gateway to Blockchain apps*. metamask.io. https://metamask.io/ (accessed Oct. 29, 2021).

[8] *Paper Wallet*. Investopedia.com; https://www.investopedia.com/terms/p/paper-wallet.asp#:~:text=A(accessed Oct. 29, 2021).

[9] B. Dean, *Coinbase Usage and Trading Statistics (2021)*. backlinko.com. https://backlinko.com/coinbase-users (accessed Nov. 21, 2021).

[10] *Custody*. bitcoinsuisse.com. https://www.bitcoinsuisse.com/vault (accessed Nov. 17, 2021).

[11] *Your Swiss Cryptocurrency Investment Partner*. bitcoinsuisse.com. https://www.bitcoinsuisse.com/ (accessed Dec. 13, 2021).

[12] *Bug Bounty Program*. hackerone.com. https://hackerone.com/coinbase?type=team (accessed Nov. 27, 2021).

[13] B.Nibley, *What Is a Crypto Wallet? Understanding the Software That Allows You to Store and Transfer Crypto Securely*. businessinsider.com. https://www.businessinsider.com/crypto-wallet?r=US&IR=T (accessed Nov. 26, 2021).

[14] *Securely Start Your Crypto Journey.*. ledger.com. https://shop.ledger.com/pages/ledger-nano-x (accessed Nov. 26, 2021).

[15] H. K. Baker, J. R. Nofsinger, V. Puttonen, *Other Frauds and Scams that Lure Unsuspecting Investors. In The Savvy Investor's Guide to Avoiding Pitfalls, Frauds, and Scams*, Emerald Publishing Limited, Bingley, 2020. pp. 149–177, doi: https://doi.org/10.1108/978-1-78973-559-820201010.

[16] M. Froehlich, F. Gutjahr, F. Alt, *Don't lose your coin! Investigating Security Practices of Cryptocurrency Users*, in ACM Designing Interactive Systems Conf. (DIS '20) Proceedings, 2020. pp. 1751-1763, doi: https://doi.org/10.1145/3357236.3395535.

[17] *How Bitcoin Wallets Work*. river.com. https://river.com/learn/how-do-bitcoin-wallets-work/ (accessed Oct. 29, 2021).

[18] *Coinbase security*. coinbase.com. https://www.coinbase.com/security (accessed Nov. 29, 2021).

[19] *Elliptic Curve Cryptography*. avinetworks.com. https://avinetworks.com/glossary/elliptic-curve-cryptography/ (accessed Nov. 17. 2021).

[20] A. M. Antonopoulos, *Mastering Bitcoin: Unlocking Digital Cryptocurrencies*, 2nd ed. California, CA, USA: O'Reily Media, 2017, pp. 116-158.

[21] A. Raghuvanshi, *Production Threshold Signing Service*. coinbase.com. https://blog.coinbase.com/production-threshold-signing-service-b16017c09661 (accessed Nov. 17, 2021).

[22] A. Shamir, *How to Share a Secret*, Commun. ACM 22, 612–613. Nov. 1979, doi: https://doi.org/10.1145/359168.359176.

[23] *Vaults*. coinbase.com. https://help.coinbase.com/en/coinbase/getting-started/other/vaults-faq (accessed Nov. 17, 2021).

[24] *Vaults*. coinbase.com. https://www.coinbase.com/ (accessed Dec. 13, 2021).

[25] *Can my Transaction be Canceled or Reversed?*. blockchain.com https://support.blockchain.com/hc/en-us/articles/211162263-Can-my-transaction-be-canceled-or-reversed- (accessed Nov. 17, 2021).

[26] P. Rogaway, *Nonce-Based Symmetric Encryption*, in Bimal Roy. Willi Meier (eds.). Fast Software Encryption. Lecture Notes in Computer Science. 3017. pp. 348–358, doi: 10.1007/978-3-540-25937-4_22.

[27] *Quality of Life*. usnews.com. https://www.usnews.com/news/best-countries/quality-of-life-rankings (accessed Nov. 22, 2021).

[28] *Data Danger Zones*. artmotion.eu. https://artmotion.eu/_Resources/Persistent/d883b8e573e4b8bc838af1793d07aeaf67416f48/DataDangerZones.pdf (accessed Nov. 22, 2021).

[29] J. I. Wong, *Millionaires Are Stashing Their Bitcoin Fortunes in a Fortified Bunker Under a Swiss Mountain*. CBC Radio. https://www.cbc.ca/radio/day6/episode-371-iran-nuclear-deal-plastic-toy-waste-kaepernick-bitcoin-bunkers-spotify-vs-composers-and-more-1.4470486/millionaires-are-stashing-their-bitcoin-fortunes-in-a-fortified-bunker-under-a-swiss-mountain-1.4470521 (accessed Nov. 22, 2021).

[30] *Unparalleled Security*. xapo.com. https://www.xapo.com/vault (accessed Nov. 22, 2021).

[31] O. Dale, *Coinbase Safety*. blockonomi.com. https://blockonomi.com/is-coinbase-safe/ (accessed Nov. 23, 2021).

[32] J. Clark, *What Is a Mantrap and Do You Need One?*. www.datacenterjournal.com. https://web.archive.org/web/20181220135005/datacenterjournal.com/what-is-a-mantrap-and-do-you-need-one/. (accessed Nov. 22, 2021).

[33] V.Motolani, *Trading on a CEX vs. DEX. What's the Difference and What Are the Risks?*. https://trustwallet.com/blog/trading-on-cex-vs-dex. (accessed Nov. 22, 2021).

[34] C. Wells, *Americans Still Don't Understand How Bitcoin Works*. www.bloomberg.com https://www.bloomberg.com/news/articles/2021-02-19/bitcoin-btc-and-cryptocurrencies-prices-surge-but-understanding-is-limited (accessed Nov. 22, 2021).

[35] J. Edwards, *Investopedia: Bitcoin's Price History*. www.investopedia.com https://www.investopedia.com/articles/forex/121815/bitcoins-price-history.asp. (accessed Oct. 29, 2021).

[36] *E*uromoney, What Is Blockchain?. Euromoney Inc. https://www.euromoney.com/learning/blockchain-explained/what-is-blockchain (accessed Oct. 29, 2021).

[37] S. Likens, *Bitcoin, Cryptocurrency, Blockchain... So What Does It All Mean?*. PWC.com. https://www.pwc.com/us/en/industries/financial-services/fintech/bitcoin-blockchain-cryptocurrency.html (accessed Oct. 29, 2021).

[38] I. A. Frincu, *How Do Crypto-Vaults Work?*. datadriveninvestor.com. https://medium.datadriveninvestor.com/how-do-crypto-vaults-work-c8609853ad90 (accessed Dec. 13, 2021).

Chapter 6

Decentralized Finance on Public Blockchains: Hype or Economic Revolution?

Dominik Kajinic, Manu Narayanan

Abstract *Decentralized Finance (DeFi) has been growing at a fast pace in the past year, reportedly reaching a market capitalization of $150 billion in April 2021 [1]. Seeing how DeFi has been steadily gaining relevance, our work tries to define it as opposed to Centralized Finance (CeFi) and assess its strengths, use-cases, weaknesses, and risks. In doing so, we try to inform the readers on the question: "DeFi on public blockchains (BC): hype or economic revolution?". We try to shed some light on how DeFi architecture is designed, important structural elements of DeFi such as smart contracts (SC), decentralized applications (DApps), decentralized exchanges (DEX), etc. We also look at the opportunities presented by DeFi and the various use-cases emerging out of these opportunities. As an innovative use-case that could only have been conceived with DeFi, we look at flash loans. On the flip side, we also analyze some of the challenges of DeFi and some threats and attacks that DeFi can fall victim to. We zoom in on these challenges and attacks by looking at some specific examples, namely, governance attacks, crypto-mixers, and money laundering. Finally, we conclude by making a broad sense of all the information presented and thinking about the direction that DeFi can take in the future.*

Contents

6.1 Introduction to Decentralized Finance

In this section, we will introduce Decentralized Finance (DeFi). Manly, we will focus on what DeFi is, what it stands for, how it works and in which context it gets used.

6.1.1 Definition and Terminology

Decentralized finance is a financial system that is based on blockchain (BC) technology through smart contracts (SC). It is therefore an alternative to the traditional financial system. In DeFi, the financial products are available on a public decentralized BC network. As a result of these propositions, users can interact directly peer-to-peer with each other without any intermediaries in between. The only thing we need to use the financial products offered by DeFi is a device with an internet connection, so it is open to everyone. DeFi also brings new financial inventions with it which were not possible to use in traditional finance. An example is the so-called "Flash Loan" in which you borrow tokens without any collateral as long as you can return them in the same transaction. For this, only a small transaction fee is paid. We will focus on Flash Loans in a later section.
DeFi has four main characteristics. Those are "efficiency", "transparency", "accessibility" and "composability". Efficiency in connection with DeFi means that the offered financial services are successfully executed correctly in an acceptable amount of time. So e.g. when tokens get transferred from one peer to another, the correct amount of tokens reach the correct peer in a short time. This efficiency gets mostly gained through the SC on which DeFi relies on.
The second characteristic of DeFi, which is "transparency", addresses the fact that the whole system is based on code. All have access to the code of the system and therefore everyone can inspect the rules and logic of the whole system which makes it easier to trust the system and makes it also a ground for a fair market.
Further, anyone who has access to a device with an internet connection has also access to the DeFi. Therefore Defi is theoretically accessible to everyone. No application is needed and the services can be used immediately.
Finally, DeFi is composable. On the one hand, this means that the DeFi tokens and apps are standardized. Therefore the tokens can be connected even if they come from different platforms and apps which does not isolate the apps/tokens from each other but it makes them able to connect with each other, which is crucial for the whole system to work. On the other hand, composability means also atomicity. This means that the connection between the apps/tokens has to be done in one single transaction. Without composability, a DeFi platform would only be able to run apps/tokens isolated from each other so the benefits of the system could not be used.

6.1.2 Blockchain and Smart Contracts

As mentioned before Defi is based on BC technology through SC. To get a better understanding we will first now introduce you to the two topics of BC and SC.
"Blockchains are distributed ledgers that enable peers to transact without the need to entrust third-party intermediaries" [2]. There exist two categories of BCs: permissionless BCs and permissioned BCs. In permissionless BCs, users can enter and leave the system without any permission. On the other hand, permissioned BCs consist of a group of authenticated participants who are able to use the system [2].
DeFi is based on permissionless BCs. Therefore the users of DeFi can enter and leave the system without any permission of other parties. Further, transactions that are documented on BCs have three main characteristics. Firstly, they are valid, as they are executed correctly, secondly, they are immutable, as they cannot be modified or changed

after successful execution. Lastly, they are verifiable as they are public and therefore visible for everyone, which makes them easy to check [3]. BCs are also protected through advanced cryptography. Those aspects enable trust in the system of BCs and therefore also in the DeFi.
The characteristics of the BC technology, therefore, enable a new form of finance to get different parties to interact with each other. The different parties can make the transaction directly peer-to-peer between each other. Consequently, intermediaries are no more used. Since the BC technology is decentralized and does not need any intermediaries, costs connected with search, contracting, and enforcement can be reduced. This is a result of the expanding transaction possibilities of the innovative system, where the peers get connected directly to peers in new, different ways [3].
"Some blockchains, for example, Ethereum, offer generic computation capabilities through smart contracts" [2]. SC are programs written in form of code. Those programs consist of pre-specified conditions and as soon all conditions are met, the code gets executed automatically [3]. Therefore, SC are executed without user interference [7]. So basically, SC have the logic of a real contract implemented and as soon as all the conditions are met, they automatically get executed.
The new thing about SC is that e.g. when a user interacts with normal server-based web applications a regular user cannot access the code and the whole logic of the application. In a second point, the user cannot control the execution environment of the application. These two points lead that the user has to trust the application service provider because either one or both of these problems can be insecure or manipulated by the service provider [2]. SC on the other hand address both problems. The code is on the BC that the SC are based on, which is public, and therefore the logic can be inspected by the users. SC also ensure that the applications work as they are expected to. The behavior of the contracts is deterministic, which means that no randomness is involved in the process and the system will give every time the same output from the initial point. The function calls of the SC (e.g. transactions) are handled simultaneously which ensures the validity of the execution [4].

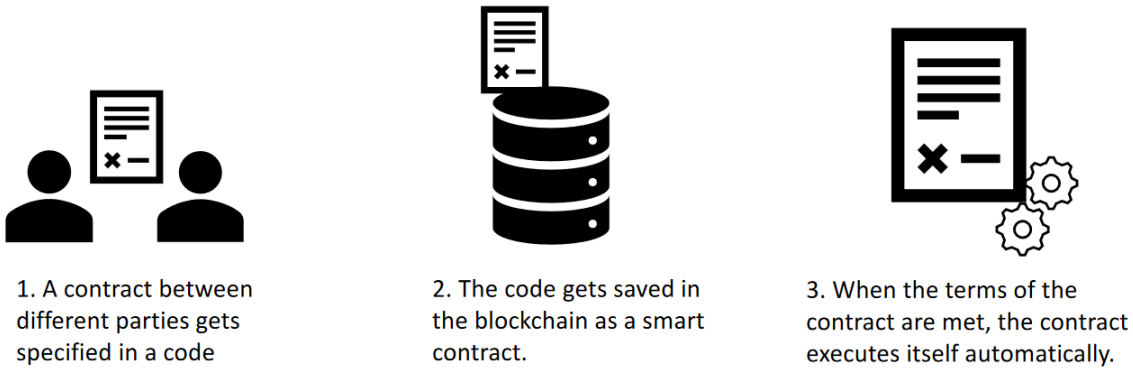

Figure 6.1: Smart Contract example

SC also can store crypto-assets which lets them work as a custodian as they can be flexibly modified when, how, and to whom the assets should be released [2]. This enables the user to be the custodian of its own assets and the user gains full control over his assets.

Due to the characteristics of SC - which are: transparency, immutability, automaticity, and programmability - the complexity and cost of contracting can be reduced [3]. To make an example of the characteristics: The contracts get executed automatically, which reduces time, they can't be changed when they get executed so they are immutable, the logic of the contracts is transparent for all, so everyone can see the code behind and the contracts and they are flexibly modified before the execution which is an example of the programmability.

6.2 From Traditional to Decentralized Finance

After the introduction to DeFi, in this section, we explain Traditional Finance also called Centralized Finance (CeFi), and make a comparison between DeFi and CeFi. This comparison shows the benefits but also disadvantages of one to the other financial system. Further, in this section, we will make some examples of the use cases of the DeFi.

6.2.1 Traditional/Centralized Finance

Traditional/centralized finance (CeFi) is the form of finance that we have already used for several thousand years. "Since then, humans have used a wide range of goods and assets as currency (such as cattle, land, or cowrie shells), precious metals (such as gold, which have enjoyed near-universal global cultural acceptance as a store of value), and, more recently, fiat currencies" [5]. Either the currency gains its value intrinsically (e.g. land) or the humans have given it a value (like fiat currency). That is basically how modern finance works, the currency gets the value that people believe it is worth, through supply and demand and also other factors like the stability of the country in which the currency is used. So the stability of a currency and financial system are made upon a centralized entity e.g. a government that protects the value of the currency and the financial system with the military at its command. [5]
In CeFi the user does not exactly hold its assets on their own. They get controlled by a different party, a financial institution e.g. a Bank. The user, therefore, executes transactions with the help of an intermediary. So the financial market gets created through three parties. The seller, buyer, and the intermediaries, which connect the buyer with the seller and execute the transactions for them.

6.2.2 DeFi and CeFi - Classification

Now that both financial systems (DeFi and CeFi) got introduced and we got a brief understanding of both systems, we make a comparison between the two systems, which clarifies first when a system is CeFi, when it is a hybrid form of CeFi and DeFi, and finally when something is completely DeFi. Secondly, it points out the advantages and disadvantages of both systems.
If we look at Table 2.1, we can see the clarification when something is CeFi, DeFi, and a hybrid form of both. We can see that if the financial assets are not controlled by the user but by a third party we are in a fully centralized financial system. If the assets are controlled by the user and the transaction execution and/or the protocol execution can get censored single-handedly (e.g. by the user or an intermediary) then we speak of a mixed form of CeFi and DeFi. Lastly, if the financial assets are controlled by the user but neither the transaction execution nor the protocol execution can get censored by someone, so they happen automatically, we speak of a fully decentralized financial system. So to be fully decentralized means that there is no need for human interaction during the execution of the transaction.

Table 6.1: CeFi vs DeFi, based on [5]

	CeFi	CeFi Intermediary, DeFi Settlement	Centrally governed DeFi	DeFi
Financial Assets are controlled by the user (non custodial)	No	Yes	Yes	Yes
The transaction execution can single handedly get censored by someone	Yes	Yes	No	No
The protocol execution can single handedly get censored by someone	Yes	Yes	Yes	No

So one difference between CeFi and DeFi is the use of intermediaries. While CeFi uses them DeFi does not. Both options have their advantages and disadvantages. In DeFi the user has full control and a clear overview of their assets. In CeFi the user lets the intermediary control their assets. The user gives input of an idea of how the assets should be invested and the intermediary tries to reach this wished goal. This can be a good aspect if the user does not have time to deal with the assets and lets an "expert" do the work for him. For this, the user has to trust the intermediary. The user can anytime give input and knows whom to make responsible for errors either of transactions or the system. Finally, the intermediary gets a fee for his work from the user.
In DeFi on the other hand the user does not have to pay those fees as he controls his assets by himself. But he is also responsible himself if he executes the wrong transactions. Further, there is no help if the whole system of decentralized finance crashes and the user has all his savings in this financial system. The user will have to deal with the consequences by himself. So there is in a way a trade-off between controllability and security.

6.3 DeFi Opportunities and characteristics

Now that we have a clear idea about what counts as DeFi and how much, we look at some properties which are characteristic of DeFi. Most of the novel opportunities and use-cases from DeFi can be traced back to one or more of these properties. They are:

- **Efficiency:** As DeFi replaces the traditional central intermediaries such as custodians and escrow agents with SC, it significantly reduces counterparty risk and thus makes financial transactions more efficient. It is also expected that due to lower trust requirements, there will be a lesser need for regulatory pressure and third-party audits. Similarly, efficiency gains are possible at every juncture of financial infrastructure with DeFi.

- **Transparency:** All transactions on DeFi are publicly observable and the SC code can be analyzed on the BC [4] . Because of this abundant data, it is expected that there will be more scholarship on DeFi, and we will be better positioned to mitigate undesirable events or even prevent them before they emerge.

- **Accessibility:** DeFi can be used by anyone with a moderate internet connection. Infrastructure requirements are relatively low and the risk of discrimination is practically non-existent in DeFi [4]. This high level of accessibility leads to many opportunities in DeFi for people who are particularly vulnerable to some form of

censorship or inaccessibility to CeFi. Moreover, anyone with some know-how can also create and deploy DeFi products or participate in the operation of DeFi applications. Due to this, it is expected that DeFi may produce a genuinely open and accessible financial system [4].

- **Composability:** DeFi protocols are often described as "money legos" due to the atomic composability of DeFi. The shared settlement layer allows different protocols and applications to interconnect, fork or rehash, and create something entirely new [4]. This composability is one of the most important and unique properties of DeFi, as it makes many radical financial products possible and worthwhile.

6.3.1 DeFi Opportunities

DeFi presents numerous opportunities in the financial system, some of which already exist in CeFi, whereas some of which have no analogues in history. These opportunities and novel use cases have been made possible by some unique properties, which have been discussed above, that are prevalent in DeFi. We look at the main opportunities in DeFi below:

- Non-stop market hours: As DeFi does not have to depend on an intermediary, such as stock exchanges, which may have strict operational hours, it can operate 24x7. As a result, there is no "pre-" or "post-market" trading in DeFi. Furthermore, there is also less likelihood of system outages that happens to CeFi marketplaces during times of high volatility.

- Novel financial instruments and mechanisms such as flash loans, autonomous liquidity pools, atomic swaps, decentralized stablecoins, etc. have been made possible in DeFi.

- Risk-free rate of return is possible if risks from underlying SC and BC consensus can be resolved.

- Borderlessness and interoperability of DeFi leads to the high accessibility of DeFi. DeFi is being speculated as being the first entry to the financial system in the future for the un-banked population in under-developed countries and countries affected by authoritarian control and government censorship.

6.3.2 DeFi Use-Cases

Now let us look at some of the use-cases of DeFi. Although there are many, we specifically look at three of them below:

i Decentralized Exchanges (DEX)

Decentralized exchanges are basically marketplaces where DeFi assets are traded. Unlike a CeFi exchange, a DEX usually achieves its exchange governance in a decentralized manner, i.e., through SC. They are more cost-efficient compared to centralized exchanges and are globally accessible. Some prominent examples of DEX's are Uniswap, Binance DEX, Sushiswap, etc. DEX's exclusively trades between cryptocurrency tokens and prices various cryptocurrencies against each other algorithmically. They use liquidity pools for trading, in which investors lock in their funds for interest-like rewards.
As an example, let us look at the case of Uniswap. Uniswap works with a design called Constant Product Market Maker, which is a variant of Automated Market Maker (AMM). In this design, liquidity providers create a market by depositing an equivalent value of two

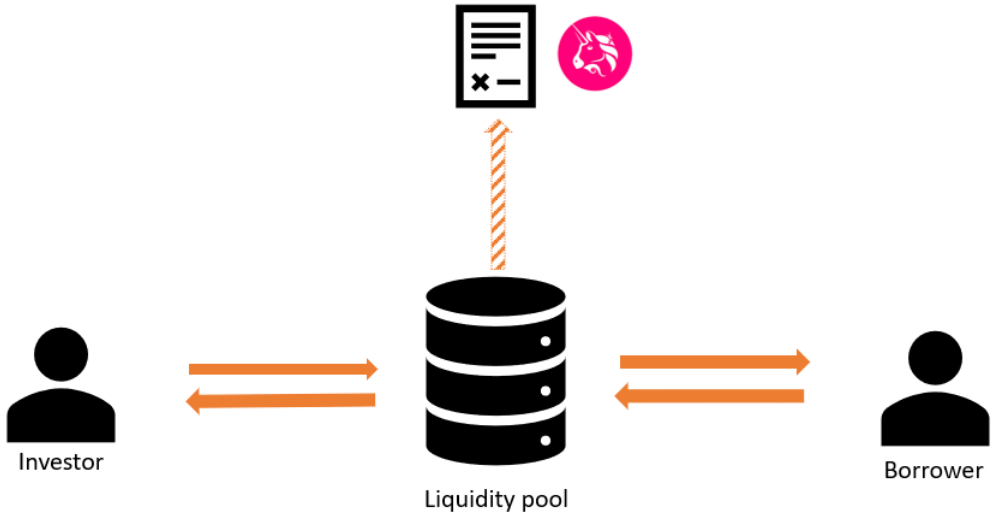

Figure 6.2: DEX example

tokens. In return, they get "liquidity tokens", which represent their share of the liquidity pool. Now, according to this design, the total liquidity in the pool is kept constant by keeping the product of the total quantities of the tokens constant. Hence, when the liquidity shifts away from a token, it becomes more expensive to take more of that token out. Thus, the liquidity pool maintains a balance between the tokens.
As mentioned above, DEX's exclusively trade between cryptocurrencies. However, these may include stablecoins which represent fiat currencies. This connects the traded assets back to CeFi, as the prices of these stablecoins would be affected by trades, regulations, and other events in CeFi.

ii Decentralized Applications(DApps)

DApps in DeFi are applications whose backend runs on BC networks, usually via SC. They can also have a frontend and user interface to make calls to its backend. As they have all the properties of DeFi such as accessibility, zero-downtime, transparency, and composability, they have been attracting significant investor interest. Although the Bitcoin network supports dApps to some extent, the vast majority of DApps are built on Ethereum. DEX's such as Uniswap, lending and borrowing protocols such as AAVE, and prediction markets such as Polymarket are all examples of DApps. Thus, we can see that the opportunities in DApps are endless and they can make use of all the properties of DeFi.

iii Flash Loans

Flash loans are a type of uncollateralized lending that has become very popular in DeFi. A flash loan is initiated and repaid within a single atomic BC transaction [6]. If the borrower cannot repay the flash loan by the end of the transaction, the on-chain state remains unchanged. Therefore, this provides a novel opportunity for DeFi participants to access liquidity without holding assets upfront. Flash loans are widely used in arbitrages and liquidations, however, they can also facilitate DeFi attacks. Although they are not the root vulnerability of these attacks, they do give adversaries access to vast amounts of liquidity which can be used for attacks.

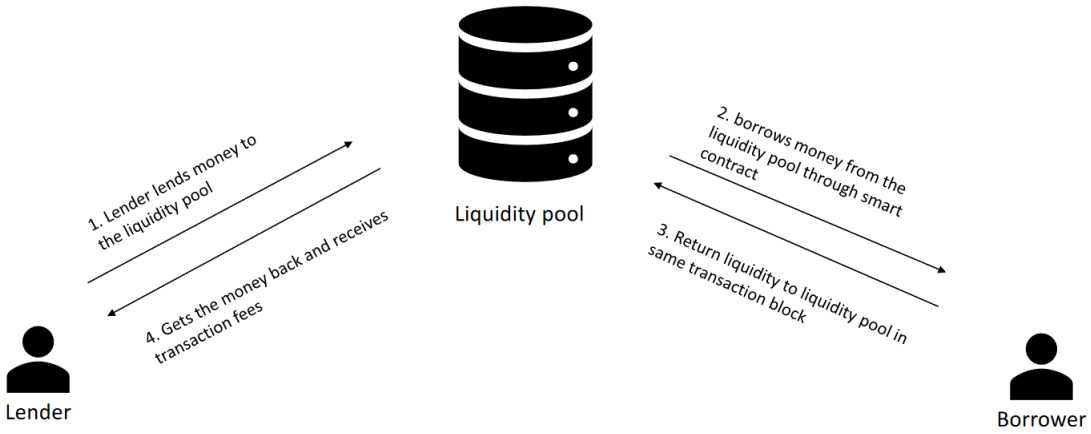

Figure 6.3: Flash Loan example

iv DeFi Stablecoins

Stablecoins are crypto-assets that fundamentally try to eliminate some of the price volatility that is common in fully decentralized cryptocurrencies. This is done by pegging their price to a fiat currency like the US dollar, which is less volatile than most cryptocurrencies [5]. This can be done in multiple ways.

One method to create a stablecoin is to back its value by collateralizing against a stable asset, such as a fiat currency. Such a mechanism usually "requires a centralized and trusted authority to manage the collateralized assets" [5]. This authority mints the stablecoins, whereas a user of the stablecoin can "burn" it at any point in exchange for the collateral at the pegged price. So, when the stablecoin price declines below the pegged price, arbitrageurs are incentivized to purchase them. On the other hand, when its price goes up, more stablecoins are minted, thereby increasing its supply and depreciating its price. USDT and USDC are examples of this type of stablecoins, and these have been shown to be more robust than other types of stablecoins.

A second method of creating stablecoins is through leveraged loans, where the value of the stablecoin is collateralized by other cryptocurrencies [5]. DAI, from MakerDAO, is the most prominent example of this type of stablecoin. These stablecoins are usually over-collateralized (1.5x over-collateralized in the case of DAI) with the underlying cryptocurrencies. Compared to the first method, this is less capital efficient due to over-collateralization. Although the price of this type of stablecoins is less robust than ones with pegged assets, they are still shown to be reasonably stable with time.

The third type of stablecoins uses autonomous algorithmic supply adjustments to keep its price stable [5]. In this method, an adjustment algorithm encoded within a SC adjusts the supply of the stable coin to drive its price towards the desired target. So far, this type of stablecoins has been seen to suffer from high price fluctuations. Examples of this type of stablecoins are AMPL and ESD.

To understand the effectiveness of the price stability of different types of stablecoins, we have included a visualization of the price trends of the USD stablecoin examples mentioned above. This has also been taken from [5], which the researchers extracted by crawling the price data from `https://www.coingecko.com/`.

To summarize, all these innovative use-cases and services arising from DeFi create a vibrant, interconnected ecosystem, that involves various kinds of DeFi assets and trades

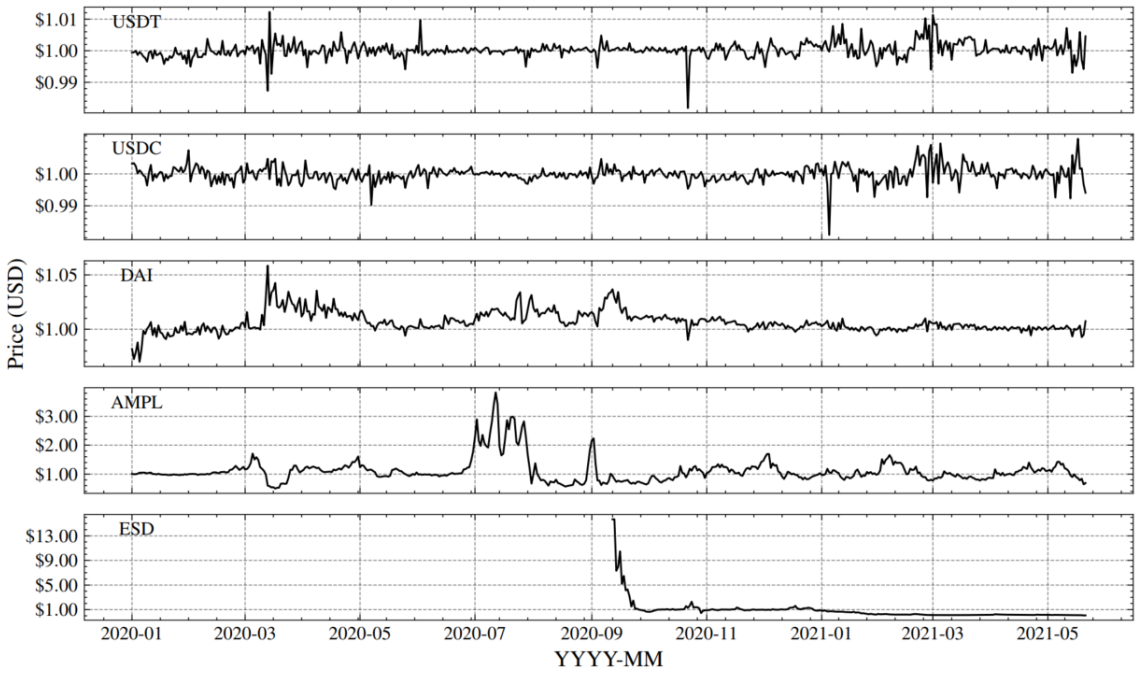

Figure 6.4: Prices of USD stablecoins [5]

such as DEX's, price oracles, crypto wallets, stablecoins, etc., which in turn would also have links to the CeFi system. To help visualize where these entities are located and how they are connected to each other, we have included a schematic diagram from [5] below.

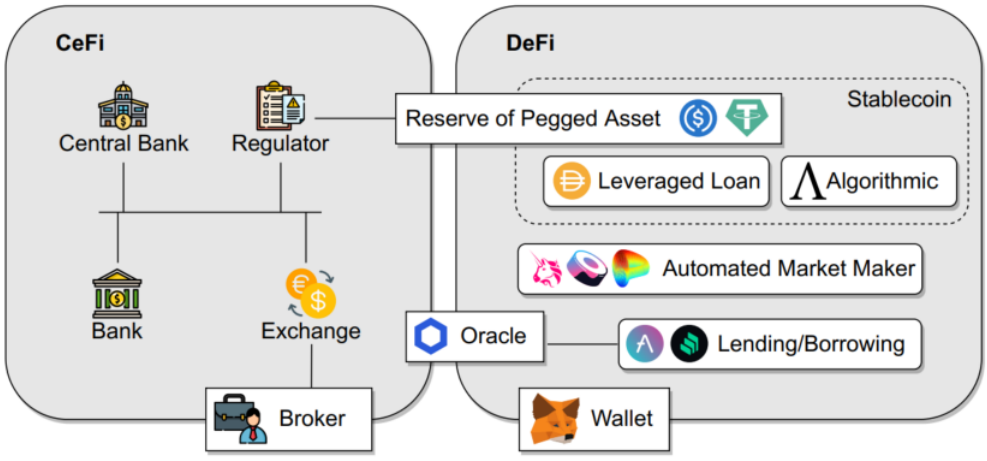

Figure 6.5: High-level service architecture of CeFi and DeFi [5]

6.4 Challenges and Security Issues in DeFi

In this section, we focus on the challenges and security issues of DeFi. We first look at the challenges and limitations of DeFi and if there are ways to resolve them. Then we look at the vulnerabilities in DeFi due to these limitations, and the various attacks that can be carried out on DeFi due to these vulnerabilities. We also look at two case studies on DeFi attacks to better understand the modality and severity of such an attack.
Firstly, let us look at some of the challenges and limitations of DeFi:

i Dependencies

This limitation is a flip-side of the composability feature of DeFi. According to [4], interactions between SC in DeFi introduce new risks by which, an issue with one SC can "potentially have wide-ranging consequences for multiple applications across the entire DeFi ecosystem". For example, we can easily imagine this when we think of liquidity pools with several cryptocurrencies and the wild fluctuations that we frequently see in the prices of these cryptocurrencies. As the DeFi ecosystem gets more and more intertwined, this limitation will become more and more important, and a shock to one part of the ecosystem can easily disrupt the whole ecosystem. Add to this the possibility of a bug in one of the SC, we can see this becoming a significant risk for the system.

ii External data

Another limitation of DeFi is that some SC rely on external data. A lot of DeFi lending services, derivatives, and some DEX's use price oracles that access external data about prices of various instruments which are not natively available on-chain. This reliance on external data and use of oracles again introduce dependencies and may lead to heavily centralized contract execution [4]. Some projects mitigate this risk by relying on decentralized oracle networks with a wide variety of data provision schemes.

iii Illicit activity

Another common concern with DeFi is its use by individuals for illicit activity. While DeFi is inherently transparent, bad actors can still make use of its pseudonymity and some creative workarounds to escape regulations. For example, mixing services such as CryptoMixer, allows a user to shuffle their coins with the coins of other users [5], which can be used for money laundering or eliminating traceability to other illicit activities. A visual depiction of the mechanism in a mixer service has been included on the following page to understand how this works. Moreover, some mixer services have also started rewarding users for participation, which makes this issue even more severe. This issue is very complicated for regulators, as reasonable solutions would have to prevent such illicit activity without undermining the privacy attribute of DeFi.

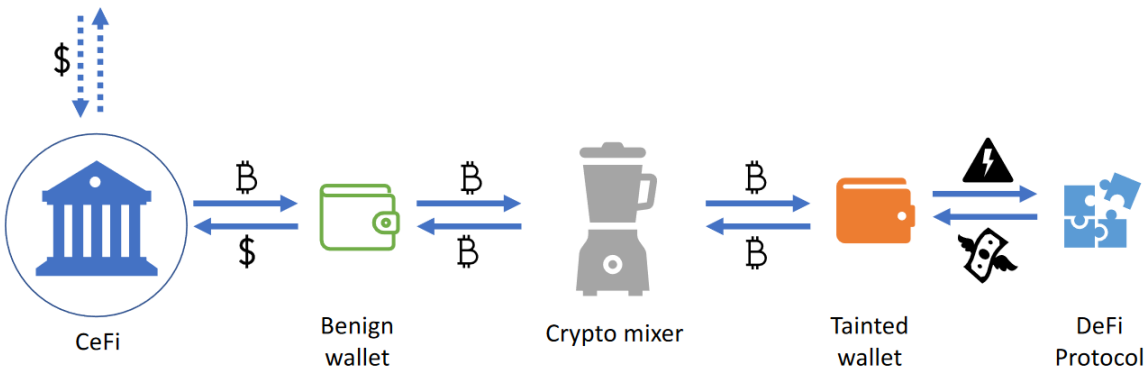

Figure 6.6: Illicit activity using mixers

iv SC execution

When it comes to SC, there is a trade-off between efficiency and security. In public BCs, the network is created such that every user can be involved in proofing the correctness of the execution of any operation. As a result of this SC are more inefficient compared to traditional centralized computing. [4] On the other hand they are very secure and enable trust since SC are transparent. Therefore they are not easy to manipulate. Another type of risk that can arise from SC is a coding error. Coding errors in the SC can potentially create vulnerabilities that can compromise the whole system. As the field develops, there are efficient solutions that are coming up which can prevent SC coding errors and bugs up to a certain extent

v Scalability

BCs face a complicated "trade-off between decentralization, security, and scalability" [4]. As networks grow and become more popular, the gas prices (transaction fees) and confirmation times also grow, which favor wealthy individuals and high-volume trades. They can be made more scalable by moving them to a more centralized base layer, which would compromise the decentralization aspect. On the other hand, other solutions such as base-layer sharding can improve scalability without affecting decentralization. This necessarily involves partitioning the entire computational and storage workload across the P2P network so that each node does not have to process the entire network's transactional load. However, this can lead to security issues over how to reach a consensus on whether the proposed transactions are authentic and can be added to the block. Thus, when a DeFi system grows, it will have to inevitably deal with this trade-off and figure out which attribute to compromise so that it can achieve its objectives.

vi Interoperability

Another aspect is interoperability. Interoperability is one of the main features in the BC technology, but full interoperability has not been reached yet in DeFi. There could be two solutions to fix this problem which could make the DeFi more useful and attractive. One way is to encourage the users to only one dominant platform and to create all the projects on this platform. This could increase interoperability as it is observable on the currently dominant platform Ethereum, which profits from very high interoperability. The downside of this is that the system gets less robust and a crash of this platform would become extremely costly. Furthermore, one single BC may not be able to serve all different needs [3]. Therefore, this way would result that in winners and losers. This would further lead to the losers again creating a new platform for their needs, which would again not be interoperable with the dominant platform.
Another way to address the problem of not full interoperability is to increase the interoperability between the different BCs. In this way, there could exist multiple different platforms and therefore projects could be built on multiple different BCs but still be fully operable between each other [3]. This would reduce the risk of one single platform crashing and the monopoly of one dominant platform and make the system more flexible. "Currently, many initiatives - such as Cosmos and Polkadot - are working on interconnecting different blockchains to achieve full interoperability" [3].

vii Frontrunning

Like CeFi, DeFi is also vulnerable to frontrunning. "Frontrunning generally exploits information asymmetries created by power structures within a financial structure" [7]. In CeFi this can happen when, for e.g., a broker has information about the user's account,

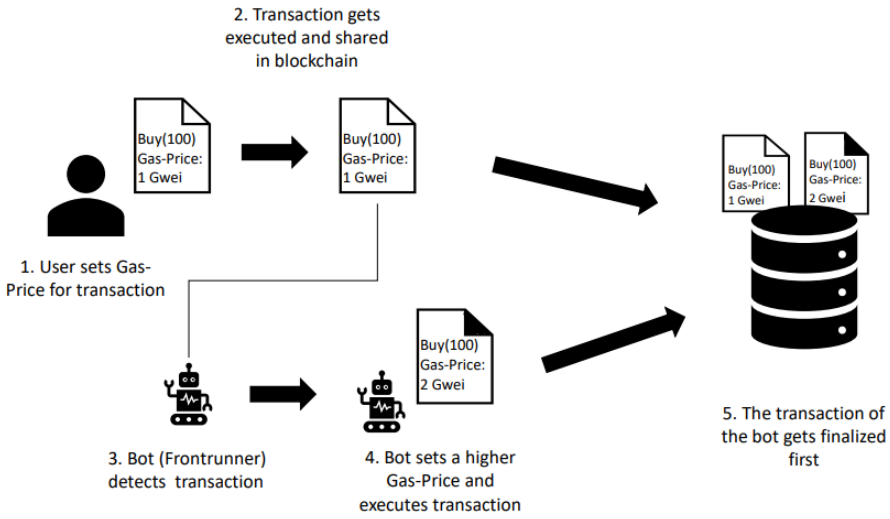

Figure 6.7: Frontrunning example

which is illegal. It can also happen because of public market information. This form of frontrunning is not illegal but it is crucial for a high-frequent trading economy [7]. In DeFi, it can happen through the block producers, the so-called "miners". It happens like this that there is a malicious node that observes a transaction. When executing a transaction the users have to pay a gas price. This fee can manipulate how quickly the miners will execute the transaction [8]. The transactions which have a higher gas price will be executed before transactions with a lower gas price due to limited space in the blocks, which is called a gas auction. The malicious node observes a transaction after it gets shared in the BC but before it gets finalized completely. Therefore the malicious node sets a slightly higher gas price for executing the same transaction and ensures that it wins the auction and that its own transaction gets executed before the observed transaction or order [8]. Therefore, also in DeFi the attacker profits from the information shared in the public BC and can use it for their own profit.
With the case of frontrunning, the trade-off of transparency and security becomes clearly visible. In one way the transparency of the system makes the system more trustable. This is because it is visible which transactions get executed and ensures a fair market. Through transparency also market manipulations can be made visible. On the flip side, it makes the system vulnerable to attacks like frontrunning. The attackers can use the information directly from the system and with programming effective bots profit from that information. This damages simultaneously the profit of the user which first set the order,

6.4.1 DeFi Vulnerabilities

DeFi is vulnerable to several attack vectors due to fundamental limitations at various levels. Based on the attack surfaces, these can be broadly classified as the network layer, consensus layer, financial institution, SC code, and DeFi protocol and composability attacks [9] [5].
Network layer attacks are carried out by undermining the BC P2P network of the DeFi instrument. Examples of this type of attack are eclipse attacks, DDoS (Distributed Denial-of-Service), MitM (Man-in-the-Middle), etc. Consensus layer attacks target the settlement layer of the DeFi instrument and thus break its integrity. Examples are double spending

and selfish mining. Financial institution attacks exploit the presence of centralized intermediaries of DeFi, such as wallet providers, BC API providers, mining pools, and oracles. These intermediaries may be affected by code vulnerabilities, unexpected failures, or even local laws and regulations, which can then affect the whole DeFi architecture that depends on them. Smart code attacks exploit vulnerabilities in the SC code such as integer overflow, reentrancy, timestamp dependencies, etc. Lastly, DeFi's atomic composability is a double-edged sword that can lead to creative and unknown attack vectors. Novel combinations of and interactions of SC and pools of capital create innovative use-cases such as flash loans but also create new vulnerabilities. Some examples of these types of attacks are the ones on "Value DeFi" which manipulated its price oracle, and on "PancakeBunny" which was a flash loan attack.
To better make sense of the mechanism and consequences of some of these attacks, we look at the following case study:

i Governance attack on DAI

This case study is a hypothetical attack mechanism, which was laid out in [10]. This attack makes use of vulnerabilities in the DeFi protocol of MakerDAO and also exploits flash loans to extract profit from MakerDAO's liquidity pool. MakerDAO is one of the largest DeFi projects by market share. They mint DAI, which is a USD stablecoin that works on over-collateralized leveraged loans of other cryptocurrencies, as mentioned in the previous section on stablecoins. MakerDAO manages its governance process through the MKR token, where participants have voting rights proportional to the amount of MKR tokens that they lock within the voting system. To put it briefly, participants elect an executive contract by staking with their MKR tokens, which define the governing rules of the system.
In the hypothetical threat model discussed in [10], if an adversary can amass enough MKR tokens, they can elect a malicious executive contract. The malicious contract can be written to transfer all the assets locked as collateral against DAI, and then mint more DAI tokens which will again be transferred to the adversary. These DAI tokens can also be then traded until their price crashes and the Maker system itself goes down. According to the literature cited, the total amount of MKR staked against a new executive contract, can at times go down to below 50,000. To get an understanding of the trends of MKR tokens staked against various executive contracts, we have taken a visual depiction of it that we took from [10].
Now, there are many ways by which an adversary can amass enough MKR tokens so that the entire attack is as profitable as possible. The researchers consider two strategies: crowdfunding and flash loans. In the crowdfunding strategy, a group of adversaries can pool their tokens and stake the required amount on a malicious executive contract. They can collaborate trustlessly, by doing this using a contract, which then stakes all its funds on the malicious executive contract once a required amount is reached. using such a contract would also ensure that the participants will be compensated for the attack without. However, such a crowdfunding attack requires high coordination among the participants and is likely to alert benevolent MKR members.
The second strategy involves utilizing a flash loan to provide the required liquidity for the adversary to fund the MKR tokens. An advantage of a flash loan attack is that no collateral is needed to obtain this liquidity, and if the execution of the attack fails at some point, the adversary suffers no loss as the whole transaction will be canceled. This type of attack also significantly reduces the capital lock-up cost for the adversary.
The researchers predict in [10] that such an attack would require around 378,940 ETH to acquire the required amount of MKR tokens. They arrived at this cost by taking into account the rise in the price of MKR tokens due to the rise in demand for MKR tokens

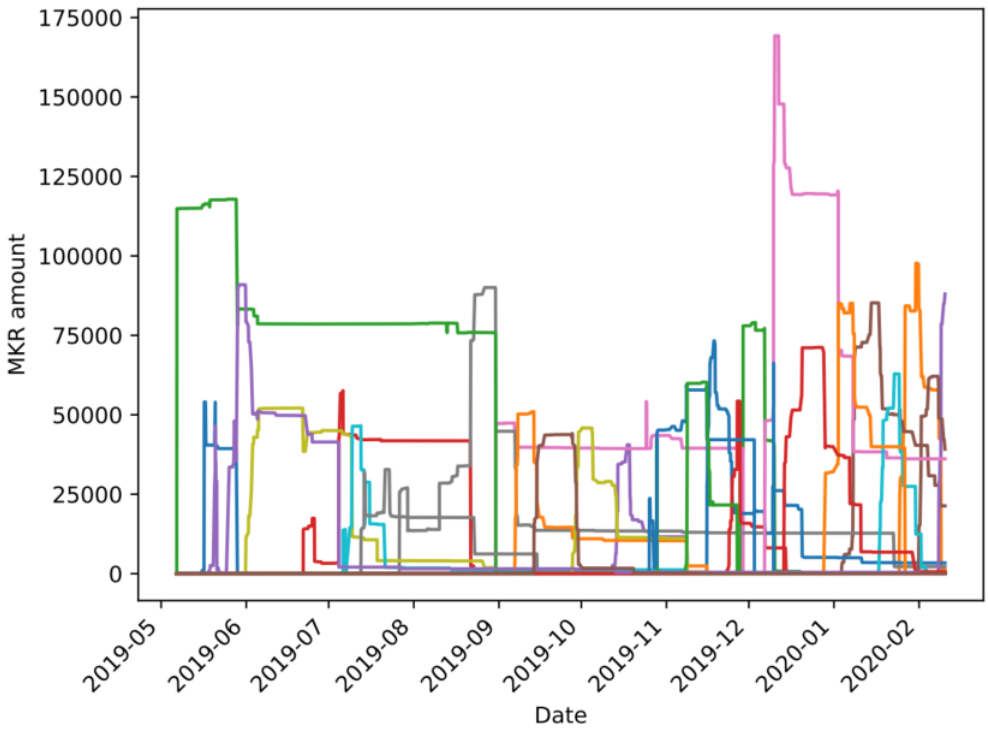

Figure 6.8: Evolution of the amount of MKR staked on different executive contracts [10]

during such large purchases. However, at the end of the attack, the attacker would have access to around 55k ETH, 50k MKR, and 145m DAI after repaying the flash loans with interest. Even after excluding the gas fees for the transactions, we can see that this would be a highly profitable attack for an adversary, not to mention the shock to the DeFi system due to one of its market systems going under.
There are also several defense mechanisms that can effectively mitigate such a governance attack. Firstly, the *Governance Security Module* encapsulates the successfully elected contract for a certain period, after which it takes effect. Increasing this period is an effective defense mechanism against such attacks. Secondly, an *Emergency Shut Down* provision can allow a set of participants holding a sufficient amount of MKR to halt the system.

6.4.2 Market Manipulation in DeFi

Other than the above-mentioned DeFi attacks, DeFi is also frequently subject to various kinds of market manipulation. DeFi Market manipulation, unlike a malicious attack on a vulnerability in the DeFi architecture, involves the act of intentionally manipulating the price of DeFi financial instruments. It undermines market fairness and can erode the trust in the DeFi market. Market manipulation can be broadly classified into exchange-based manipulation and external manipulation.
Exchange-based manipulation involves an exchange and can be further classified into trade-based and order-based manipulation. Trade-based manipulation involves buying and selling financial instruments in a predetermined manner. Insider trading and front running are examples of trade-based manipulation, where the manipulator uses unethically obtained information to then trade on the exchange and make money. Order-based manipulation mainly involves canceling placed orders before execution, as in the case of spoofing or quote stuffing.
External manipulation is carried out without the involvement of exchange and it can be further classified into action-based and information-based manipulation. In action-based

manipulation, the manipulator usually drums up the value of an instrument without any trading activities. Examples of this kind of manipulation are Ponzi schemes and honeypots. Information-based manipulation is carried out by disseminating false information or rumors about a DeFi instrument or its price and thus manipulating its price.
As external manipulation is not reliant on the technical details of the underlying financial system, it follows the same ways in DeFi as in CeFi. However, exchange-based manipulation in DeFi can differ from CeFi in terms of its methods and scale, as it relies on the underlying financial system. As DeFi heavily facilitates high-frequency trading, it is highly vulnerable to exchange-based manipulation.
To understand how market manipulation works in DeFi, and how it differs from market manipulation in CeFi, we look at one brief example which is described below.

i Frontrunning with Priority Gas Auctions (PGAs)

Despite all the benefits of DeFi, on-chain, SC-mediated trades are slow. Combine this with the inherent transparency of DeFi, adversaries can front-run orders on DeX's, observing the pending orders and placing their own orders with higher fees to ensure they are mined first [7]. In the literature, the researchers observe a community of arbitrage bots, which competitively bid up transaction fees to obtain priority ordering. This is called Priority Gas Auctions (PAGs).
Frontrunning is also prevalent in traditional CeFi exchanges and it is often illegal. In CeFi, actors with privileged access to user information, such as brokers, are the ones who exploit this information asymmetry to front-run orders and make a profit. In contrast, DeFi does not have any single party with privileged information. However, automated market bots from high-frequency trading firms can still obtain privileged information due to the superior resources available to them. They can exploit information asymmetry about price discrepancies across exchanges trading the same or correlated assets and then make a profit out of it.

6.5 Summary, Conclusion and Future Research Directions

In this section, we make a summary of our research and conclude our findings. Further, we will point some topics which from our perspective are interesting for future research and therefore for the future of DeFi.
In this paper we focused on the topic of DeFi. First, we made an introduction that clarified DeFi and explained how it works and therefore that it relies on the BC technology through SC. We continued with the definition of the most used financial system the CeFi and made a comparison between those two financial systems. This comparison clarified the advantages and disadvantages of each financial system which lead further to the opportunities that DeFi has. In the next step, we explained some use cases of DeFi. This made clear that DeFi does not only offers the financial services and products that are already used in CeFi, but it offers also new services and products (e.g. Flash Loans) which are not possible to have in CeFi. We pointed out that this is because of the innovative technology of the DeFi. Finally, we also pointed out the challenges that DeFi has and also the vulnerabilities of DeFi. This made clear that DeFi can and has to be improved further to make it less attractive for attacks and therefore to be a more secure system.
We conclude that DeFi has a big potential to grow and therefore to expand to an even bigger and more used way of finance. To use this potential DeFi has to be further improved. At this moment there are still some security issues that make DeFi vulnerable and attractive for attacks as the attacker gets away with no or small consequences. On the other hand, those vulnerabilities make DeFi attractive for innovators and entrepreneurs to

develop new products addressing those vulnerabilities and therefore to make DeFi more secure. DeFi is in general attractive for innovators and entrepreneurs due to the new possibilities of the DeFi technology. Using these possibilities new products (e.g. Flash Loans) which do not exist in CeFi can further be developed which then later can make DeFi more and more attractive for new users. DeFi is now still in the beginning and it will be interesting to observe the development of DeFi in the future if it will get more robust, secure, and through that also more popular, or if it will stop growing. Finally, the question of responsibility rises, who can be made responsible if DeFi crashes and what will be the consequences from it.

Bibliography

[1] CoinGecko; `https://www.coingecko.com/buzz/q2-2021-quarterly-cryptocurrency-report`, November, 2021

[2] Qin, K., Zhou, L., Gamito, P., Jovanovic, P., & Gervais, A.: *An Empirical Study of DeFi Liquidations: Incentives, Risks, and Instabilities*; June, 2021, arXiv preprint arXiv:2106.06389.

[3] Chen, Y., & Bellavitis, C.: *Blockchain disruption and decentralized finance: The rise of decentralized business models*; Journal of Business Venturing Insights, 13, 2020, e00151, doi: 10.1016/j.jbvi.2019.e00151.

[4] Schär, F.: *Decentralized Finance: On Blockchain- and Smart Contract-Based Financial Markets*; Federal Reserve Bank of St. Louis Review, Second Quarter 2021, 103(2), 153-74. doi: 10.20955/r.103.153-74.

[5] Qin, K., Zhou, L., Afonin, Y., Lazzaretti, L., & Gervais, A.: *CeFi vs. DeFi - Comparing Centralized to Decentralized Finance*; June, 2021, arXiv preprint arXiv:2106.08157.

[6] Qin, K., Zhou, L., Livshits, B., & Gervais, A.: *Attacking the DeFi ecosystem with flash loans for fun and profit*; March, 2021, arXiv preprint arXiv:2003.03810.

[7] Daian, P., Goldfeder, S., Kell, T., Li, Y., Zhao, X., Bentov, I., Breidenbach, L., Juels, A.: *Flash Boys 2.0: Frontrunning in Decentralized Exchanges, Miner Extractable Value, and Consensus Instability*; 2020 IEEE Symposium on Security and Privacy (SP), May, 2020, 910-927, doi: 10.1109/SP40000.2020.00040.

[8] Eskandari, S., Moosavi, M., & Clark, J.: *SoK: Transparent Dishonesty: Front-Running Attacks on Blockchain.*; April, 2019, arXiv preprint arXiv:1902.05164.

[9] Saad, M., Spaulding, J., Njilla, L., Kamhoua, C., Shetty, S., Nyang, D., & Mohaisen, A.: *Exploring the Attack Surface of Blockchain: A Systematic Overview*; April, 2019, arXiv preprint arXiv:1904.03487.

[10] Gudgeon, L., Perez, D., Harz, D., Gervais, A., & Livshits, B.: *The Decentralized Financial Crisis*; June, 2020, arXiv preprint arXiv:2002.08099.

[11] Perez, D., Werner, S. M., Xu, J., & Livshits, B.: *Liquidations: DeFi on a Knife-edge*; April, 2021, arXiv preprint arXiv:2009.13235.

Chapter 7

An Overview on Decentralized Exchange Platforms: Flow of Funds in Different Order Types

Carlos Kirchdorfer, Nitharsan Yoganathan

While Centralized Finance (CeFi) and Centralized Exchanges (CEX) have lasted for a long time, a new exchange type emerged from the popularity of the blockchain architecture. This new type of Finance, called Decentralized Finance (DeFi), provides through its Decentralized Exchanges (DEX) new assets and services. In our work, we provide a high-level overview of different assets that are available on DEXs, the different services that are used and possible market manipulations. We find that smart contracts play a crucial role in DEXs, since they are the fundation of a working decentralized exchange. Furthermore, they are the basis of an Automated Market Maker which is the smart contract responsible for the exchange to work. We also find that DEXs allow trading with assets that did not exist in CEX and that DEXs are still vulnerable to various market manipulations. These finding were supported by our investigation of the two exchanges Uniswap and Binance.

Contents

7.1 Introduction

In recent years, Blockchain has raised in popularity in the world. The revolutionary components of removing the intermediaries and allowing the users to own their funds respectively transfer the assets without any centralized, trusted authorities, is a critical value proposition for an open finance network. At the moment, there are hundreds of assets, which are blockchain-based, and many more are added, hence there is a demand to have a platform, to exchange these assets [1]. With the rise of the second generation of blockchain like the Ethereum platform, the functionality of smart contracts is possible [2]. These smart contracts enable the transaction between two or multiple parties in a decentralized and secure manner.
This introduced decentralized exchanges in the ecosystem, which was filled before with only centralized exchanges [1]. These exchanges in turn are critical milestones for decentralized finance also called as DeFi, which is a blockchain-based financial infrastructure. With the mission of replicating existing financial services, to create an open platform and transparency for the users. DeFi does not rely on centralized, trusted authorities or any intermediaries due to the invention of decentralized exchanges [3].
This seminar thesis will give an overview of the different services of a decentralized exchange and how various market manipulations can affect these exchanges, to see the flow of funds in different order types. Afterward, the seminar topic will have a closer look at the implementation of Uniswap and Binance regarding the decentralized services and market manipulation to evaluate the different strengths and weakness' of a centralized and decentralized exchange.

7.1.1 Order-Book Exchanges

When trying to understand how Decentralized Exchanges (DEX) work, one should understand how Centralized Exchanges (CEX) work. One of the most famous models that emerged from Centralized Finance (CeFi) is the order-book model. An order-book is in essence an electronic list of buy and sell order which are mapped to a specific asset. These orders are typically structured around a price level. Order books are used in CEX and DEX and are used for various assets. A order-book mainly consists of three parts: buy orders, sell orders and the order history. The market price is the result of the last matched buy and sell orders. Therefore, a counterparty discovery mechanism is needed. This matching of buy and sell orders is nowadays performed by a Matching Engine. The Matching Engine calculates which orders can be fully or only partially done [35].
Matching engines usually support multiple order types. The three most famous order types are limit orders, market orders and stop orders (stop-loss) [36].

1. Market Order: This order is placed with the idea to be immediately executed with the current market price. Since the market price changes quickly when using an electronic Matching Engine, a market order will be executed on a different market price as to the time the order was specified. This happens because processing the transaction takes time. Nevertheless, the price difference should be relatively small [36].

2. Limit Order: In a limit order, one can specify the price on which the order should be executed. Therefore, the order will be executed as soon as the market price reaches the specified price [36].

3. Stop Order: When specifying a stop order, the asset will be automatically sold as soon as the market price reaches a predefined price level. When using this type of order, one can minimize the financial losses [36].

7.1.2 Asset Standards

In the cryptocurrency world do not exist globally approved definitions for the terms like cryptocurrency, crypto-asset, digital asset, coin or token. In practice, it is often observed that people use these terms in an interchangeable way. Nevertheless, there exist useful definitions for the terms coin and token which will be explained in the sections below [6].

i Digital Asset

The definition of digital assets is the concept of representing an asset in a unique and digital form [4]. Importantly, there is also value associated with digital assets. In the context of blockchain, digital assets include cryptocurrency and crypto tokens [5].

ii Native Coin

The term coin in the crypto context means that the digital asset is a standalone cryptocurrency that can operate on its own blockchain platform [6]. Examples are Bitcoin and Ethereum which both work on their own blockchain.

iii Token

In general, a token is a non-mineable unit of value in a digital form. They are entries in blockchains and exist in many kinds of forms, for example as currencies for ecosystems or to cryptographically encode data and therefore enabling unique identification (NFTs are an example for this). Furthermore, there are tokens that can be exchanged with real-world, off-chain assets [7].
Different to a native coin, a token on the other hand is a cryptocurrency that uses not its own, but rather a different blockchain to work [6]. An example for such as third-party blockchain used for tokens is the Ethereum blockchain where ERC-20 tokens are being issued through ICOs and then being traded on the secondary market. Therefore, in a tighter definition, tokens are not actual cryptocurrencies, but are exchangeable units of values that are traded and issued while using an already existing, native blockchain [7].

iv ERC-20 standard

To issue and program tokens on the Ethereum blockchain, ERC-20 as a technical standard was developed in November 2015 by Fabian Vogelsteller. In the ERC-20 standard, general rules are proposed so that a token can work correctly on the Ethereum blockchain. The ERC-20 standard should be considered as a technical specification and not as software code [8]. In its core, the standard facilitates and states the prediction of the interaction between applications and ERC-20 tokens more precisely. Furthermore, the standard defines how ERC-20 tokens are exchanged within the Ethereum ecosystem and how the consistent record management of supply and address balances of tokens should work [8].

v Fungible Tokens

Tokens can be issued on a decentralized or on a centralized exchange platform. Since centrally governed exchanges mostly will not list any of these tokens, trading these assets on the secondary market is often the only option [10].
In DEX, tokens need to interact with the protocols through which the exchanges work. Therefore, compatibility with the protocols is needed. To achieve this, most of the tokens are using the ERC-20 standard on the Ethereum (ETH) blockchain [10].

vi Non-Fungible Token (NFT)

A non-fungible token (NFT) is a cryptographic token that acts as a representator of a unique asset and is stored on a blockchain. NFTs may also be called crypto collectibles. With NFTs, one has the possibility to tokenize an already existing assets or one may implement a new digital asset from scratch. Each NFT is unique and NFTs in general are only available in a limited number. Importantly, one can only replicate or transfer a NFT with the consensus of the NFTs owner [12].
Fungibility is context of an asset means that its units can be interchanged with each other. Therefore, the units of the asset are indistinguishable, and each unit contains the same valid value [9]. Fungibility is needed for any currency that shall be used as an exchange unit. However, fungibility is not desirable for collectible items and is therefore not a trait of NFTs. NFTs are not fungible which means that the different NFT-units are not interchangeable with each other. Since fungibility is not available with NFTs, they can be used to validate authentication and ownership in the digital world. Furthermore, even partial ownership of items of different value can be implemented with NFTs [12].
Because of their innovative ownership and fungibility ideas, NFTs have already been used in a great deal of projects. Examples of the use of NFTs can be found in licensing, gaming, digital identity, and art [12].
NFTs find usage in decentralized applications (also called DApps) for issuing purposes, for example for issuing crypto collectibles or unique digital items. As already mentioned, NFTs can be used to tokenize real, already existing assets. Since NFTs are storable on a blockchain and tradable, the tokenized real-world asset can now be traded with the same properties as already existing tokens. The token economy might provide a new way of gaining liquidity to markets that so far struggled in gaining and maintaining liquidity. The area of digital identity may also profit from the attributes of NFTs. The concepts of unique identification and ownership within blockchain databases provide value since it benefits user privacy and data integrity [12].
When evaluating the value of the NFT, one might think that NFTs do not contain any value, since they are not physical entities made of a material, but only exist in digital form. Nevertheless, NFTs contain value to some subjects because the people who trade with NFTs have a mental model where they assign value to NFTs. This underlines the fact that value is created by people who share the same belief. An analogy for this is also the commonly used fiat money which also does not contain an inner value, but provides value to the people of a country, society, or group [12].

7.1.3 Smart Contract

Smart contracts build the foundation for decentralized exchanges. These are self-executing contracts where the terms of agreement of two or multiple sides of buyer and seller are written in a programming language. The characteristic of a smart contract is that the caused transaction is traceable, transparent, and irreversible [13].

i DeFi Stack

The reason, why smart contracts are integral for the entire DeFi ecosystem can be explained by the DeFi stack (see Figure 7.1): The first layer, which builds the foundation of the stack is called the base settlement layer (Layer 1). Ethereum can be considered as a base settlement layer for DeFi created on its network. Then the asset layer follows, which covers all assets, which are on top of the first layer. This will include the native protocol assets combined with its token issued (Layer 2). The third layer is called the protocol layer (Layer 3), which offers said DeFi services by using a set of smart contracts. Possible DeFi use cases are exchanges, lending, derivatives, stablecoins, asset management, and many

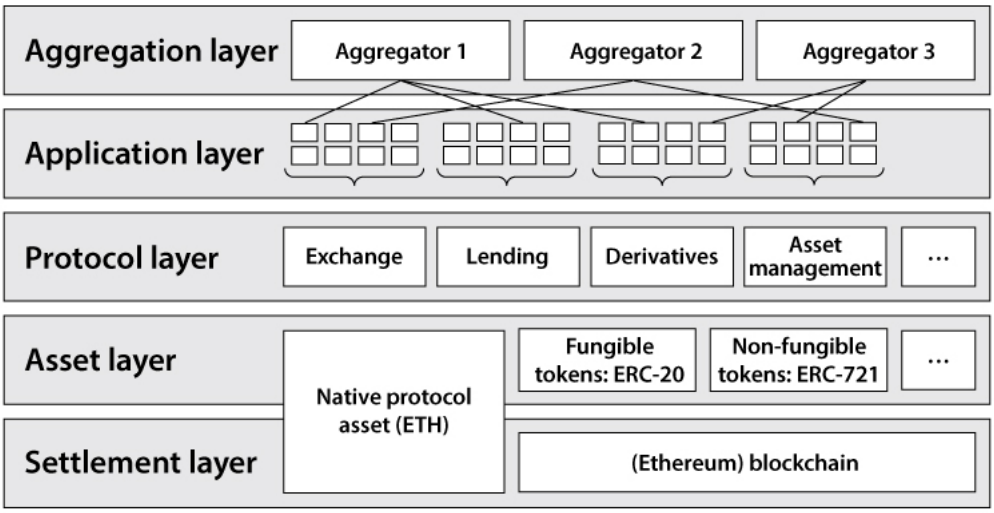

Figure 7.1: DeFi Stack [3]

more. This is possible due to the nature of protocols being highly interoperable, consequently, it is enabling a near frictionless financial market. The application layer (Layer 4) is bridging the gap between end-users and the protocols by providing a consumer-friendly application. Finally, the aggregation layer (Layer 5) contains aggregators. These user-centric platforms are connected to various applications and protocols such that it enables the users to perform a complex task by using several protocols simultaneously [3].

7.1.4 Decentralized Exchange

Decentralized exchange (DEX) is defined as the following: "An online P2P service termed as a decentralized exchange is the one by which direct transactions of cryptocurrency are allowed between two parties involved in the process [14]."
The advantages of DEX are that they don't require any ID verification or know your customer (KYC) restriction. Additionally, there is no counterparty risk involved which means that the users don't have to put their funds on the exchange itself respectively can store them in their own wallet. This will mitigate the risk of theft or loss of funds in DEXs due to hacker attacks. Besides that, DEXs are popular due to offering a higher variety of tradeable tokens compared to their counterpart, which only provides the most popular crypto assets [14].
However, DEXs bring also some disadvantages with them. On one hand, there is no recovery ability due to the nature of anonymity. This means a user can't retrieve its fund back if it's lost or ask for a refund of a transaction from support. Low liquidity is another weakness of DEXs compared to their centralized counterparts and liquidity is often an indicator, how attractive an exchange is for a trader. Additionally, DEXs are plagued by low transaction speed, since the validation process is dependent on the miners' network. Due to DEXs being new, it brings rather simple services and doesn't provide advanced trading functionalities like centralized exchanges. Lastly, DEXs suffer from scalability issues, there having often network congestions because most of them use Ethereum as a platform, which has not resolved the issue [14].

i On- and Off-chain

There are two types of DEXs: On-chain and off-chain. Some DEXs process all their orders on-chain. This is also the most decentralized form since it offers complete transparency and does not use any centralized entities. Unfortunately, this type is not practical in reality, since all phases of an order are processed on the blockchain, which causes low

transaction speed and high fees. This low transaction speed makes it vulnerable for certain market manipulation eg. front-running [14].
On the other hand, off-chain-based DEXs are more centralized than the previous mentioned type but still belong to decentralized exchanges. This type process the orders on a centralized entity that is responsible for the governance of the order book, such that it is exploitable for front-running or misinterpreting orders. But this type offers faster transaction speed, therefore has a lower chance of network congestions [14].

7.2 An Overview of DEX Services

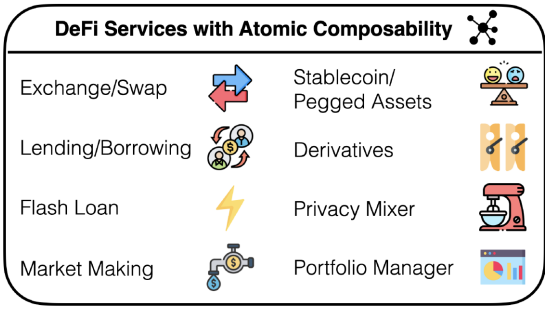

Figure 7.2: Services Provided By DeFi [15]

In the figure above, one can find an overview of services which DEXs provide. In the following, we will explain the portrayed services.

7.2.1 Exchange

An exchange is defined as a marketplace where financial assets are dealt. In former times, a financial exchange often happened on a real-world location where participants of the trade met to carry out the deal. Recently, trading has shifted from physical locations to the digital world. At the beginning of this shift, mostly centralized electronic exchanges have been used by traders to support their deals [15]. An electronic exchange consists per definition three main components. The first component is a mechanism to discover prices, the second one is an algorithmic matching engine for trades and the third component is a system to clear trades. Whether an electronic exchange can be classified as a decentralized one depends on how decentralized those three components are [15].
Furthermore, there are four dimensions across which exchanges can be decentralized. These dimensions are the mechanism for finding a counterparty, the blockchain platform, the order matching algorithm, and the transaction settlement [11].
In addition, exchanges can also be classified regarding the exchanged asset type. Usually, on decentralized exchanges with blockchain-based clearing systems one can only trade with cryptocurrencies (coins), tokens and stablecoins which represent existing fiat currencies [15].

i Financial Instrument Listing

In DEX, exchanges are handled in a decentralized way. Through this exchange architecture, the listing of assets works in a transparent way. A commonly used requirement for token listings in DEX is that the financial asset should fulfil the ERC-20 standard [15].

ii High-frequency Trading (HFT)

High-frequency Trading corresponds with the idea of having trading strategies which work in an automated way and have the goal to benefit from market volatility within a short period of time. Despite being different to CEX, DEX reveals that HFT strategies that work in CEX, also work in DEX. In summary, HFT strategies are very competitive. In the following example of two-point arbitrage, to be able to profit from this HFT strategy, a fast execution speed on both markets and exchanges results in a profit in expectation [15].
A suiting example is the concept of two-point arbitrage. In this financial concept, a trader buys a financial asset in one specific market, only to sell this asset at a higher price (and therefore making profit) in a different market. With this process, two-point arbitrage ends short-term value differences between different markets which results in a better efficiency of the market. There exist other HFT strategies than two-point arbitrage, but these strategies are relying on it. Because notifications in blockchain-based DEX spread on the public peer-to-peer network, there is no way of knowing which miner will be the one responsible for executing the transaction. To profit from arbitrage, one needs to minimize its latency to every miner and to the mining pools. Optimally, a transaction that tries to benefit from arbitrage executes atomically, so that the risks of price volatility are reduced [15].
When focusing on two decentralized exchanges on the same blockchain, an arbitrage can be seen as risk-free, when not looking at the transaction costs. This is because DEX traders can use the atomicity functionality provided by the blockchain to develop a smart contract that executes the arbitrage transaction. But if the transaction did not yield any profit, it would be reverted. Nevertheless, a trader still needs to pay the transaction costs if the transaction was reverted [15].
When analysing two DEX on different blockchains, the previously mentioned risk-free property does not hold anymore and the arbitrage is as risky as it is on a CEX or hybrid exchange. On CEX and hybrid exchanges, arbitrage is exposed to price fluctuations on the market, except the arbitrage performing traders collaborate with the exchanges to gain execution atomicity [15].

7.2.2 Liquidity Pools and Swaps

DeFi relies on liquidity pools which are controlled by smart contracts that allow trading a variety of different tokens. Users that provide tokens in such a liquidity pool get a liquidity provider (LP) token which rewards the user for sharing liquidity with the pool [15]. In general, liquidity providers provide proportionate amounts of two cryptoassets to determine the structure of a trading pair [17].
Within liquidity pools, traders can swap the tokens they like with the help of an Automated Market Maker (AMM). This AMM may incentivise swapping or liquidity provision on the DEX. Often, this happens through so-called protocol tokens which the traders can receive as a reward [18].

7.2.3 Lending and Borrowing

In CeFi, lending and borrowing are services that already exist for a long time. In DeFi, since there are no creditworthiness systems and default enforcement tools, the system needs over-collateralization in the lending and borrowing procedures. Over-collateralization refers to the idea of the borrower providing a larger amount of collateral (for his debts) than he is borrowing. Typically, lending and borrowing in DeFi occurs in lending pools. This pool is a smart contract that governs the assets that can be lent and borrowed and

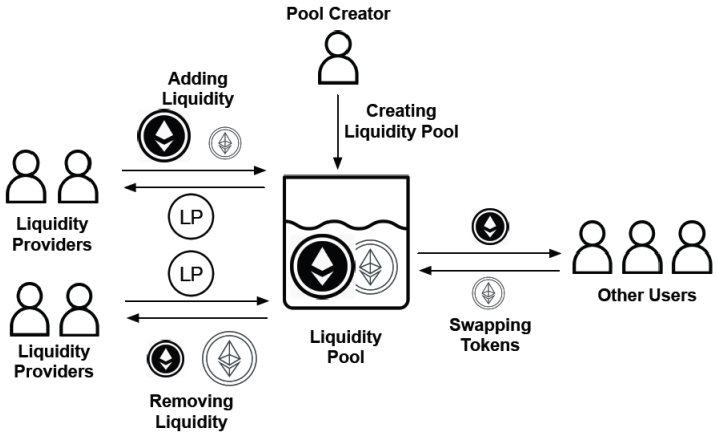

Figure 7.3: Overview Of Interacting With A Liquidity Pool [16]

that organises the way different actors collaborate. Normally, a lender deposits the cryptocurrencies which he wants to provide into the lending pool. The borrower then can collateralize into the pool before borrowing from it. While in CeFi the assets placed by the users in the lending pools are under the protection of CeFi regulations, there doesn't exist such a security in DeFi. To ensure that the loan is still over-collateralized while it is being used, lending pools often query the prices and exchange rates of cryptocurrencies from price oracles. As soon as the collateral is not sufficient anymore, the position will be secured with the method of liquidations [15].
The term liquidation describes the moment when a liquidator repays his debts and then receives its provided collateral assets. There are mainly two DeFi liquidations techniques. While the first one is the fixed spread liquidation, that can be executed in one blockchain transaction, the second one relies on auctions which need actions within several transactions [15].
In DeFi, users are allowed to borrow assets in an under-collateralized state, but this comes with some restrictions. The loan then is controlled by the lending pool and can only be used in a restricted way [15].

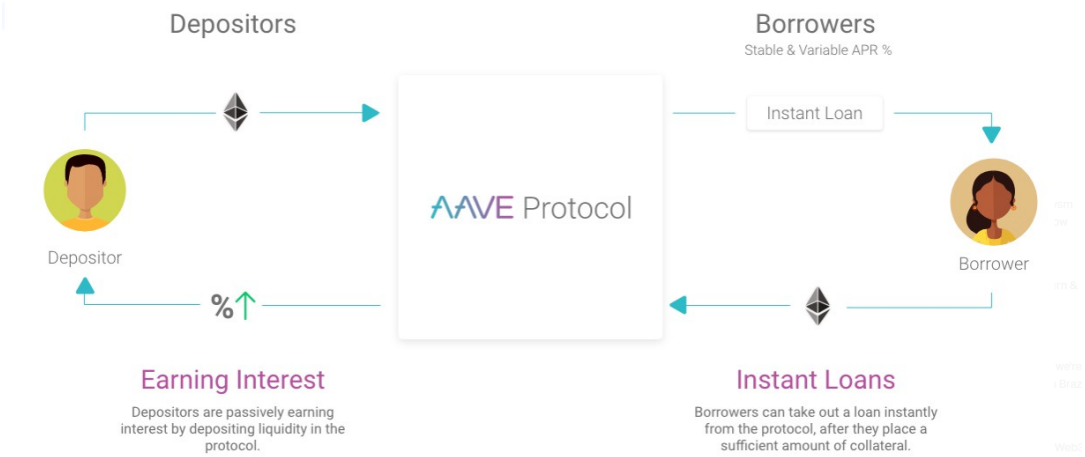

Figure 7.4: Lending And Borrowing Example With Aave [34]

7.2.4 Flash Loans

The concept of flash loans only exists in DeFi. A flash loan has the property that it is started and repaid within one atomic blockchain transaction. For achieving this, the borrower needs to execute three actions. First, he needs to request assets from a flash

loan lending pool. Then, the borrower may use the assets as he likes. Lastly, the borrower repays the loan with interest added to the lending pool [15].
With the transaction atomicity property, the on-chain state of the blockchain stays the same, if the borrower does not have the means to repay the flash loan by the termination point in time. This means that lenders can be confident that borrowers will not default on their liabilities, although the borrowers do not deposit collateral for the loaned assets [15].

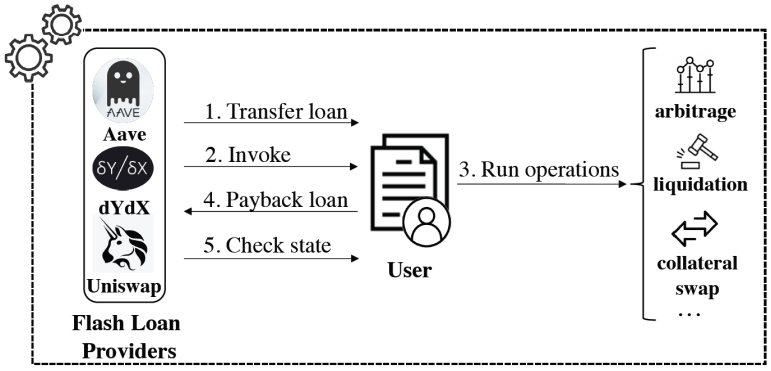

Figure 7.5: The Workflow Of A Flash Loan Transaction [15]

7.2.5 Market Making

While in CeFi the order-book model is very famous, a new exchange technique with the name Automated Market Maker (AMM) is used in DeFi. An AMM is in essence a smart contract which receives financial instruments from liquidity providers [15]. The AMM works with a liquidity pool that mostly contains two assets which are valued relative to each other [30]. Therefore, traders deal against the AMM smart contract and not with the liquidity providers in a direct way. Through AMM, less interactions are needed from the market makers than required in the CeFi order-book model [15]. For the exchange to work, AMMs of DEX use a so-called peer-to-pool method, where liquidity providers (LPs) provide financial assets to liquidity pools. Traders then may exchange assets with pools which have the corresponding assets. Therefore, traders receive the needed liquidity without the need of finding a counterparty for the exchange. At the same time, liquidity providers can benefit from their asset supply with exchange fees from traders [18]. Using the peer-to-pool method, AMMs set the asset price with algorithmic functions. To be more precise, AMMs use a conservation function which determines the price of the assets by enforcing the price to move along predefined graphs. The conservation function is responsible for encoding a prefered invariant property of the ecosystem [18].

i Asset classes incorporated by AMM

In AMM protocols, some different distinct asset classes for operations, governance and controlling are used [18]. *Pool shares*: Pool shares also are called liquidity shares or LP shares. They are held by LPs and are an indicator for who owns which assets within a pool. According to the distribution of corresponding shares, trading feeds will be split proportionally and can be used to withdraw assets from the pool [18]. *Protocol tokens / governance tokens*: With this class of tokens, voting rights on the protocol are represented. When debating a protocol governance subject, traders can use this token to vote [18].

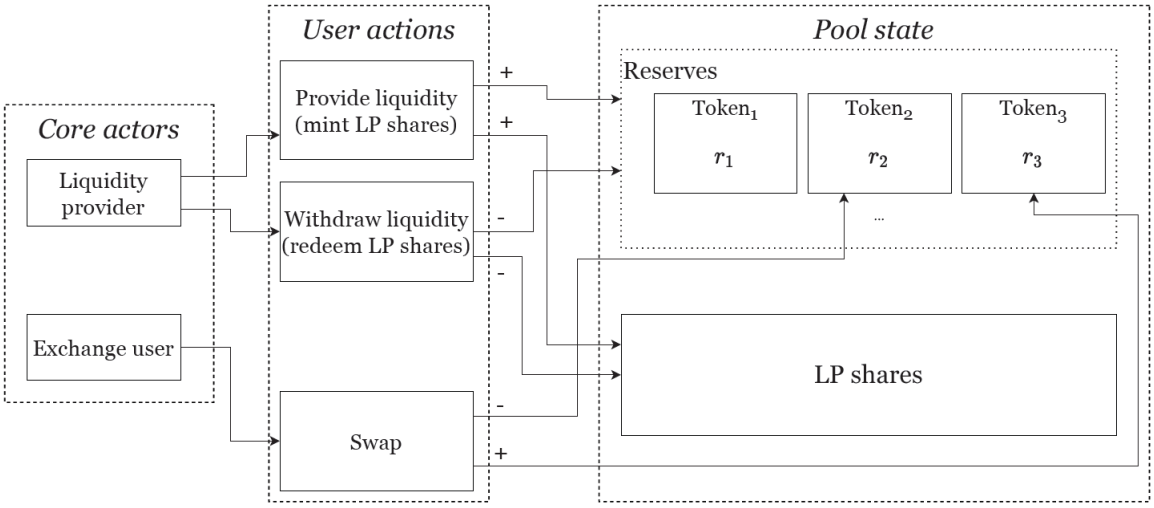

Figure 7.6: AMM Mechanism [18]

7.2.6 Stablecoin/Pegged Asset

Cryptocurrency are often associated with price volatility, which attracts speculators. In order to have more conservative or sophisticated traders who prefer price stability in an asset, stabelcoins were designed for satisfying such a demand. In the following figure, the different types of DeFi stablecoin mechanisms are described [15].

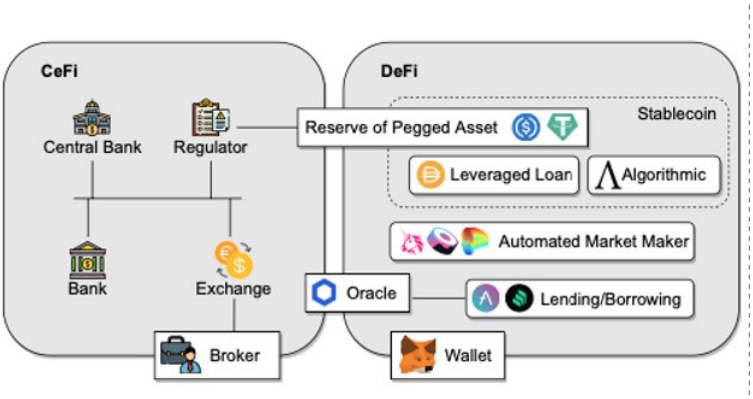

Figure 7.7: Stablecoin Types [15])

i Reserve of Pegged Asset

The most known type is the reserve of pegged asset, which takes a fiat currency for example USD as a collateral that the stablecoin has to be pegged to in a reserve for backing the value of the stablecoin. This type of mechanism requires a centralized and trusted authority, hence, the type overlaps with the CeFi ecosystem. The drawback of this type is the centralized governance structure, which is useful for regulation compliance by blacklisting addresses, however, it harms the ideals of DeFi [15].

ii Leveraged Loan

The leveraged loan mechanism uses similar to the previous type also collateralization to secure the value. In contrast, though, it uses a cryptocurrency as collateral, which also does not require anymore a centralized entity, which is also more in accordance with decentralization in the DeFi ecosystem. One example of a prominent stablecoin is the DAI that facilitates a such mechanism [15].

iii Algorithmic

The last type uses a completely different mechanism instead of using any collaterals of fiat or cryptocurrencies, it is based on algorithms that maintain the price of the stablecoin autonomously. Basically, this means that the algorithm adjusts the supply of stablecoin in response to the price fluctuation, to have the desired target price respectively preserve price stability. Since this type works autonomously by using smart contracts, it requires no central entities, however, algorithmic stablecoins are the most volatile stablecoins, because the supply of the stablecoin is manipulated [15].

7.2.7 Privacy Mixer

Privacy mixer is a service or technique that offers privacy in cryptocurrency transactions. The mechanisms can be explained the following way: If a person wants to send a cryptocurrency coins to another person, but said person would like to preserve its privacy by hiding the address, where the sent asset comes from, then a privacy mixer service can be used by paying an additional transaction fee. The mixing service will shuffle the sent asset with others to delink the trail of the origin transaction address, then the recipient will receive the exact amount of coins but from a different source, so it can't be traced back [20].

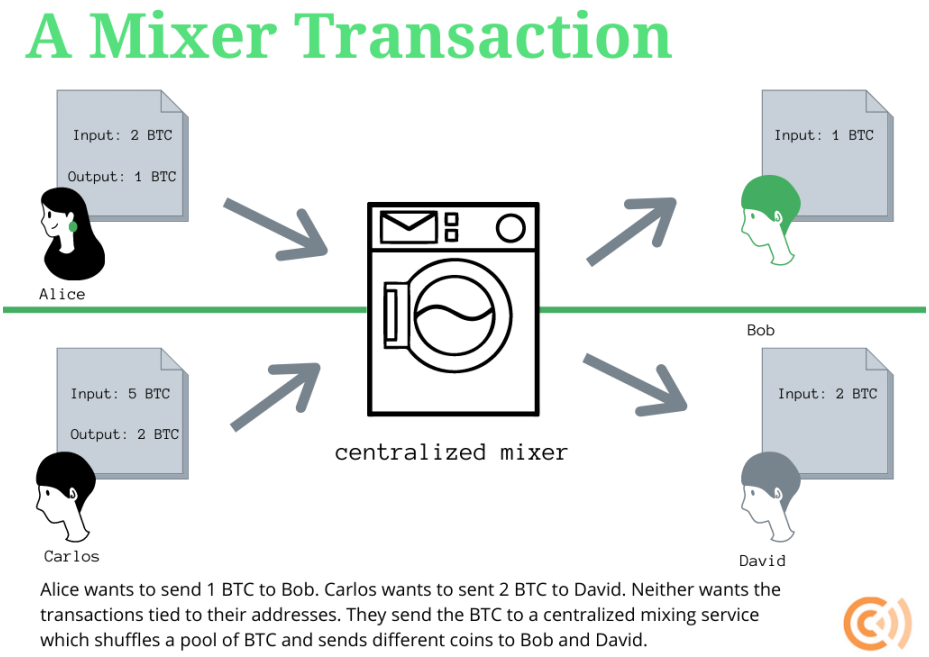

Figure 7.8: Privacy Mixer Service [20]

7.2.8 Derivatives

Derivative trading is very popular on exchanges based on looking at the trading volume numbers of coinmarketcap.com [21]. Derivative trading can be understood as trading with cryptocurrency tokens, but instead of trading directly with the assets, the multiple parties enter a financial contract, that speculates on the future price of a cryptocurrency asset. This advanced trading functionality offers many opportunities for traders. One one hand, you can reduce your risk as a long-term investor by using the crypto derivative types of futures and options. For example, an investor can enter a financial derivative contract,

which will trade its asset as soon as the price goes to a certain low limit, in order to protect itself from a specific amount of loss [22].

7.2.9 Portfolio Management

Since DeFi removes any central governance or someone who manages the assets of an user, this will shift the responsibility to the user. As a consequence, there is a demand for portfolio or asset management tool, where it tracks the portfolio of investors or traders [23].

7.3 An Overview of Market Manipulations

In the following section, the different types of market manipulations will be covered, which has been an issue for DEXs in the DeFi ecosystem [15].

Figure 7.9: Market Manipulation Types [15]

7.3.1 Front-Running

Front-running is a well-known concept, because it already occurs in high-frequency trading of traditional stocks. Basically, if someone got insider information that someone else will buy a large amount of a specific asset, then this person can use this knowledge to its advantage by buying the assets before occurence of the large trade since this will drive the value of said asset [24]. In the DeFi ecosystem, such market manipulation is prevalent, especially when exchanges have low transaction speed respectively are vulnerable to network congestions. This makes it easier for the front-runner to anticipate large trades since transactions are transparent in the network. If the front-runner discovers a large trade to profit from, a higher transaction fee can be paid, such that the transaction of the front-runner is prioritized by the network miners since miners order the transactions by the highest gas price paid. There is a limit to this market manipulation, because the more front-runners there are, the higher the transaction fee has to be, in order to do front-running [25].

7.3.2 Back-Running

The concept of back-running works the following way: Here, the back-runner tries to do the transaction after a large trade has been done. The reason is that the sold coins of the large trade are less valuable respectively the new, bought coins gained in value. If

the back-runner buys the sold token, when the price is low and sells it to another DEX for a premium price, then the back-runner makes a profit by creating arbitrage [25]. The issue of back-running is that in contrast to the previous market manipulation type, it has no financial barrier by paying more transaction fees and when more traders use such a strategy, then the exchange is vulnerable to network congestions [25].

7.3.3 Pump and Dump

Pump and dump schemes are often occurring in cryptocurrency markets due to the lack of regulations and there are a plethora of crypto assets with low market capitalization, which are easier to manipulate the price. Here, the illegal scheme tries to drive the crypto asset based on misleading or false information to make a profit fro. For example, there are social media influencers, who have bought some worthless cryptocurrencies, when the price was low and then these influencers try to convince their followers to buy that crypto asset to drive the price despite the asset having no real value. As soon as the asset hits a certain price level, the influencer will sell its asset to make a profit [26].

7.3.4 Wash Trading

Wash trading is a process where a trader buys and sells the same asset, in order to feed false signals to the market. This can be either achieved by buying and selling the same token back on forth to drive the trading volume. This has again an effect on the liquidity, which is an indicator, how attractive an exchange is for a trader. By falsifying the signal, a trader can't properly determine, how attractive an exchange really is. This illegal process is prohibited, but still common on DEXs [27].

7.3.5 Market Cornering

Market cornering means that someone or an institution has a number of shares of a specific asset, in order to manipulate the price in the market, usually a niche market. For such market manipulation, it requires deep pockets from the manipulator side [28].

7.3.6 Bear Raiding

Bear rading is similar to pump and dump, but the other way around. It is as well an illegal practice, where you try to push down the asset price via collusion or spreading false, negative rumors about the said asset. The bear raider makes a profit from a short position, this means that the asset will be sold high and bought low [29].

7.4 Implementation of DEXs

In the following section, two different types of exchanges will be presented, Uniswap and Binance. These exchanges can be differentiated by their degree of decentralization.

7.4.1 The Uniswap Protocol

The Uniswap protocol was created by Hayden Adams on 22. November 2018 [31]. Then, the company Uniswap Labs continued to develop the Uniswap protocol together with a web interface.To this date, the latest version of Uniswap is the third one (Uniswap V3) [30].

i Introduction

The Uniswap protocol is a system of persistent smart contracts which are non-upgradable. The Uniswap protocol contains an AMM which simplifies peer-to-peer market making and swapping functionalities of ERC-20 tokens which operate on the Ethereum blockchain. Furthermore, the Uniswap protocol was designed to resist against censorship, to be secure, to strengthen self-custody and to work without institutions like central banks that may block certain users from interacting with the exchange [30].
In the Uniswap protocol, all services are open for public use. There is no way to (arbitrarily or selectively) restrict users from accessing the protocol. Therefore, all traders are allowed to provide liquidity, to swap assets or design and create new markets. This is a major difference to CeFi where access restriction can be applied [30].

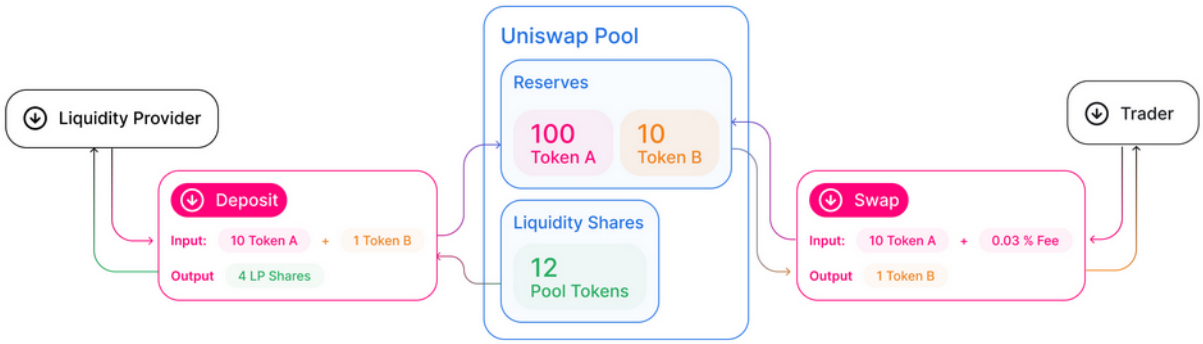

Figure 7.10: Overview Of The Uniswap Exchange [32]

ii Uniswap V3 and concentrated liquidity

The constant product function of Uniswap governs trading movements between pool assets. The mentioned function conserves the product of value-weighted quantities of the two tokens. When performing a trade, the value removed in the first asset must equal the value added in the second one. Through this design, Uniswap supports one of its invariant property goals, namely the previously mentioned weight-preserving characteristic [18].

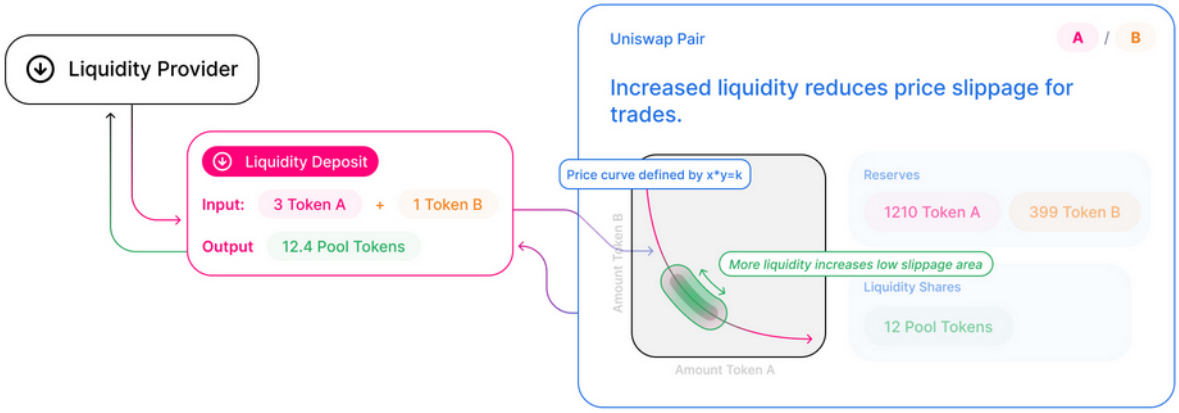

Figure 7.11: Liquidity Provision And The Price Curve [32]

In Uniswap V3, the new concept of concentrated liquidity was added. Previously, on Uniswap V2, liquidity in a liquidity pool was distributed on the whole price curve which resulted in inefficient liquidity provision, since for example for stablecoin pairs the relative price did not move a lot and therefore a lot of the liquidity was rarely used. In Uniswap

V3, traders may specify a price range in which their liquidity position shall be activated. This results in the fact that traders may offer liquidity around the mid-price of the asset and they gain more trading fees. Every liquidity provider can have multiple liquidity positions with custom price ranges within different liquidity pools [40].

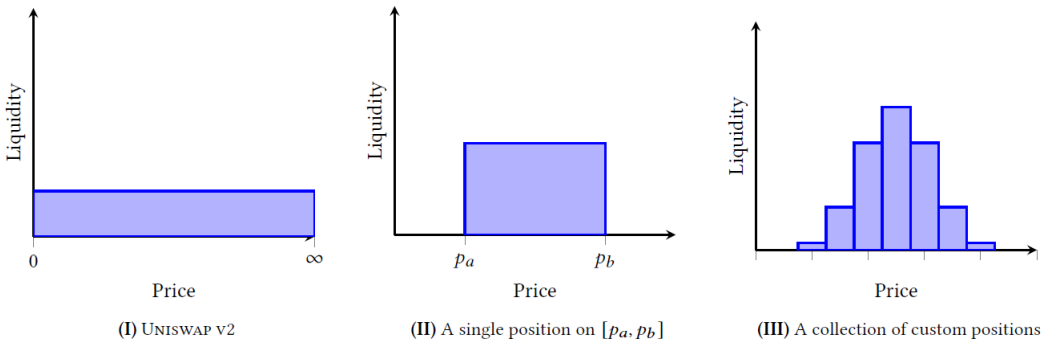

Figure 7.12: Liquidity Distribution: Uniswap V2 And V3 [41]

iii Uniswap V3 and the Smart Contract Architecture

The architecture of Uniswap V3 is in essence a binary smart contract system which is divided into Core and Periphery Contracts [33].
The core contracts are guaranteeing safety for all entities that interact with the Uniswap protocol. These contracts define the sequence and constraints of pool generation, the attributes of the pools, the flow of funds and how traders can interact with the funds [33]. Periphery contracts on the other hand were built to provide functionalities for interacting and working with the core contracts. The goal is to strengthen clarity and the safety of the traders. Usually, external network calls will firstly interact with the periphery interfaces [33].

7.4.2 Binance

In this section, the implementation of Binance will be introduced. How Binance has developed over time and what its characteristics are as an exchange. Then, the architecture of Binance will be presented.

i Introduction

Binance (BNB) is one of the leading cryptocurrency exchanges in the world, but in contrast to Uniswap, it is a more centralized exchange. The exchange has launched in the year 2017 the Binance coin, which is based on the Ethereum platform as an ERC-20 token. This utility token enabled users to pay lower transaction fees on the exchange by using the token for transaction [42].
The Binance Coin was moved in the year 2020 to the Binance Smart Chain, which can be contributed to the following reasons. On one hand, Binance has scalability issues due to being based on the Ethereum platform and thus, the exchange had network congestion problems. On the other hand, the exchange wanted to build smart contract functionalities off-chain, in order to not sacrifice performance in terms of transaction speed and fee. The Binance Smart Chain, which is a hard fork of Ethereum, enabled through their smart contract capabilities a platform that offers DeFi services for their users. Binance Smart Chain uses a Proof-of-Staked-Authority (PoSA) as a consensus mechanism [42].
With this transition, Binance CEO categorized its exchange as CeDeFI, because it has several centralized components. On one hand, it has an ID verification process and saves

your personal information and on the other hand, the governance structure is centralized, which will be discussed in the architecture section. On top of these centralized components, it offers all DeFi services compared to other DEXs like Uniswap [42].

ii Architecture

Binance Smart Chain's off-chain solution enables Binance to still offer high performance in transaction speed. The cross-chain and oracle module enables the communication between two independent blockchains like Binance and Binance Smart Chain [43]. On top of the Binance Smart Chain, there is an Ethereum virtual machine also called EVM, which allows any smart contract that runs on EVM to be transferred to the Binance Smart Chain [43]. It was mentioned before that the consensus mechanism is PoSA. This mechanism works the following way: All transactions are validated by a set of nodes. Validators can either be active or inactive. The maximum amount of validators is 21. Only active validators can validate transactions and are ranked based on the amount of BNB tokens held. The top 21 active validators with the highest amount of BNB are eligible for validating transactions and taking turns. In summary, it is a combination of proof-of-staking and -authority. A limited set of validators are producing blocks and are ranked based on staking governance. These validators take turns based on a proof-of-authority manner. The small set of validators and the financial requirements of becoming a validator has a centralized component to it [42].

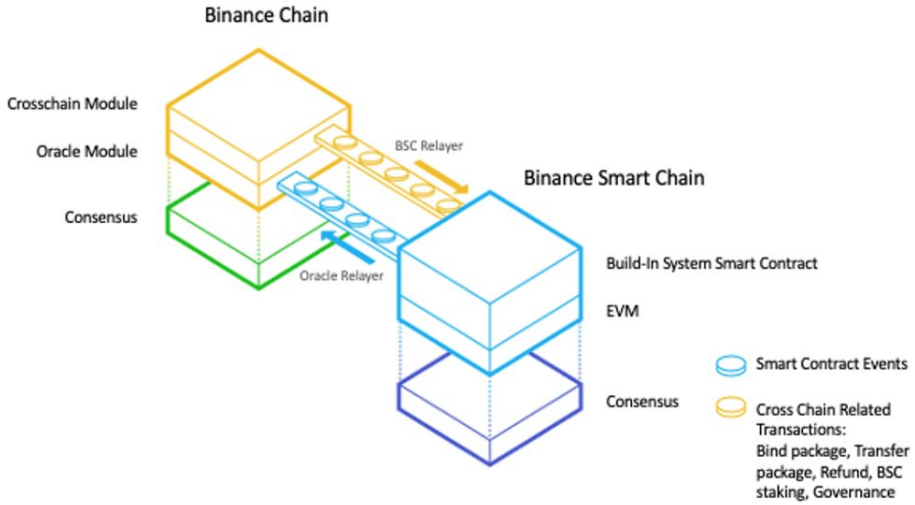

Figure 7.13: Binance Architecture

7.5 Comparison of Binance and Uniswap regarding DEX Services

In this section, the two exchanges, Uniswap and Binance, will be compared based on the DEX services they provide. The comparison for both exchanges was done based on the mentioned services in section 7.2.

Table 7.1: Comparison of Binance and Uniswap with regards to their Services (own representation)

Functionality of Service/Entity	Uniswap	Binance
On-chain or off-chain?	On-chain AMM	Off-chain
Swapping possible?	Yes	Yes
Lending and borrowing possible?	No	Yes
Do Flash Loans exist?	Yes	No
How does Market Making work?	Using AMM	Using AMM
Do Derivatives exist?	No	Yes
Can you trade stablecoins?	Yes	Yes
Does it contain a privacy mixer?	No	No
Does it contain portfolio management?	Yes	Yes
Whats the role of smart contracts?	Basis of AMM	Basis of AMM

7.5.1 Uniswap Services

The Uniswap protocol makes use of an on-chain AMM [18]. This AMM is responsible for managing the liquidity pools which are used for swapping functionalities. Therefore, when one asset is exchanged for the other one, there is a shift in the relative prices of these two financial instruments. The AMM system then determines a new market rate for both assets [30].
For listing a token on Uniswap, the only requirement the token has to meet is the ERC-20 standard [15]. For liquidity providers, it is possible to earn swapping fees which are proportional to their provided capital [39].
To our knowledge, classic borrowing and lending concepts as known in CeFi do not exist in Uniswap, as they appear on Aave for example. We assume that this service was never prioritized by the leading developers of Uniswap, but may be added in a later stage, since classic borrowing and lending concepts are well-known to people around the world. This simplicity of understanding may increase the number of usages of this service.
The concept of flash loans was introduced in Uniswap for version 2 (UniswapV2). In Uniswap, Flash Loans are called flash swaps and can be called on the *swap()* function. Uniswap V2 charges a swap fee of 0.3 percent which is based on the financial assets which the trader borrowed [19]. Flash Loans are also implemented in Uniswap V3.
Furthermore, Uniswap does not offer derivatives [17]. We assume that Uniswap did not add derivatives, because derivatives (as known in CeFi) may have a complex nature and depend on liquid markets. Since DeFi markets may encounter a lack of liquidity, derivatives may be difficult to implement.
Trading with stablecoins is possible on Uniswap. Prominent examples are Tether (USDT), the US Dollar Coin (USDC) or DAI. We assume that trading with stablecoins is possible because Uniswap wants to give users the possibility to trade with all kinds of tokens and wants to increase the number of risk averse traders. Risk averse traders may profit from the nature of stablecoins, since they are less volatile and pegged to a currency.
We also find that Uniswap does not contain a privacy mixer. We assume that this is the case, because privacy mixer may be complex to implement and to understand.
Also, Portfolio management is possible with Uniswap. A trader may hold one or several tokens and can manage them when using the Web interface or communicating directly with the smart contracts.

7.5.2 Binance Services

In contrast to Uniswap, Binance processes its transactions off-chain. Unlike their counterpart, they offer advanced derivatives functionalities and portfolio management features. One advantage of Uniswap is that it offers flash loan services. Otherwise, in all other DEX service categories, it resembles Uniswap with for example the basis of AMM. Worth mention is the fact that the comparison table would have looked different before Binance moved to Binance Smart chain because back then Binance used an orderbook and did not have all the DeFi services ready [44].

7.6 Uniswap and Binance: Market Manipulation Comparison

In this section, the two exchanges, Uniswap and Binance, will be contrasted based on the vulnerability of market manipulations. The comparison for both exchanges was done based on the mentioned market manipulations in section 7.3.

Table 7.2: Comparison of Binance and Uniswap with regards to possible Market Manipulations (own representation)

Market Manipulation type	Uniswap	Binance
Front-Running	Possible	Slim chance
Back-Running	Possible	Possible
Pump and Dump	Possible	Possible
Wash Trade	Possible	Possible
Market Cornering	Possible	Possible
Bear Raid	Possible	Possible

7.6.1 Market Manipulations on Uniswap

When analysing Uniswap with regards to possible Market Manipulations, one can find out that front- and back-running are possible. Both market manipulation types can appear in combination on Uniswap, which is then called a sandwhich attack [18].
Also, Pump and Dump may appear on Uniswap, as it happened in the case of the Youtube Patrick Shyu that was accused of using Pump and Dump techniques on Youtube when offering the token *million token* [37].
Wash Trading is a market manipulation type that can appear on Uniswap too. The Website theblockresearch.com states that wash trading happened on Uniswap [38].
Since it is possible on Uniswap to acquire enough shares of an asset to manipulate its price, we state that Market Cornering is possible on Uniswap.
Since Bear Riding is nowadays highly reliant on Social Media, it cannot or be prevented by the Uniswap protocol. We therefore state that Bear Riding is possible on Uniswap.

7.6.2 Market Manipulations on Binance

In contrast to Uniswap, Binance has an in-build anti-frontrunning system, which prevents such market manipulation by avoiding any network congestion respectively making it easier for the frontrunner [45]. Otherwise, it is similar to Uniswap regarding the vulnerability of market manipulation.

7.7 Conclusion

We conclude that DEXs in general have the same basic idea which is to handle and work with transactions in a decentralized way but often differ in their system architecture. While they often use an AMM for creating liquidity or to determine the price for an asset, some DEXs still use the order-book model to handle the transactions. There are many different kinds of AMMs and we state that the full potential of AMMs has not been explored fully yet. Even though blockchains are highly important for DEXs, some exchanges (off-chain exchanges) decide to not store all information immediately on the blockchain, while other exchanges do (on-chain exchanges). We state that off-chain exchanges may increase the energy efficiency of the system. Off-chain exchanges introduce the concept of centralization again, which DEXs originally tried to avoid. Therefore, we state that exchange developers must be careful not to design a centralized system when designing a decentralized one. Furthermore, we conclude that DEXs can make use of a large variety of different services, but often do not implement all of them. The reason may be the complexity of implementing those services or the lack of demand for them.
In conclusion, Binance and Uniswap represent two different governance structures. On one hand, Uniswap does all its transactions on-chain, which is the most transparent approach. However, this idealistic view sacrifices in terms of performance. This means lower transaction speed and higher fees for the users. This avoided Binance by introducing Binance Smart Chain to do the orders on a different blockchain (off-chain), as a consequence, Binance could retain the performance, which might be the reason for being the leading cryptocurrency exchange, when looking at the trading volume or liquidity. Additionally, Binance offers advanced derivatives features, hence the crypto derivatives market is flourishing according to coinmarketcap based on trading volume [21]. But, both exchanges show weaknesses in terms of market manipulations. As a future outlook, there needs to be more improvement done similar to the traditional stock market. There are already potential solutions ready to be further explored and implemented.

Bibliography

[1] Malamud, Semyon, and Marzena Rostek: *Decentralized Exchange* https://www.aeaweb.org/articles?id=10.1257/aer.20140759, November 2017.

[2] Shubhani Aggarwal, Neeraj Kumar: *Chapter Fifteen - Blockchain 2.0: Smart contracts* https://www.sciencedirect.com/science/article/abs/pii/S006524582030070X, October 2020.

[3] Fabian Schaer: *Decentralized Finance: On Blockchain- and Smart Contract-Based Financial Markets* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3843844, May 2021.

[4] IBM; https://www.ibm.com/blockchain/solutions/digital-assets, November 2021.

[5] Forbes; https://www.forbes.com/sites/philippsandner/2021/08/24/digital-assets-the-future-of-capital-markets/, November, 2021.

[6] Ryan Amsden, Denis Schweizer: *Are Blockchain Crowdsales the New 'Gold Rush'? Success Determinants of Initial Coin Offerings*, April 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3163849#references-widget.

[7] Binance Academy Glossary: Tokens; https://academy.binance.com/en/glossary/token, November 2021.

[8] Binance Academy Glossary: ERC-20; https://academy.binance.com/en/glossary/erc-20, November 2021.

[9] Binance Academy Glossary: Fungibility; https://academy.binance.com/en/glossary/fungibility, November 2021.

[10] Angelo Aspris, Sean Foley, Jiri Svec, Leqi Wang: *Decentralized exchanges: The wild west of cryptocurrency trading*, International Review of Financial Analysis, Volume 77, October 2021. https://doi.org/10.1016/j.irfa.2021.101845.

[11] Lindsay X. Lin, Legal Counsel at Interstellar and Stellar Development Foundation: *Deconstructing Decentralized Exchanges*, Stanford Journal of Blockchain Law & Policy, January 2019. https://stanford-jblp.pubpub.org/pub/deconstructing-dex/release/1.

[12] Binance Academy: A Guide to Crypto Collectibles and Non-fungible Tokens (NFTs); https://academy.binance.com/en/articles/a-guide-to-crypto-collectibles-and-non-fungible-tokens-nfts?utm_source=BinanceAcademy, November, 2021.

[13] JAKE FRANKENFIELD: *Smart Contracts* https://www.investopedia.com/terms/s/smart-contracts.asp, November 2021.

[14] DXC Learn: *What is a Decentralized Exchange (DEX)?* https://dcxlearn.com/trading/what-is-a-decentralized-exchange-dex-2/#Ownership_suspense, November 2021.

[15] Kaihua Qin, Liyi Zhou, Yaroslav Afonin, Ludovico Lazzaretti, Arthur Gervais: *CeFi vs. DeFi – Comparing Centralized to Decentralized Finance*, June 2021. https://arxiv.org/abs/2106.08157v2.

[16] Pengcheng Xia, Haoyu wang, Bingyu Gao, Weihang Su, Zhou Yu, Xiapu Luo, Chao Zhang, Xusheng Xiao, Guoai Xu: *Trade or Trick? Detecting and Characterizing Scam Tokens on Uniswap Decentralized Exchange*, November 2021. https://arxiv.org/abs/2109.00229.

[17] Yuen C Lo, Francesca Medda: *Uniswap and the Emergence of the Decentralized Exchange*, October 2020.https://dx.doi.org/10.2139/ssrn.3715398.

[18] Jiahua Xu, Krzysztof Paruch, Simon Cousaert, Yebo Feng: *SoK: Decentralized Exchanges (DEX) with Automated Market Maker (AMM) Protocols*, October 2021. https://arxiv.org/abs/2103.12732.

[19] Dabao Wang, Siwei Wu, Ziling Lin, Lei Wu, Xingliang Yuan, Yajin Zhou, Haoyu Wang, Kui Ren: *Towards A First Step to Understand Flash Loan and Its Applications in DeFi Ecosystem*, May 2021. https://doi.org/10.1145/3457977.3460301

[20] Andrea O'Sullivan: *What are mixers and "privacy coins"?*, July 2020. https://www.coincenter.org/education/advanced-topics/what-are-mixers-and-privacy-coins/

[21] CoinMarketCap: *Top Cryptocurrency Derivatives Exchanges*, November 2021. https://coinmarketcap.com/rankings/exchanges/derivatives/

[22] Cem Dilmegani: *What are Crypto Derivatives? Types, Features & Top Exchanges*, November 2021. https://research.aimultiple.com/crypto-derivatives/

[23] Diego Geroni: *An Overview On Asset Management In DeFi*, July 2021. https://101blockchains.com/defi-asset-management/

[24] CORY MITCHELL: *Front-Running*, November 2021. https://www.investopedia.com/terms/f/frontrunning.asp

[25] Marius Van Der Wijden: *Backrunning in DeFI*, August 2020. https://medium.com/@m.vanderwijden1/backrunning-in-defi-301f3cade30a

[26] RAJEEV DHIR: *Pump and Dump*, June 2021. https://www.investopedia.com/terms/p/pumpanddump.asp

[27] Guénolé LePenneca, Ingo Fiedler, Lennart Ante: *Wash trading at cryptocurrency exchanges Author links open overlay panel*, November 2021. https://www.sciencedirect.com/science/article/abs/pii/S1544612321000635

[28] JAMES CHEN: *Corner A Market*, April 2021. https://www.investopedia.com/terms/c/corneramarket.asp

[29] CORY MITCHELL: *Bear Raid*, June 2021. https://www.investopedia.com/terms/b/bearraid.asp

[30] The Uniswap Protocol: Introduction; https://docs.uniswap.org/protocol/introduction, November 2021.

[31] Uniswap Blog: History https://uniswap.org/blog/uniswap-history/, November 2021.

[32] The Uniswap Protocol: How Uniswap works https://docs.uniswap.org/protocol/V2/concepts/protocol-overview/how-uniswap-works, November 2021.

[33] The Uniswap Protocol: Smart Contracts Overview https://docs.uniswap.org/protocol/reference/smart-contracts, November 2021.

[34] Monolith: Aave, the DeFi lending protocol https://medium.com/monolith/monolith-spotlights-aave-the-defi-lending-protocol-a45edb3f0da0, November, 2021.

[35] Will Kenton; *Order Book* https://www.investopedia.com/terms/o/order-book.asp, November 2021.

[36] Ivan Jericevich, Dharmesh Sing, Tim Gebbie: *CoinTossX: An open-source low-latency high-throughput matching engine*, February 2021. https://arxiv.org/abs/2102.10925.

[37] John Kumi; *YouTuber soll Uniswap fuer Pump-and-Dump-Betrug benutzt haben*, https://www.crypto-news-flash.com/de/youtuber-soll-uniswap-fuer-pump-and-dump-betrug-benutzt-haben/, November 2021.

[38] Igor Igamberdiev; *A close look at trading volumes on Uniswap and SushiSwap*, https://www.theblockresearch.com/a-close-look-at-trading-volumes-on-uniswap-and-sushiswap-101532, November 2021.

[39] The Uniswap Protocol: Swaps https://docs.uniswap.org/protocol/concepts/V3-overview/swaps, November 2021.

[40] The Uniswap Protocol: Concentrated Liquidity https://docs.uniswap.org/protocol/concepts/V3-overview/concentrated-liquidity, November 2021.

[41] Hayden Adams, Noah Zinsmeister, Moody Salem, River Keefer, Dan Robinson: *Uniswap v3 Core* https://uniswap.org/whitepaper-v3.pdf, March 2021.

[42] jakub: *Binance Smart Chain and CeDeFi Explained*, April 2021. https://finematics.com/binance-smart-chain-and-cedefi-explained/

[43] ibdi.it: *Binance Smart Chain: Ein kompletter Leitfaden zur Binance Smart Chain (BSC)*, May 2021. https://ibdi.it/de/binance-smart-chain-a-guide-complete-to-binance-smart-chain-bsc-2/

[44] Binance: *FAQ*, November 2021. https://www.binance.com/en/support/faq

[45] Binance Chain Docs: *Anti Front-Running*, November 2021. https://docs.binance.org/anti-frontrun.html

Chapter 8

The Market of Digital Espionage and Forensics

He Liu and Rathes Sriram

The Market for Digital Espionage and Forensics is a growing part of the cyber security industry and impacts our daily lives mostly without us even knowing. Although it is claimed that it only serves the governments to protect their citizens and the fight against criminals and terrorists, concerns about the actual and potential misuse are justified. The existing regulations have not caught up yet. This paper will introduce leading forensic solutions based on the number of targets, focus on the most relevant ones, and explain how new fields in digital forensics such as Internet-of-Things forensics and cloud computing forensics were created. Further, we analyzed the techniques for Mobile Forensics, then categorized their tools and described their different features. We continue to introduce key players of this Market, their flagship products, and the suspected customers. In taking an economic point of view, we analyze the Digital Forensic Market and the Digital Espionage Market separately to their current state and trends and see that the Market will continue to grow.

Contents

8.1 Introduction and Motivation

Most of our lives have virtually become wholly intertwined with digital devices and information systems: virtually everything we do today is done through or in conjunction with a digital device or platform. In the digital age, information security and assurance issues abound, and with increased technological advancements, criminals are also improving their skills and causing more and more havoc. Digital forensic investigations are used to ensure information assurance and security by discovering how an incident connected to an electronic device occurred and possibly tracing and apprehending those behind it [1].
Digital forensic science is a branch of forensic science that focuses on recovering and investigating material found in digital devices related to cybercrime. The process of digital forensics is identifying, preserving, analyzing, and documenting digital evidence. It is done to present evidence in a court of law when required. The term digital forensics was first used as a synonym for computer forensics. Since then, it has expanded to cover the investigation of any devices that can store digital data [2].
Digital forensics is becoming more challenging due to the tremendous increase in computing devices and computer-enabled paradigms, providing new challenges to the distributed digital data processing and adding to the investigative complexity. Nowadays, there are different types of digital forensics like disk forensics, memory forensics, network forensics, mobile forensics, malware forensics, and so on [1].
In contrast to the largely stationary internet of the early 2000s, people all over the world today are increasingly connected to the world of digital information while on the go via smartphones and other mobile devices [3]. Figure 8.1 gives an example of mobile phone ownership over time in the United States. By February of 2021, around 97% of adults in the United States owned a cell phone. The smartphone share was 85%, which increased from 35% in 2011. Figure 8.2 shows that more than 15% of American adults are smartphone-only internet users - meaning they own a smartphone but do not have traditional home broadband service [3]. As smartphones have become an integral part of peoples' lives, criminals also rely on them. Therefore, mobile forensic is playing an essential role in modern digital investigations [4].
The increasing use of IoT devices has led us to save a lot of personal data on them [5]. So while we used to keep phone numbers in telephone directories, collect our images in photo albums, and write our appointments in an agenda, we began in the last decade to save them all in one place - our smartphones. Smartphones have developed themselves into a tool such that it reveals what kind of person their owner is. For example, what this person's hobbies are, where they drink their coffee, or when and where they have a regular meeting. Consequently, if one tries to access as much information as possible about a person, electronic devices (especially smartphones) are becoming targets. This is where Digital Espionage tools come into play.
Digital espionage is one way of hacking with political or economic intentions. For example, stealing secret information to develop new technologies based on the stolen data or political reasons falls under this category. This is mainly done without the victim not even knowing or not realizing for a period [6].
Since the unveiling by Edward Snowden in 2013 of the extent of espionage activities by US agencies, data privacy and digital espionage has been brought to our consciousness. This way of hacking and extracting information is facing much criticism from society. Even though this method is seen as critical, the demand is growing. On the other side, possible victims try to defend themselves by using encrypted web pages and applications. The attempt to outdo each other seems to continue [7].

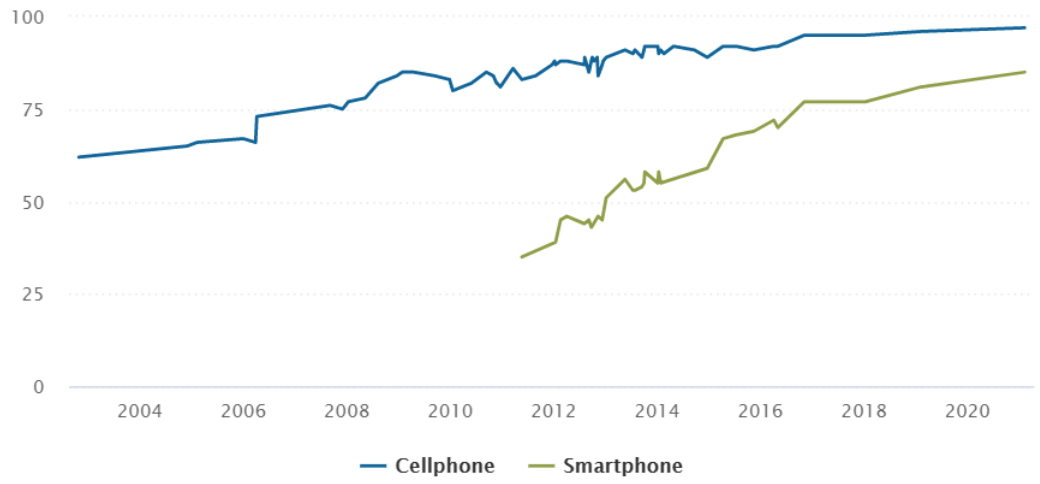

Figure 8.1: Mobile Phone Ownership (% of U.S. adults) [3]

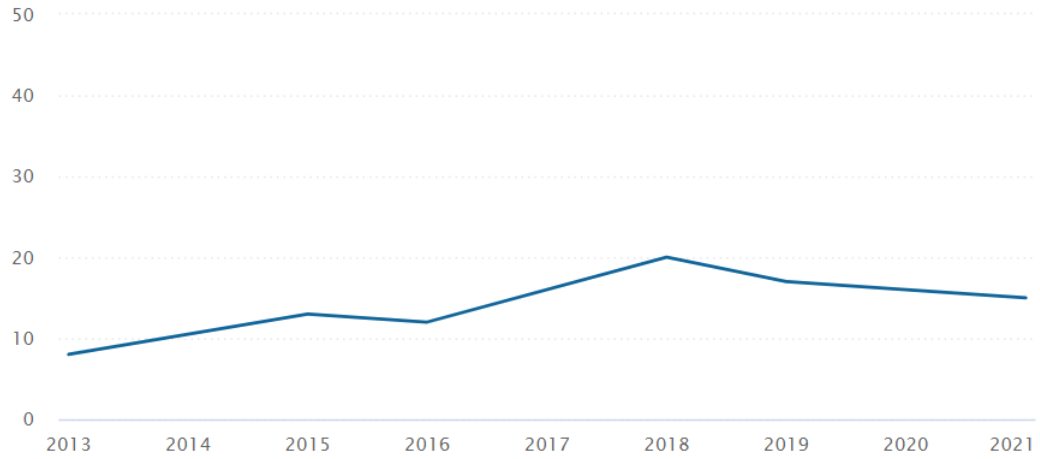

Figure 8.2: Smartphone Dependency (% of U.S. adults) [3]

8.2 Digital Forensic Targets and Techniques

Digital evidence is the information stored or transmitted in binary form that may be relied on in court [8] and can exist on several different platforms and in many different forms. Therefore, forensics investigation can be done on any kind of device storing or transferring data.

The sources of digital evidence are the targets of digital forensics. All the devices that can store, port, extract or transmit data, such as storage devices, random-access memories, and routers, can be the targets of digital forensics. Besides hardware components and devices, people can also be the targets of digital forensics. Media and elements can sometimes be the targets of state-sponsored hacking.

The goal of digital forensics is to preserve any evidence in its most original form while performing a structured investigation by collecting, identifying, and validating the digital information to reconstruct past events [9]. The techniques used in this process are digital forensic techniques.

This section introduces the leading different forensic solutions from the most common targets. Also, the most relevant used for uncovering and interpreting electronic data are covered.

Non-volatile Storage Device

Non-volatile storage device here means all types of devices that can store and retrieve digital data and can retain stored data even when powered off [10]. Although several

types of these devices, hard disk drive (HDD) and solid-state drive (SSD), are most commonly used in personal computers.

HDD is a legacy low-cost data storage device with spinning disks inside where data is stored magnetically, while SSD is a newer and faster device that stores all the data in integrated circuits. In other words, the main difference between them is the architecture. There are two approaches to forensics investigation on storage devices: hardware recovery and software recovery. In case of hardware failures, hardware recovery is necessary to retrieve the data. The broken components, such as an arm, slider, head, or electronic board, can be replaced for HDDs to restore data. As long as the replacement work is done very carefully and in a clean environment, an investigator can retrieve the data in the end in most cases [12]. However, recovering data from SSDs is more complicated. SSD typically uses multiple NAND flash memory chips, and all the memory chips and control chips are soldered to one board. Therefore, replacing them by hand is a fragile and challenging job and nearly impossible for multiple chips [13]. Although chip readers can separately access the data on each chip, it is hard to restore the SSDs with multiple chips because different manufactures often use different strategies on how to address chips, how to perform wear leveling and garbage collection, and how to distribute data [11].

Software recovery is the most common way to recover information from a storage device that is still working correctly. Generally, when a file is deleted from a file directory or removed from the computer Recycle Bin, it is only marked for deletion and is not removed from an HDD [15] until being over-written by a new file [14]. That is because the magnetic status of the untouched block on an HDD remains the same until being altered. Therefore, removed data can be recovered easily from HDDs, and data recovery software usually utilizes this fact to restore data. Unlike HDDs, deleted files and data on SSDs the suspect attempted to destroy may be lost forever in a matter of minutes. Although some forensics software still provides the function for restoring data on an SSD, the recovering rate is usually unsatisfied [11].

Random-access Memory

Random-access Memory (RAM) is a volatile storage device much faster than HDD and even SSD. It gives programs a place to store and access actively-using data on a short-term basis. The forensic information in RAM is often overlooked [16]. Although the information on it will be lost when the power is switched off, any program must be loaded in memory to execute when the power is on. Lots of precious information, such as open network connections, account credentials, and running processes, is solely on it [17]. Another advantage of memory analysis is that the size of RAM is usually much smaller than a disk, so it can be much quicker when transferring the output. Therefore, RAM forensics is an integral part of digital forensics analysis.

Network

Network forensics is a comparatively new and growing field of forensic science. With the growing popularity of the Internet, more and more things in our daily life are becoming on the cloud, and their data is outside of a disk-based digital form. Similar to the data in RAM, the data on the Internet is volatile, as it will be lost when its transmission is over. Unlike memory and other areas of digital forensics, network forensics needs to deal with highly dynamic data. Therefore, it is often a pro-active investigation [8].

Network forensics is vital for investigators as it can reveal the data exclusively online. It is also helpful for network administrators to deal with network issues. There are two types of network forensics: *catch-it-as-you-can* and *stop, look and listen* [18].

Catch-it-as-you-can , or capturing all of the data, is the more thorough version of network forensics. In which all packets passing through a certain traffic point are captured and written to storage, with analysis being done subsequently in batch mode. In-

vestigators can look backward for all the logs within a reasonable time. However, it needs ample storage space to house the data while it is being processed.

Stop, look and listen , or temporary detention, is a more sophisticated approach. In which each packet is analyzed in a rudimentary way in memory, and only suspicious or predetermined information will be saved for future analysis. It needs less storage than *catch-it-as-you-can* and can keep track of ongoing information. However, it may require a quick processor to continually keep the data flow moving.

Both of these types are helpful for investigators, but the sources might be various. Generally, there are two main resources for investigation:

Network packet is a formatted unit of data the Internet carries. It contains control information and user data. In other words, from packet analysis, investigators can playback the activities of Internet users with details [19].

Log files reside on web servers, proxy servers, Active Directory servers, firewalls, Intrusion Detection Systems (IDS), DNS, and Dynamic Host Control Protocols (DHCP).

Besides security-related usage, Internet forensics can also be used in some exciting areas, such as *forensics on espionage.* With the growing demand for monitoring Internet activities, several companies develop and supply spyware to governments or other customers and help them to collect signals intelligence. Although almost all of those companies claim that they respect everyone's privacy and their products are only used to investigate terrorism and crime and leave no traces whatsoever, they never, and might not be possible to give any proof of it.
Some research labs, companies, and non-profit labs are identifying what those surveillance companies are doing. Methodologically, it is rarely possible for researchers to have access to the devices of both producers and their customers. Network forensics become the only possible solution in this case. Pegasus is spyware developed by NSO Group, an Israeli surveillance firm, and is capable of reading text messages, tracking calls, collecting passwords, location tracking, accessing the target device's microphone and camera, and harvesting information from apps [20]. The details of NSO will be introduced in Section 8.4.1. Since it is hard to find any information other than the advertisements and reports from NSO Group itself, an interdisciplinary laboratory, Citizen Lab, in the University of Toronto, tried developing Internet scanning techniques to identify what Pegasus spyware may be conducting operations.
The research from Citizen Lab between August 2016 and August 2018 tried to find out the suspected infections of Pegasus spyware. The researchers in this project used DNS Cache Probing to study the matching domain names to identify the infected countries [21]. Specifically, they hypothesized that devices infected with Pegasus would regularly look up Pegasus's servers using their ISP's DNS servers, then probed ISP DNS caches around the world to find out the targets. The findings from Citizen Lab will be discussed as an application scenario in Section 8.6.

Mobile Device

A mobile device can be described as a computer small enough to hold and operate in hand, but most existing computer forensics techniques cannot be applied directly on a mobile device. With the increased availability and popularity of such devices on the consumer market and the wider range of information stored or transmitted on them, the demand for mobile device forensics is growing [22].
The phrase *mobile device* usually refers to *mobile phone* in daily life. However, it can also relate to any portable digital device such as a tablet, smartwatch, GPS device. This report mainly focuses on smartphones.

Mobile device forensics is a branch of digital forensics relating to recovering digital evidence from a mobile device. It dates from the late 1990s, and early 2000s [8]. Early efforts to examine mobile devices focused on acquisition techniques and general forensic analyses of computer and other intelligent devices: analyzing phone contents directly via the screen and photographing important content [8]. Enterprising mobile forensic examiners often synchronized mobile device data to a forensic computer, then performed it as computer forensics. However, as the number of mobile devices increased, investigators called for more efficient means of extracting data. Additionally, computer forensic tools could not read or write a mobile device directly and could not retrieve removed data [23].
In 2002, Burnette discussed the forensic examination of two older types of RIM (BlackBerry) devices. In this paper, he described the hardware and software used for acquisition, and several examination methods, including the use of hex editors and emulators. He also discussed the difficulty of analyzing always-on, push messaging devices: without truly turning the radio off, the removed data might be overwritten by coming information [24]. Grand introduced a forensic tool for memory imaging and forensic acquisition of data from the Palm OS, a ceased mobile operating system, in the same year [25].
The development of smartphones makes our lives convenient and easily connected with other people, but it is also true for criminals. The proliferation of mobile phones in society has led to a concomitant increase in their use in and connected to criminal activity, so mobile forensics has become an essential aid to law enforcement in the investigation of crime [26]. From an investigative perspective, digital evidence recovered from a cell phone can provide a wealth of information about the user, and each technical advance in capabilities offers a more significant opportunity for recovery of additional information [27].
With the global pandemic due to the novel Corona Virus (COVID-19), people worldwide moved their work or study online, which increased the use of mobile applications. In 2020, the number of smartphone subscriptions surpassed 6 billion [28]. Meanwhile, people spent over 180 billion collective hours in the third quarter of 2020 using mobile applications [29]. During that same time, several social networking and Voice-over-IP (VoIP) apps were being downloaded or actively used included Facebook, Facebook Messenger, Spotify, Telegram, TikTok, Twitter, WhatsApp Messenger, and ZOOM Cloud Meetings [29].
This uptick in mobile app usage and cyberattacks requires digital forensics investigators and researchers to keep pace with mobile application forensics. By the December of 2020, over 72.5% of the smartphones in the global market are running Android while the Apple iOS accounts for 26.9% [30]. Therefore, examiners must understand operating systems, file systems, and folder structures. They also need to know what items of potential evidentiary value could be found on each device as well as where data is stored and different forensic data extraction tools and techniques that could be of use when examining both Android and iOS devices [31]. Meanwhile, with the rapid and frequent updates in OS and applications, data storage and structure may change when a significant update is installed in mobile devices, which causes potential disruptions to the forensic tools' ability to acquire and interpret user data [32]. Therefore, mobile forensics should always be adapted to newer systems. The following section will introduce and discuss several mobile forensic techniques and tools adapted to relatively new mobile operating systems.

8.3 Scenario of Mobile Forensics

This section introduces mobile devices' forensic techniques and tools and starts with the data acquisition techniques. The data on a cell phone should be acquired for analysis at the beginning of mobile forensics. The two most common techniques for data extraction are physical acquisition and logical acquisition.

Physical Acquisition captures all of the data on a physical piece of storage media. It is a bit-for-bit copy, like the clone of a hard drive [33]. It is often done through JTAG or cable connection. It allows the forensic tool to collect remnants of deleted data.

Logical Acquisition is a technique for extracting the files and folders without any deleted data from a mobile device. However, some vendors describe logical extraction narrowly as the ability to gather a particular data type, such as pictures, call history, text messages, calendar, videos, and ringtones [34]. It often occurs via Bluetooth, infrared, or cable connection.

Several digital forensic tools were created to examine mobile devices efficiently to observe data on a device without damaging it. Generally, mobile forensic tools can be categorized as Figure 8.3

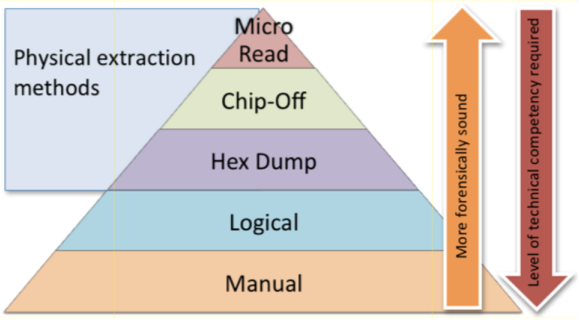

Figure 8.3: Mobile Forensic Tool Analysis Pyramid [35]

Manual Extraction allows investigators to extract and view data through the device's touchscreen or keypad. At a later stage, this data is documented photographically. Furthermore, manual extraction is time-consuming and involves a great probability of human error. For example, the data may be accidentally deleted or modified during the examination.

Logical Extraction as already mentioned before, the investigators connect the cellular device to a forensic workstation or hardware via Bluetooth, Infrared, RJ-45 cable, or USB cable. The computer with logical extraction tools sends a series of commands to the mobile device. As a result, the required data is collected from the phone's memory and sent back to the forensic workstation for analysis purposes.

Hex Dump , or physical extraction, as mentioned before, extracts the raw image in binary format from the mobile device. The forensic specialist connects the device to a forensic workstation and pushes the boot-loader into the device, instructing the device to dump its memory to the computer. This process is cost-effective and supplies more information to the investigators, including recovering the phone's deleted files and unallocated space.

Chip-off allows the examiners to extract data directly from the flash memory of the cellular device. They remove the phone's memory chip and create its binary image. This process is costly and requires ample knowledge of hardware. Improper handling may cause physical damage to the chip and renders the data impossible to retrieve.

Micro Read involves interpreting and viewing data on memory chips. The investigators use a high-powered electron microscope to analyze the physical gates on the chips and then convert the gate level into 1's and 0's to discover the resulting ASCII code. This process is expensive and time-consuming. Also, it requires ample knowledge of hardware and file systems. There is no tool available for micro read [36].

There are also various types of tools available for mobile forensic purposes. They can be categorized as open-source, commercial, and non-forensic tools. Both non-forensic and forensic tools frequently use the same techniques and protocols to interact with a mobile device.

Before using mobile forensic tools, investigators should first understand the architecture of mobile operating systems. For Android devices, the Linux kernel is responsible for managing the core functionality of Android, such as process management, memory management, security, and networking. The file hierarchy is designed as a single tree with the top being denoted as the root, where specific folders are only visible through root access. Rooting is the process of gaining privileged access on an Android device, which is very beneficial for examiners [37]. Examiners need to make an informed decision about where to look for data. The several vital partitions that are common to most Android devices and key for examiners to understand are as follows [38]:

Boot This information and partition is required for the phone to boot and contains the kernel and RAM disk. Data residing in the RAM contains important information for an examination and should be captured.

System Contains the system-related files other than the kernel and RAM disk and should never be deleted as it will make the device unbootable.

Recovery This is designed for backup purposes and allows the device to boot into recovery mode.

Data This is where each application's data is stored, which is of great importance to an examination. Data belonging to the user, such as contacts, SMS, and dialed numbers, is stored in this partition.

Cache This is where frequently accessed data and app components are stored, which is beneficial to examiners to know which data is regularly accessed by the device user.

Misc This partition contains information about miscellaneous settings, and information about hardware settings and USB settings can be accessed from this partition.

SD card This is where all the information present on an SD card is held. It is valuable to an examination as it could contain pictures, videos, files, documents, and more.

SQLite database is the most common data storage format used on Android devices to save crucial data, storing all user information in the form of files and holding a wealth of valuable data for examiners to extract and analyze. SQLite is the database for internet browsers, web applications, and software products to keep their data. A database may have usernames, passwords, account numbers, and even deleted data. It may also have browser history, including downloads, keywords, and URLs. SQLite database allows for analyzing and extracting key artifacts from social networking applications such as Whatsapp, Facebook, and Skype. [39]

Although Android and iOS are both smartphones, they are vastly different device types in how data is stored. iOS filesystem is configured into two logical disk partitions: system partition and user data partition. The system partition contains the OS and all preloaded applications used with the iPhone but contains little useful evidentiary information. User-created data is in the user data partition and can provide the most evidentiary information pertinent to examiners.

Examiners can also find a wealth of information from the backups of both Android and iOS devices. From logical acquisition, examiners may extract users' sensitive data such as contacts, SMS, photos, calendar, call logs, configuration files, documents, and the data

in password managers. The backup also contains the device details such as the serial number, Unique Device Identifier, SIM details, and phone number [38].
To extract data from filesystems or backups, investigators need to observe forensics tools and select a suitable one (or more). The research was done by Lwin et al. [41] compared the performance of an open-source tool and a commercial tool on two Android devices, a Galaxy Note 4 running Android 6 and an Oppo A83 running Android 7. In order to get access to the data directory, both devices were rooted. They first compared physical acquisition tools then compared analyzing tools. Both open-source tools and commercial tools are included in two comparisons.
In order to acquire logical data, the authors used three tools: ADB Backup, Belkasoft Acquisition, and Magnetic Acquire tools. Linux DD, Belkasoft Evidence Center, and Magnetic Acquire tools were used for physical acquisition. Belkasoft Evidence Center and Magnetic Acquire can be used for logical and physical acquisition. Then the researchers used Autopsy and Belkasoft Evidence Center to analyze the data acquired.

ADB Backup is an official way to leak private data in certain apps. It is also useful for backing up the entire device, as what the authors did in this research.

Magnetic ACQUIRE is a commercial but free forensic tool from Magnet Forensics. It is a community version of Magnet AXIOM and is designed to use the ADB process to acquire the application data from the device. It uses the agent application to acquire select application data that may be available to be obtained in addition to the ADB-recovered data (for example, SMS/MMS, Contacts, browser history) if it was not found in the ADB backup [42].

DD command is a utility on Unix and Unix-like operating systems for duplicating disks, and it is free of charge, the primary purpose of which is to convert and copy files. Researchers of this paper used it to get the physical image of the tested phones.

Belkasoft Evidence Center is the flagship digital forensic suite. It supports both computer and mobile forensics. The product makes it easy for an investigator to perform all steps of modern digital investigation such as Data acquisition from various devices and clouds, artifact extraction and recovery, analysis of extracted data, reporting, and sharing evidence. It also offers a comprehensive analysis of devices. Figure 8.6 gives an example of geolocation analysis on the GPS-enabled data from photos, URLs, mobile apps like Uber.

Autopsy is a free and open-source analysis tool that can analyze the most common Android file systems. It is a GUI-based, trusted, and easy-to-use digital/mobile forensic platform. Basis Technology Corp. largely maintains the tool with the assistance of programmers from the community. Figure 8.4 shows a configuration step after DD file and Figure 8.5 shows the automatically categorized results.

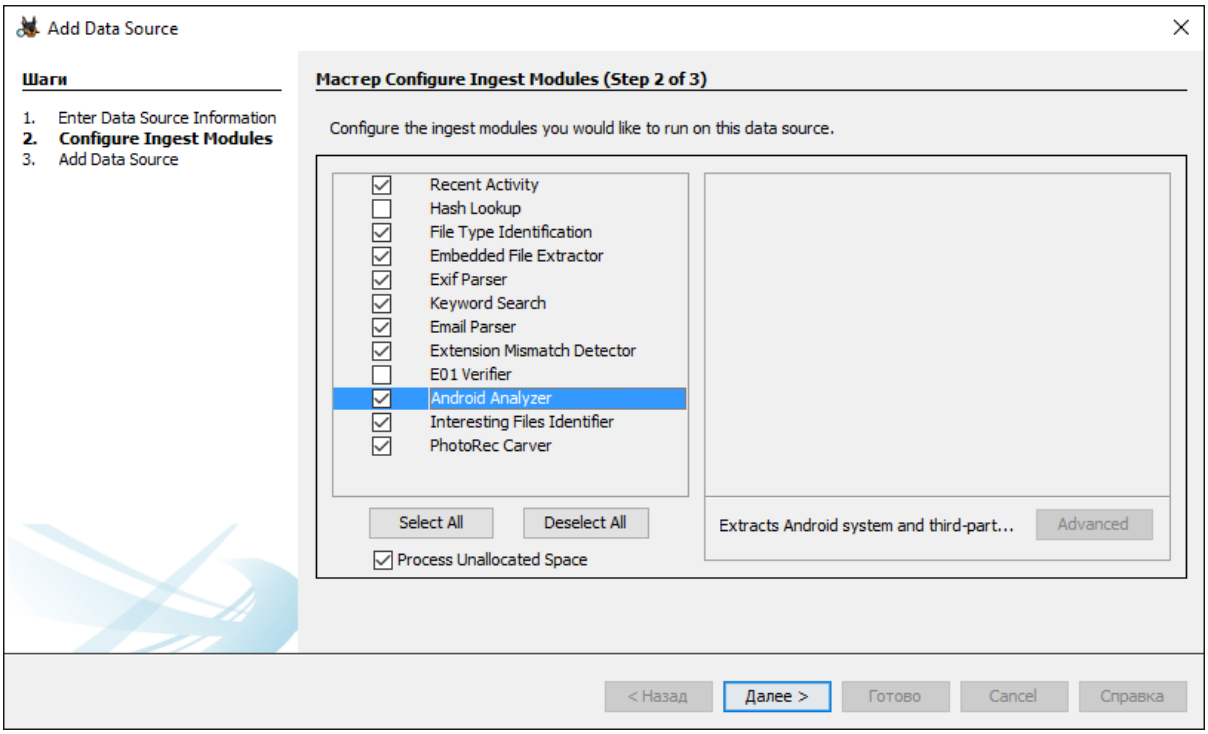

Figure 8.4: A configuration step in Autopsy [40]

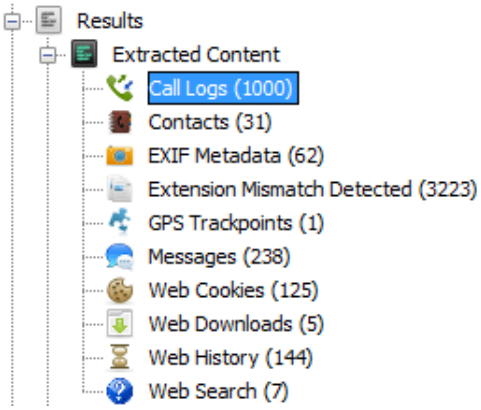

Figure 8.5: An extracted result from Autopsy [40]

Tables 8.1 and 8.2 compares the performance of logical acquisition tools and physical acquisition tools respectively. Tables 8.3 and 8.4 are for two analyzing tools. During analysis, Belkasoft could find encrypted passwords, but Autopsy could not do it in both cases. According to the results, it is recommended to use Magnet Acquire for logical acquisition and DD command for physical acquisition. For analyzing tools, they found that Belkasoft worked the best and claimed that there was no single tool that could get and analyze all sorts of data. In general, the researchers concluded that commercial tools could save time and provide the investigator with more accurate results and more extracted data, especially in the analysis process. In the end, they suggested researchers use multiple tools to get integrity and accurate result.

Table 8.1: Comparison for Logical Acquisition Tools [41]

	ADB Backup	Magnet ACQUIRE	Belkasoft
Size (GB)	11.6	15.0	1.41
Time (hh:mm)	03:00	01:00	00:10
GUI	No	Yes	Yes
Cost	Free	Free (Request needed)	Trial

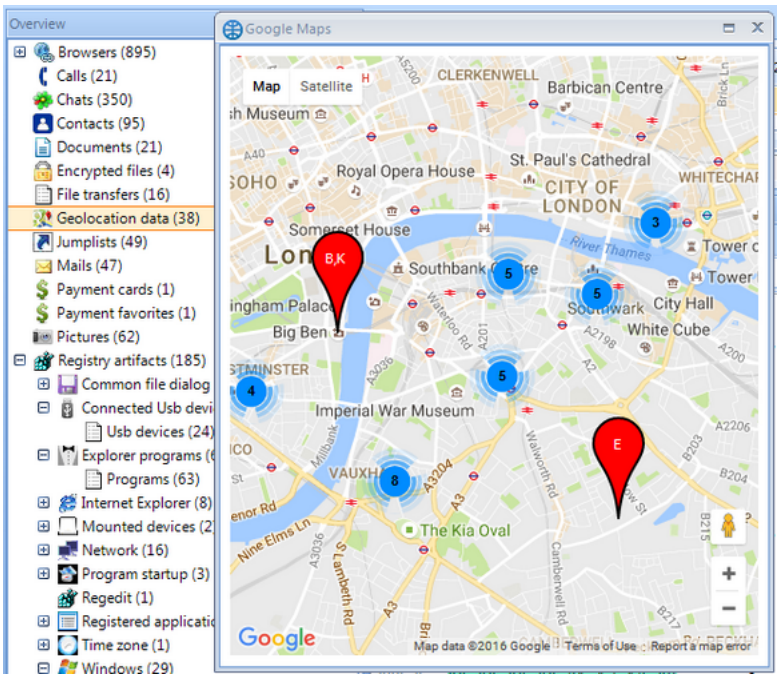

Figure 8.6: An example of Belkasoft geolocation analysis [43]

Table 8.2: Comparison for Physical Acquisition Tools [41]

	DD command	Magnet ACQUIRE	Belkasoft
Size (GB)	8.3	9.7	8.3
Time (hh:mm)	01:40	01:15	00:40
GUI	No	Yes	Yes
Cost	Free	Free (Request needed)	Trial

Table 8.3: Comparison for Logical Data (Based on [41])

	Autopsy (Samsung)	Belkasoft (Samsung)	Autopsy (Oppo)	Belkasoft (Oppo)
Categories	14	31	14	31
Artifacts	25113	107004	9912	51237
Report	Yes	Yes	Yes	Yes
Time (hh:mm)	00:39	00:40	00:15	00:20

Table 8.4: Comparison for Physical Image (Based on [41])

	Autopsy (Samsung)	Belkasoft (Samsung)	Autopsy (Oppo)	Belkasoft (Oppo)
Categories	23	31	23	31
Artifacts	68258	110425	30971	52070
Report	Yes	Yes	Yes	Yes
Time (hh:mm)	01:49	00:45	00:49	00:30

Another research [44] analyzed 27 Android and 33 iOS mobile applications comprehensively. The applications ranged from instant messaging and social networking apps to VoIP and vault apps. They used an iPhone SE running iOS 13 and a Google Pixel 3 running Android 10. They first used Cellebrite UFED and Magnet Acquire to acquire the data and then used Autopsy and Magnet AXIOM to validate it. The validation process can be found in Figure 8.7.

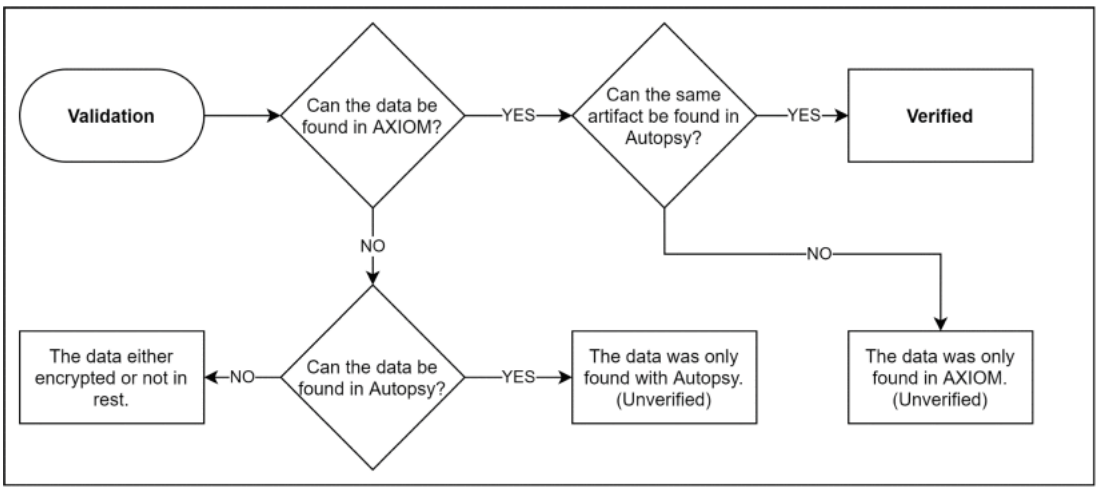

Figure 8.7: Validation Methodology [44]

Sensitive data can be retrieved in their research. One example is the artifacts related to the Gallery Vault application in Android. Encrypted file name, authentication email, number of launch times, and the number of times the user was discovered in an XML file. Meanwhile, the database file containing several tables with invaluable information could be restored. Lastly, the user's personally identifiable information (PII), such as geolocation, IP address, and home address, can also be recovered from a JSON file.

ZDATECREATED	ZBODY	ZVOICEMAILURL	ZCONTACTPHONENUMBER
608346235.487	Hey. This message is coming from my Burner number.	(null)	+19195790479
608346378.287	Got it. It's showing up as a 704 number. Is that right?	(null)	+19195790479
608346564.502	Yes. I can not send pictures or videos.	(null)	+19195790479
608346578.011	My subscription is not high enough.	(null)	+19195790479
608346681.811	That is fine. Just do a call and we will call it a day.	(null)	+19195790479
608346898.438		(null)	+19195790479
608346921.556		https://s3.amazonaws.com/burner-voicemail/prod/a77b40c9-dd98-492b-8b1b-96a91a842c32.wav	+19195790479

Figure 8.8: Sent and received messages on the Burner app [44]

Another example is the information found in the Burner application in iOS. This application allows users to create temporary disposable phone numbers. Although the phone numbers are temporary, Burner's users are only semi-anonymous because their temporary numbers are linked to their actual phone number and email address. In the analysis, the most relevant forensic data from this app was recovered from several tables within a database by Magnet AXIOM, which included the user's email address, actual phone number, Burner number, and chat messages (see Figure 8.8). Additionally, each record's numerical timestamp (i.e., ZDATE) can be easily converted to a human-readable date. This app also records a URL link to any voicemails left for the app user. These links were confirmed to still be live some three months after being created.

Besides these two examples, invaluable data can also be recovered in other applications. Tables 8.6 and 8.7 shows the summary of artifacts recovered from Android 10 and iOS 13 applications. The tables in this report only retain applications in both Android and iOS. The meanings of the symbols in the table are in Table 8.5.

From their findings, several conclusions can be made. First of all, except for passwords and IP addresses, artifacts can be recovered in most of the tested applications. Secondly, the artifacts recovered from Android devices are more than iOS devices. Lastly, most artifacts could be recovered by both tools, but the number of artifacts that could be found

Table 8.5: Meanings of the symbols

Symbol	Meaning
Yes-M	Artifact was found with Magnet Axiom only
Yes-A	Artifact was found with Autopsy only
Yes-V	Artifact was found with both Magnet Axiom and Autopsy
Yes-I.S.	Artifact was found if shared by the user in chat
*	Artifact was partially recovered; Missing relevant data
No	Artifact was not found with Magnet Axiom nor Autopsy
N/A:	Not applicable
Enc	Encrypted
U	Unpopulated

by Magnet AXIOM is slightly more than Autopsy's, which is similar to the research [41] where commercial tools performed better than open-source tools.

Table 8.6: Summary of recovered user credentials, contact, and PII (Based on [44])

App\Artifacts	OS	Username	Password	User Contact	User ID	User Location	User IP Address
Discord	Android	Yes-V	No	Yes-V	Yes-V	No	No
	iOS	Yes-V	No	Yes-V	Yes-V	No	No
Dust	Android	Yes-V	N/A	Yes-V	Yes-V*	U	No
	iOS	Yes-V	No	Yes-V	No	No	No
Facebook Messenger	Android	Yes-V	No	Yes-V	Yes-V*	Yes-V	No
	iOS	Yes-V	No	Yes-V	No	No	No
Imgur	Android	Yes-V	No	Yes-V	Yes-V	Yes-V	Yes-V
	iOS	No	No	No	Yes-V	No	No
Imo	Android	Yes-V	No	Yes-V	No	No	Yes-V
	iOS	Yes-V	No	Yes-V	No	No	No
Instagram	Android	Yes-V	No	No	Yes-V	No	Yes-A
	iOS	Yes-V	No	No	Yes-V	No	No
MeWe	Android	Yes-V	No	Yes-V	Yes-V	No	No
	iOS	Yes-V	No	No	Yes-V	Yes-V & I.S.	No
Signal	Android	Yes-V	N/A	Yes-V	Yes-V	No	No
	iOS	No	No	No	No	No	No
Silent Phone	Android	Yes-V	No	Yes-V	Yes-V	No	No
	iOS	Yes-V	No	Yes-V	Yes-V	No	No
Skout	Android	Yes-V	Yes-V	Yes-V	Yes-V	Yes-V	No
	iOS	No	No	No	Yes-V	No	Yes-V
Skype	Android	Yes-V	No	Yes-V	Yes-V	Yes-V*	Yes-V
	iOS	Yes-V	No	Yes-V	Yes-V	Yes-V & I.S.	Yes-V
Snapchat	Android	Yes-V	No	Yes-V	Yes-V	Yes-V	No
	iOS	Yes-V	No	Yes-V	Yes-V	No	No
Spotify	Android	Yes-V	No	Yes-V	Yes-V	No	No
	iOS	Yes-V	Enc	No	Yes-V	No	Yes-M
Telegram	Android	U	No	Yes-V	Yes-V	No	Yes-V
	iOS	N/A	No	Yes-M	Yes-M	Yes-V & I.S.	Yes-V
TextNow	Android	Yes-V	No	Yes-V	Yes-V	Yes-V	Yes-V
	iOS	Yes-M	No	Yes-V	Yes-V	N/A	No
Threads	Android	Yes-V	No	Yes-V	Yes-V	No	Yes-A
	iOS	Yes-V	No	No	No	No	No
TikTok	Android	Yes-V	No	Yes-V	Yes-V	No	No
	iOS	No	No	No	Yes-V	No	No
Twitter	Android	Yes-V	No	U	Yes-V	No	No
	iOS	Yes-V	No	No	Yes-V	No	No
WhatsApp	Android	Yes-V	No	Yes-V	Yes-V	Yes-I.S.	Yes-V
	iOS	Yes-V	No	Yes-V	Yes-V	Yes-V & I.S.	No
Wickr Me	Android	Yes-M	No	No	No	Yes-M & I.S.	No
	iOS	No	No	No	No	No	No
Wire	Android	Yes-V	No	Yes-V	Yes-V	Yes-V	No
	iOS	Yes-V	No	Yes-V	Yes-V	Yes-M & I.S.	No
Venmo	Android	Yes-V	No	Yes-V	Yes-V	No	No
	iOS	Yes-V	No	Yes-V	No	No	No
Viber	Android	Yes-V	No	Yes-V	Yes-V	Yes-I.S.	No
	iOS	No	No	Yes-V	No	No	No

8.4 Digital Espionage and Forensics Market

The use of IoT devices (Internet of Things) has increased tremendously over the past few years [5]. Depending on more such devices, be it at work, in a smart home, or while browsing the internet, the data we generate becomes of more interest for third parties. These are, in particular, governments or police forces that try to extract personal data from persons of interest to prevent certain events (e.g., protests, terror attacks) or to prove guilt by seizing all possible information from these electronic devices. The result of this is the existence of the Digital Espionage and Forensics Market. This section introduces this market by presenting providers, products, and customers. Then analyzes the market by describing the status quo and its existing regulations and continues to explain the observable trends.

Table 8.7: Summary of other recovered artifacts (Based on [44])

App\Artifacts	OS	Text Messages	Shared Media	Phone Calls	Others's Info	Logs
Discord	Android	Yes-V	Yes-V	U	Yes-V*	Yes-V
	iOS	Yes-V	Yes-V	No	No	No
Dust	Android	U	U	N/A	Yes-V*	Yes-V*
	iOS	Yes-V*	No	No	No	No
Facebook Messenger	Android	Yes-V	Yes-V	Yes-V	Yes-V*	Yes-V*
	iOS	Yes-V	No	No	Yes-V	No
Imgur	Android	N/A	Yes-V*	N/A	Yes-V*	Yes-V
	iOS	Yes-V	No	No	No	No
Imo	Android	Yes-V	Yes-V	N/A	Yes-V*	Yes-V
	iOS	Yes-V	No	No	No	No
Instagram	Android	Yes-V	Yes-V*	Yes-V	Yes-V	Yes-V*
	iOS	Yes-M	Yes-V	No	Yes-V	No
Signal	Android	Enc	Yes-V	Enc	Enc	Enc
	iOS	No	No	No	No	No
Silent Phone	Android	Enc	Enc	Enc	Enc	Enc
	iOS	No	No	No	No	No
Skout	Android	Yes-V	Yes-V	Yes-V	Yes-V*	Yes-V
	iOS	Yes-V	Yes-V	N/A	No	Yes-V
Skype	Android	Yes-V	Yes-V	Yes-V	Yes-V	Yes-V*
	iOS	Yes-V	Yes-V*	Yes-V	Yes-V	Yes-V
Snapchat	Android	Yes-V	Yes-V	Yes-V*	Yes-V	Yes-V*
	iOS	Yes-V	No	No	No	No
Spotify	Android	N/A	N/A	N/A	N/A	Yes-V
	iOS	N/A	N/A	N/A	N/A	Yes-V
Telegram	Android	Yes-M	Yes-M	Yes-M	Yes-M	Yes-V
	iOS	Yes-M	Yes-V	Yes-V	Yes-M & A*	Yes-V*
TextNow	Android	Yes-V	Yes-V	Yes-V*	Yes-V	Yes-V*
	iOS	Yes-M & A*	Yes-V	Yes-M*	Yes-V	Yes-V
Threads	Android	Yes-V	Yes-V*	Yes-V	Yes-V	Yes-V*
	iOS	Yes-M	Yes-V	No	No	Yes-V
TikTok	Android	Yes-V	Yes-V*	N/A	Yes-V*	Yes-V*
	iOS	Yes-V*	Yes-V	N/A	Yes-V	Yes-V
Twitter	Android	Yes-M	Yes-M	N/A	Yes-V*	Yes-V
	iOS	Yes-V	Yes-V	N/A	Yes-V*	Yes-V
WhatsApp	Android	Yes-V	Yes-V	Yes-V	Yes-V	Yes-V*
	iOS	Yes-V	Yes-V	Yes-V	Yes-V	Yes-V
Wickr Me	Android	Yes-M	U	Yes-M*	Yes-M*	Enc
	iOS	No	No	No	No	Yes-V*
Wire	Android	Yes-V	Yes-V	Yes-V	Yes-V	Yes-V*
	iOS	Yes-V	Yes-V	No	Yes-V	Yes-V*
Venmo	Android	Yes-V*	N/A	N/A	Yes-V	Yes-V*
	iOS	Yes-V	No	No	No	No
Viber	Android	Yes-V	Yes-V	Yes-V	Yes-V	Yes-V
	iOS	Yes-V	Yes-V	Yes-V*	Yes-V	No

There are a lot of different providers in this market. We can, however, differentiate between the ones in the Espionage Market and the ones in the Forensics Market. The former focuses on providing products that, as an example, help extract data from devices that have been seized, while the latter offers solutions to spy on persons of interest without their knowledge. The following sections will focus on the NSO Group, which provides solutions in the Espionage Market and became known for their infamous product Pegasus and Cellebrite, one of the leading companies in the Digital Forensics Market.

8.4.1 NSO Group

The NSO Group is a technology company based in Herzliya, Israel, focusing on selling spyware software products (e.g., Pegasus). Founded in 2010 by Niv Carmi, Shalev Hulio, and Omri Lavie, it got international attention in 2016 when one of its products could hack into Apple's iPhone. By 2017 they had nearly 500 employees and are now among the most prominent players in this market. Renowned for providing governments internationally with such solutions, they have been linked with human rights abuses in Mexico, UAE, and Rwanda. Nevertheless, to the present day, the NSO Group never admitted doing something wrong and affirms to be selling its technology with caution and only to prevent terrorism and crimes [45]. This is why the company itself describes them as

> *NSO Group develops best-in-class technology to help government agencies detect and prevent terrorism and crime.* [46]

on their webpage. They develop these products intending to help governments, which are licensed by the Israeli Ministry of Defense, to support the intelligence and law enforcement agencies to fight the most dangerous threats, as the prevention of terrorism, the exposure of crime rings, support in finding of abductees and the assistance in emergent rescues. They further describe that their functioning as a firm is based on the four principles *Accountability, Integrity, Excellence*, and *Boldness*, and emphasize its importance. We are going to elaborate on the meaning of these principles [46].

Accountability says NSO, is essential to set high ethical standards for all their activities. During a strict process, they will decide on whom to license their products. They claim that they set the benchmark and act as pioneers for this industry. In an article in The Economist, a spokeswoman of NSO says that the firm did not close deals from 2016 to 2019 worth $100m because the clients did not meet the ethical standards. [7].

Excellence is a factor that has been assigned a high value by NSO Group. Despite the company's growth in the last decade, they want to offer excellent service to their customers. In addition, they are convinced that their technology has made the world a better place, saving thousands of lives and preventing crimes. However, the accusation that NSO's product Pegasus was used to spy on the journalist Jamal Kashoggi (who was later killed brutally in the consulate of Saudia Arabia) was not admitted and is still being denied despite the evidence from investigations. [47].

Integrity is being claimed by them since they assure that their software is only used to help governments and protect citizens. Furthermore, they are committed to assisting them against terror, crime, and other security risks. They guarantee to investigate any misuse of their product. Nevertheless, despite providing proof of the use of Pegasus software in hacking journalists' mobile devices, NSO never took action [48].

Boldness is required by the NSO Group. Even if the principles mentioned before are restricting the business, they emphasize that they want to be bold while adhering to the regulations and not contradicting their values.

Their mission summarizes how NSO sees its role in this world: the company and its employees save lives and create a better, safer world [46].
NSO provides information on their webpage on how they function and what their Governance consists of. They write that they have a Governance, Risk, and Compliance Committee board that monitors all their sales, conducts a detailed analysis and provides recommendations based on that. The Committee can deny sales or demand investigations on suspicion of misuse. It is a very recent process that was triggered by recent negative headlines.
The governance of the NSO Group consists of three crucial policies, namely: *Human Rights Policy* [49], *Transparency* [50] and *Whistleblower Policy* [51][52].

Human Rights Policy is in orientations towards the United Nations *Guiding Principles on Business and Human Rights* guidelines [53] such that every process and employee must comply with those rules. They have summarized thirteen points to which standards they adhere. As a result, to this date, the NSO Group has rejected over $300m in sales opportunities, disconnected five customers from the system following an investigation of misuse since 2016, and discontinued business with five customers due to concerns with human rights [54].

Transparency has a high standard within the company, and they commit to promoting this wherever possible. For the first time, NSO published a Transparency and Responsibility Report for the year 2021 [54], in which they analyzed the effectiveness of their policies and are being transparent with intern decision makings.

Whistleblower Policy consists of an internal [51] and an external policy [52]. There are clear instructions on reporting the misuse of any suspected wrongdoing or dangers concerning the company's activities in these documents. Once the company gets a report from a whistleblower, they will investigate as described in the Product Misuse Investigation Procedure. In the past year, they opened twelve investigations of product misuse [54].

Even if NSO offers a vast range of products, one of them is very well known in this market for its frightening capabilities, usage by different countries, and devastating consequences.

Pegasus is a software product for spyware, which government clients of NSO mainly use. It intends to collect user data from mobile devices of persons of interest. These could be criminals, terrorists, or opponents of the ruling government. The program has different methods to infect mobile devices. One way may be using SMS or iMessage with a link that forwards the user to another webpage. By clicking on it, one downloads malicious software, and the phone is compromised. The more concerning one is, however, using Zero-Click.

Zero Click is an attack using software that takes advantage of a gap in the security system of an electronic device to cause damage. The unique feature is that there is no need for the victim to interact (i.e., no keyboard commands or mouse clicks).

Vulnerabilities in the iMessage application of iPhones allowed to infect the device using the Zero-Click method, i.e., by simply receiving a message, Pegasus hacked the phone. Similarly, one could get hacked by just receiving a Whatsapp call (even if not picked up). The Pegasus-user has now gained complete control of the mobile device and can extract data, take screenshots, turn on the camera or the microphone, and even erase traces. It is observed that the software is more effective on iPhones than on Android-driven devices [55]. Since the sale of this product has to be in alignment

with Israel's Foreign Policy, it has certain restrictions. As an example, Pegasus can not be sold to Iran or Qatar, and it is prohibited to attack any US-American phone numbers [55].

In the last couple of months, the NSO Group has received some attention from the media, thus highlighting the harmful activities supported by this kind of company, directly or indirectly. A summary of related events is provided below.

July 18, 2021: The Pegasus Project is a global consortium of journalists around the globe, which ForbiddenStories coordinated in cooperation with Amnesty International Security Lab. They had access to the leak of 50'000 phone numbers which NSO clients monitored. Analyzing them, they found out that several journalists, human rights activists, political opponents, or heads of states were being surveilled. NSO did not admit any fault [56].

October 24, 2021: In a recent article in the New York Times, Ben Hubbard described how his phone was hacked, probably by the software Pegasus of NSO. Saudi Arabia may be behind these attacks. They seem interested in his profile and contacts since he covers news about the Middle East. In collaboration with the research institute Citizen Lab at the University of Toronto, they tried to figure out what had happened. They couldn't conclude anything with certainty, but it was likely that a suspicious text message Ben Hubbard received had been sent by Saudi Arabia - using Pegasus - to hack his phone. The NSO Group, on the other hand, denies the use of their software in this case. Figure 8.9 shows how the message which the Saudi regime sent by using the Pegasus software looks [48].

Figure 8.9: Screenshot of Ben Hubbard's chat [48]

November 3, 2021: The US government puts NSO, as distributor of Pegasus, on a trade blacklist. According to them, NSO acted *contrary to the national security or foreign policy interests of the United States.* This ban is a rather surprising move the US government took since they are close allies with the Israeli government. NSO Group defended themselves by saying that they look forward to providing complete information on why they still satisfy the US conditions. They would like to continue to maintain US national security [57].

8.4.2 Cellebrite

With its headquarters in Israel, the company has operated since 1999 with fourteen offices worldwide. They claim to be the leader in working with public and private organizations

to support ongoing investigations, assure data privacy, and more. As of now, in over five million analyses, Cellebrite has been used. They serve five thousand public safety customers and about 1'700 enterprise customers.
Cellebrite aims to help institutions organize complex digital investigations connecting it to its Digital Intelligence Investigative Plattform and offering services in this area. They are convinced that their technology allows in bringing justice to victims of several crimes like killing, assaults, drug, and human trafficking, or any financial crimes [58].
The Board of Directors sets the guidelines to which every employee of the company must adhere. In addition, they are responsible for overseeing the management of Cellebrite and ensuring that all the activities are within the range of the policies. The following three documents form the basis for that: Code of Business Conduct and Ethics [59], Corporate Governance Guidelines [60], and Whistleblower Policy [61].

Code of Business Conduct and Ethics: In this policy, Cellebrite specifies its expectations towards its employees. Activities should be in the interest of the company and with no conflicts. In addition, the employees should permanently preserve the reputation of the company. Further, they emphasize the importance of Human Rights and the confidentiality of their data. There are case scenarios described that introduce how to react to align with these policies.

Corporate Governance Guidelines: It states the mission and responsibilities of the Board of Directors. For example, besides strategic and financial decisions, the Board is in charge of monitoring the effectiveness of their policies. It is mainly responsible for conforming to laws and regulations.

Whistleblower Policy: The primary purpose is to encourage people to report misuse. Any employee may submit their concern in a confidential, anonymous manner. If it is not possible to report it to the manager, there is a Corporate Development Officer or even the Audit Committee. There are clear instructions on how to proceed in this case. It is guaranteed that Cellebrite will take no measures against employees who decide to step out and protect their identity.

One can divide the products of Cellebrite into three categories: (a) Collect&Review, (b) Analyze&Investigate, and (c) Manage&Safeguard. Most of the products are within category (b). Therefore, this is the category of Cellebrite tools which this report will focus on.

Cellebrite UFED: This is the flagship product of Cellebrite. One can use it to get full access to the mobile data on the devices, even if there are complicated locks or encryption barriers, and extract it. In addition, Cellebrite UFED can restore deleted or unknown content as well. These options are beneficial during investigations. It supports more than 30'000 device profiles and covers a wide range of Android and Apple products.

UFED is usable in a software format (UFED 4PC) on the existing PC or laptop, or it can be bought as a tablet with specific features (UFED Touch2). Cellebrite even offers it with a computer (UFED Ruggedized Panasonic Laptop) loaded with this software and additional digital forensic accessories [62].

Cellebrite UFED Cloud: It is similar to Cellebrite UFED, but additionally, this product enables it to access cloud-based content, extract, preserve and analyze it. UFED Cloud can be purchased as a single solution or as an add-on to other products [63].

Digital Collector: This works similar to UFED but is used for Windows and Mac computers [64].

8.4.3 Business Models of Providers

To analyze the Business models of the providers in the Digital Espionage Market, we will consider the Canvas Business Model with the following points: *Key Partners, Key Activities, Key Resources, Value Propositions, Customer Relationships, Channels, Customer Segments, Cost Structure, and Revenue Streams.* Each one of these points is described as follows.

Key Partners: Companies Key Partners in this field are research teams that provide the latest results. Developers will then implement new methods based on the results of these researches to their products.

Collaboration with Intelligence agencies or local authorities is essential as well since, in this way, they get to know about the latest methods used and what - as a potential client - could be of interest. Depending on that, the companies can start to develop solutions for these problems. Further, the countries' governments in which the products are sold are vital since they set the product's framework conditions and consequently regulate providers' activities in this market.

Key Activities: One of the main activities of companies in this market is building products that make it possible to access data from digital devices. Besides programming such tools, the maintenance of it and updates considering the quickly changing environment are essential too.

Key Resources: Human Resources are of great importance in this field since creating such products requires knowledge on an extremely high level. The providers have to make sure that they can recruit talented developers. Additionally, they depend on the software and licenses that the companies will sell.

Value Propositions: Most providers focus on government authorities (e.g., intelligence agencies) and customize their products for them. They aim to help them in preventing terrorism, exposing crime rings, assistance in rescues, and other activities.

Customer Relationships: The relationship with a customer depends on how frequently the product is used. In the case of a police force of a larger city, the support will be higher than a single-user case. A customer center and technical support ensure that the users always have someone for help.

Channels: Most customers are reached remotely and rarely in person. However, it may be the case that in the period right after selling the product, the client needs support on-site. Later the demand for this may decrease.

Customer Segments: The most important customers are governments and the corresponding departments. In some instances, the products may be for private companies as well.

Cost Structure: Developing the software tools, maintaining them, and providing the technical infrastructure will take up most of the cost.

Revenue Streams: The customers mainly pay for the licenses of the products for a certain period. After expiry, it can be renewed.

8.4.4 Customers and Partners

We observe that customers of Providers in the Forensic Market do not hesitate to take responsibility for using such tools and admit their application. In contrast, clients of

companies offering spyware products deny their use in most cases. We assume the reason behind this is that the utilization of Forensic Tools is ethically acceptable since rare instances of misuse come to light. On the other hand, the potential abuse of spyware products is known in our society and will always be viewed skeptically. Hence, ethically not always justifiable. It may be why Cellebrite can display customer stories on their webpage and advertise its usage and advantages. One of them is described below.

Mexicos Jalisco State Forensic Investigators offer their expert opinions to the Jalisco state prosecutors. Indeed, applying Cellebrite products, the investigators analyze the devices they receive and try to restore as much information as possible. As a result, one noticed a substantial increase in utilization such that they seized digital evidence from nine times as many devices as they handled in 2019.The investigators mostly make use of the solution UFED 4PC (see 8.4.2) [65].

Customers in the Digital Spyware Market are not known publicly. However, many articles indicate their presence [7]. Taking the software Pegasus as an example and the revelation by Forbidden Stories [56], we learned that the customers range from autocratic regimes (e.g., Saudia Arabia, Morocco) to countries with democracy (e.g., India, Mexico) and from European countries as Hungary and Azerbaijan to Rwanda and Togo in Africa [56].

Figure 8.10: Forbidden Stories [56]

Observing Figure 8.10, we note that these countries may be using Pegasus, or there are journalists in this country who are of interest to other countries.
In Figure 8.11, we see a summarization of suspected customers sorted by the number of vendors (not only limited to NSO). The complete table can be found on [66]. We note that most of them are non-western countries.
Suppose one now focuses on what type of people are targeted by these customers. In that case, we can divide them into four groups: (a) Activists&Dissidents, (b) Political Figures, (c) Journalists, and (d) Lawyers (see Figure 8.12). We observe that persons in category (b) are mainly located outside Western Countries. The same pattern applies to (c). However, we note nearly as many Western countries as others for (a) and (d). The reason may be that Activists and Dissidents primarily act from abroad because they fear consequences in their home country and have to live in exile. In contrast, journalists and political figures live mainly in their homeland and are active from there.

Suspected End-User	EIU Ranking	Freedom of the Net Ranking	Vendors	No. of Vendors
UAE	Authoritarian	Not Free	Candiru, Circles, DarkMatter, FinFisher GmbH, Hacking Team, NSO Group	6
Mexico	Flawed Democracy	Partly Free	Ability Inc., Circles, FinFisher GmbH, Hacking Team, NSO Group	5
Morocco	Hybrid Regime	Partly Free	Circles, FinFisher GmbH, Hacking Team, NSO Group	4
Indonesia	Flawed Democracy	Partly Free	Candiru, Circles, FinFisher GmbH, Verint Systems	4
Saudi Arabia	Authoritarian	Not Free	Candiru, FinFisher GmbH, NSO Group, Undisclosed	4
Nigeria	Hybrid Regime	Partly Free	Circles, Elbit Systems, FinFisher GmbH, Hacking Team	4
Vietnam	Authoritarian	Not Free	Ability Inc., Circles, Ocean Lotus, Strategic Cyber	4
Colombia	Flawed Democracy	Partly Free	Hacking Team, Mollitiam Industries, Verint Systems	3
Ethiopia	Authoritarian	Not Free	Cyberbit, FinFisher GmbH, Hacking Team	3
Guatemala	Hybrid Regime	-	Circles, Hacking Team, NSO Group	3
Honduras	Hybrid Regime	-	Circles, Hacking Team, NSO Group	3
Israel	Flawed Democracy	-	Ability Inc., Candiru, Circles	3
Peru	Flawed Democracy	-	Circles, Mollitiam Industries, Verint Systems	3

Figure 8.11: Overview of some suspected customers [45]

Type of Target	Target Location
Activists & Dissidents	Australia, Canada, Ecuador, Egypt, Germany, Hong Kong, India, Iran, Kazakhstan, Mexico, Morocco, Nigeria, Pakistan, Philippines, Qatar, Saudi Arabia, Spain, The Netherlands, Turkey, U.S., UAE, Uganda, UK, Uzbekistan, Vietnam
Political Figures	ASEAN Group, Belgium, Cambodia, Ecuador, India, Iran, Kazakhstan, Laos, Lebanon, Mexico, Nigeria, Pakistan, Panama, Philippines, Qatar, Saudi Arabia, South Africa, Spain, Togo, Turkey, U.S., UAE, Uganda
Journalists	Cambodia, Chile, Colombia, India, Kazakhstan, Lebanon, Mexico, Morocco, Pakistan, Qatar, Saudi Arabia, U.S., UAE, Uganda, UK, Uzbekistan, Vietnam
Lawyers	Belgium, Ecuador, France, Iceland, India, Israel, Kenya, Mexico, Morocco, Nigeria, Norway, Switzerland, U.S., UK

Figure 8.12: Overview of targets [66]

8.5 The Current Market and Challenges

The Digital Espionage and Forensics market is undergoing a significant change right now. For that reason, we will describe the current situation on the market from an economic point of view and describe observable trends and challenges one faces.
Regarding transparency, we have already seen a difference between the Market for Digital Espionage and Digital Forensics. This continues to the economic aspects as well. As a consequence, we will consider them separately.

8.5.1 The Market of Digital Forensic

The Digital Forensics Market was predicted to have a significant growth in revenues from 2017 with a market value of USD 4.15 billion to 2022 with a market value of USD 9.68 billion. The increment is measured in the Compound Annual Growth Rate (CAGR).

CAGR is the relative growth of a factor in a period of time or the mean growth if one considers several periods. This yields the following formula:

$$CAGR(t_0, t) = \left(\frac{A(t)}{A(t_0)}\right)^{\frac{1}{n}} - 1$$

where t_0 is the starting point and t the endpoint of the period, then $n = t_0 - t$. $A(t)$ is the value of the factor at time t and accordingly $A(t_0)$ at time t_0.

Hence, the market was forecasted to grow at a CAGR of 15.9% - the reasons for this vary. The massive increase in the use of IoT devices has undoubtedly given the whole development a boost. By 2022 there will be 18 billion such devices that will be applied in a range of areas such as autonomous driving, smart home, smart grids, and many more. However, the increasing use attracts criminals, which is why IoT devices are strongly affected by cyberattacks [67].
We see in Figure 8.13 that North America holds the largest share of the market, followed by Europe, Asia-Pacific, Middle-East Africa, and Latin America. This distribution seems to be constant over the years. The reason for North America holding the most significant share appears to be clear if we consider that many of the essential participants in this market are located there and that most of the organizations from there adapted to cloud technologies early. This has further accelerated the number of cyberattacks. Therefore several governments in this region are taking steps to strengthen their digital forensics sector [67].

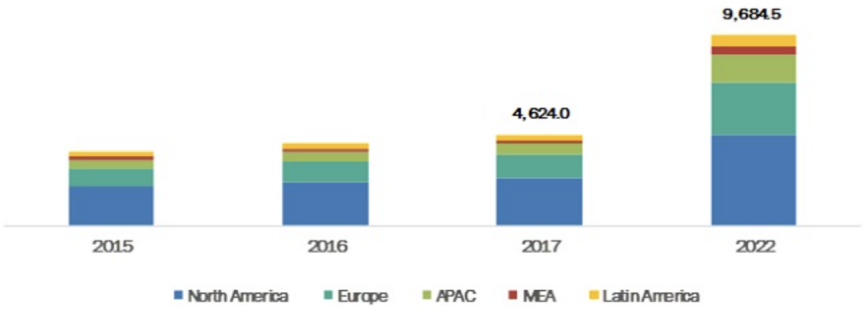

Figure 8.13: Digital Forensics Market, by Region [67]

8.5.2 The Market of Digital Espionage

Very few resources have analyzed the market itself and provided numbers because they struggle with transparency. As a result, the public depends on the firms to publish their numbers (e.g., revenues, market shares) on their own.
To have a good reference point, we take NSO, which confirmed that its revenue in 2018 was USD 250 million. To estimate the value of NSO, one of the more prominent companies in this market, we consider that a majority stack of them was bought in February 2019 by Novalpina Capital, a British private-equity firm founded just two years before the transaction. The investment was USD 1 bio. However, it is hard to estimate the total value of the market. Danna Ingleton from Amnesty International assumes that the market's worth is at least several billion dollars [7] [68]. In Figure 8.14, we see a list of the key

players in this market, the country they are located in, and their suspected clients. They are sorted in ascending order according to the number of clients.

Vendor	Vendor Country	Suspected End-Users	No. of Clients
Gamma Group / FinFisher GmbH	Germany & UK	Angola, Bangladesh, Belgium, Bosnia and Herzegovina, Czech Republic, Egypt, Ethiopia, Gabon, Indonesia, Italy, Jordan, Kazakhstan, Kenya, Lebanon, Malaysia, Mexico, Mongolia, Morocco, Nigeria, North Macedonia, Oman, Paraguay, Romania, Saudi Arabia, Serbia, Slovenia, South Africa, Spain, Taiwan, Turkey, Turkmenistan, UAE, Uganda, Venezuela	34
Circles	Israel	Australia, Belgium, Botswana, Chile, Denmark, Ecuador, El Salvador, Equatorial Guinea, Estonia, Guatemala, Honduras, Indonesia, Israel, Kenya, Malaysia, Mexico, Morocco, Nigeria, Peru, Serbia, Thailand, UAE, Vietnam, Zambia, Zimbabwe	25
NSO Group	Israel	Bahrain, Croatia, GCC Region, Great Lakes region of Africa, Guatemala, Honduras, Hungary, India, Kazakhstan, Latvia, Mexico, Morocco, Mozambique, Panama, Poland, Rwanda, Saudi Arabia, Spain, Switzerland, Togo, UAE, Uzbekistan, Zambia	23
Hacking Team	Italy	Brazil, Chile, Colombia, Ecuador, Egypt, Ethiopia, Guatemala, Honduras, Italy, Mexico, Morocco, Nigeria, Panama, Thailand, UAE	15
Ability Inc.	Israel	China, Czech Republic, Germany, Israel, Mexico, Myanmar, Singapore, Switzerland, Vietnam	9

Figure 8.14: Largest companies based on the number of suspected customers [45]

The market was created when companies had the idea to provide governments, which cannot develop software products for spyware in-house, with their solutions. This industry had already existed for a while but had a boost in 2013 when Edward Snowden revealed the surveillance capabilities of American agencies. As of now, the regulations which exist are the same laws as for the sale of weapons (see Wassenaar-Agreement [69]). These, however, seem to be less effective [7]. That is why the US took action and banned the sale of hacking tools for specific regimes [70]. David Kaye, UN's special rapporteur on freedom of opinion and expression, described the Market for Digital Espionage as out of control and not countable [7].

8.5.3 The Consequences

The impact the existence of this market has is serious. Many of the products offered have been misused in different ways, which resulted in bad events such as the killing of journalists, spying of human rights activists, silencing of political opponents, and so on. Further, a product as Pegasus, which governments may only use as of now, has a vast potential for misuse if it gets in the hands of criminals or corporates. This may accelerate cyber espionage in the different industry sectors. Consider the Figure 8.15.
We note, however, that the resistance to such spyware products has increased as well. For example, encrypted web pages have grown drastically over the past years. While in 2014 in the US, approximately 40% of the web pages loaded in Google Chrome have been encrypted, and it is in 2019 roughly 90%, which is shown in Figure 8.16.
Laws and regulations are not up to date with such products. Therefore, either one fails to present evidence of harm, or it is not traceable back to the defendant. It can be not very easy for the judges to understand what is being presented. Recently, an American with the pseudonym *MrKidane* claimed that the Ethiopian government was spying on him. The court closed the case saying that the spying did not happen entirely within America's border [7].

8.6 Application Scenarios

8.6.1 Flight Data Analysis in Aircraft Accident Investigation

To facilitate the investigation of aviation accidents and incidents, flight recorders are installed in the most crash survivable part of commercial flights. There are two types of flight recorders, flight data recorder (FDR) and cockpit voice recorder (CVR) [73].

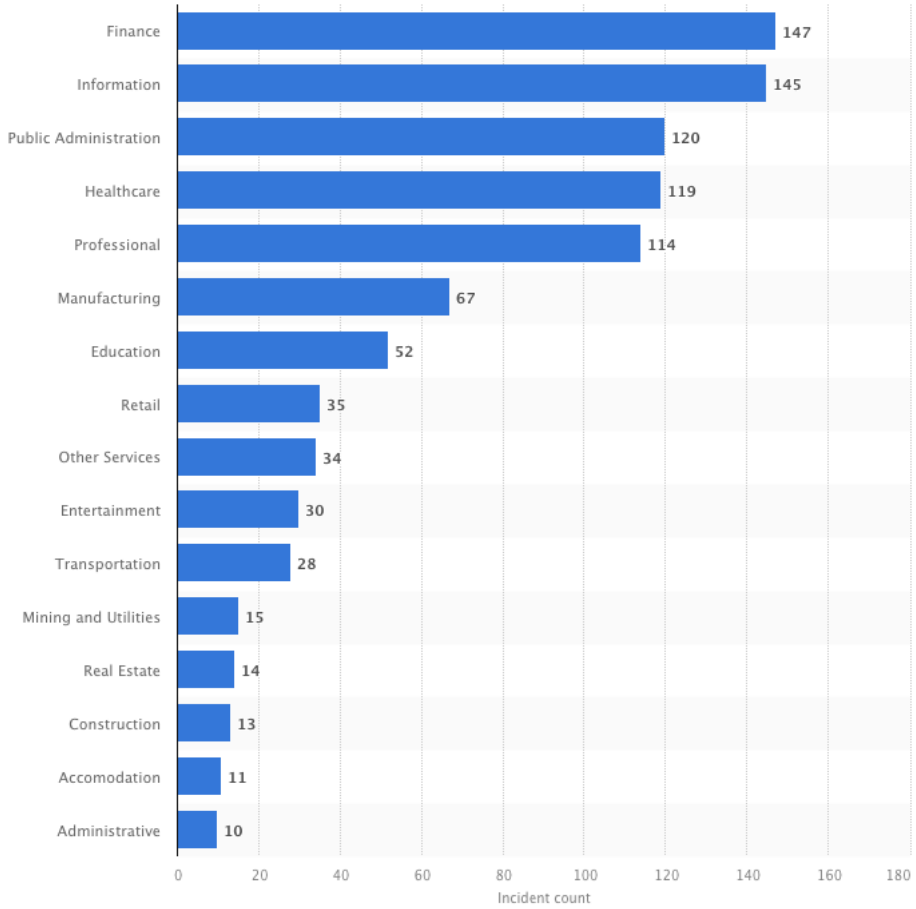

Figure 8.15: Sectors most targeted by cyber espionage in 2020 [71]

Figure 8.16: The number of encrypted web pages over the years [7]

FDR is used to record specific aircraft performance parameters. By regulation, newly manufactured FDR must monitor at least eighty-eight important parameters such as time, altitude, airspeed, heading, and aircraft attitude [73]. CVR is used to record the audio environment in the flight deck for accidents and incident investigation purposes. It records and stores the audio signals of the microphones and earphones of the pilots' headsets and of an area microphone installed in the cockpit [74].

Earlier FDR and CVR used analog magnetic tape to record the data. Magnetic tape works like any tape recorder. Magnetic-tape FDRs have the potential to record up to 100 parameters, and magnetic-tape CVRs can store at least the last 30 minutes of sound [75]. Partially damaged tape can still be read, and the result will be a data stream with holes in it [76].

The most recent FDR and CVR use solid-state memory and digital recording techniques. The solid-state uses stacked arrays of memory chips, so they do not have moving parts, which makes them much more resistant to shock, vibration, and moisture [74]. Compared with magnetic-tape recorders, solid-state FDR can track more parameters, and solid-state CVR can record more audio. For instance, in the Boeing 787, the FDR unit can log a whopping 146,000 parameters [75] and the CVR unit can record 120 minutes of audio.

The recording inspection should always be performed by specialists. Analysts should be not only familiar with the system and analysis tools but also able to identify issues from the recordings. For instance, for the CVR recordings, investigators should first retrieve the data from recorders. This extraction process is similar to digital forensics on a solid-state drive. The next step is identifying interference, then figuring out the normal (acceptable) interferences and issues. They should also deal with poor recording quality when there is a box malfunction. Figure 8.17 shows an example of having a mechanical interference occur over several minutes then disappear. This kind of interference should be detected and be further analyzed [77]. The interference in Figure 8.18 may be tolerated as the 400 Hz signal that comes from the cabin system during the use of cabin interphone is empirically acceptable.

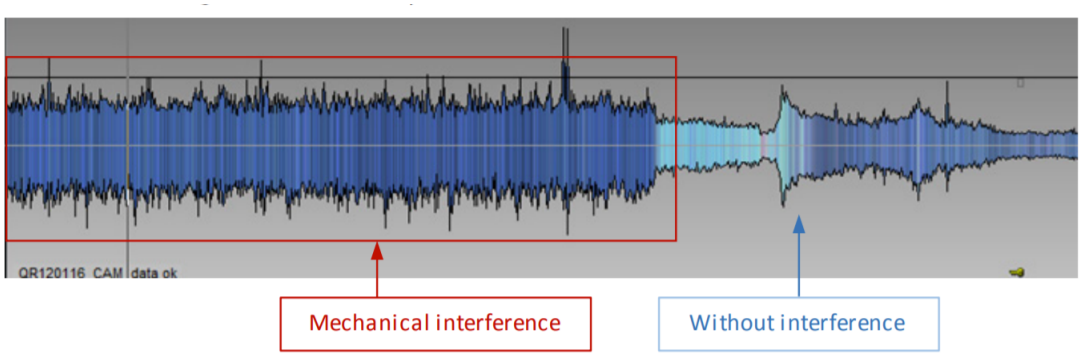

Figure 8.17: Audio level change in cruise flight phase (audio level versus time) [77]

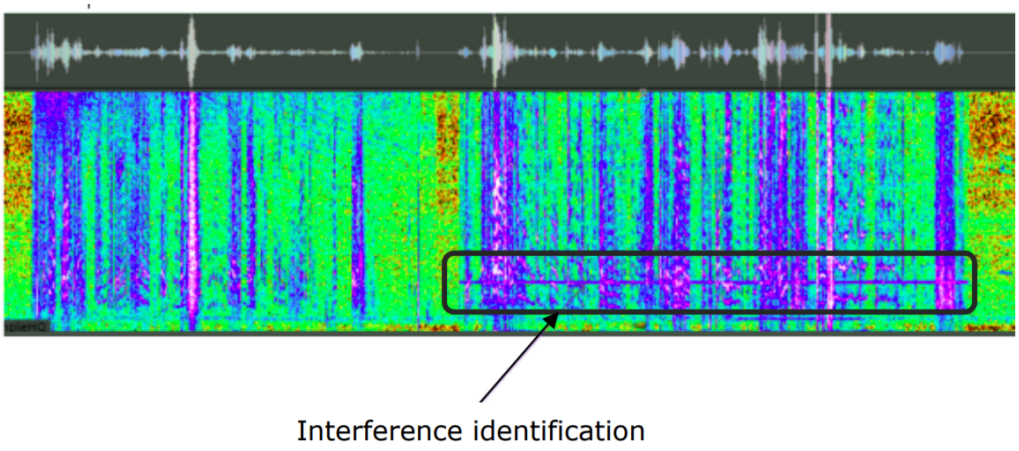

Figure 8.18: Audio waveform and Frequency data versus time [77]

8.6.2 Forensics on Espionage

In the research of finding out the suspected infections of Pegasus spyware, researchers in Citizen Lab first fingerprinted the behavior of the obtained exploit links and operator's command and control (C&C) servers to identify other servers associated with the same spyware system, including perhaps the servers of other operators if multiple operators use the same spyware system. Afterward, they performed DNS cache probing to generate a list of countries with possible infections associated with the operator. [21]
Fingerprinting is a tracking technology, which involves collecting the characteristics of a device, then assembling them into a profile that helps identify users [78]. In 2016, the research group built fingerprints and scanned the decoy pages from Pegasus servers. They published the research report in August 2016, then NSO Group upgraded its system, so there were no more decoys. After studying the behavior of several suspected new Pegasus servers, new fingerprints adapted to the change. The research result is in Figure 8.19, which shows the number of operation servers over time. The blue line is the version before adaptation, and the red line is the version after that.
DNS resolvers typically have a cache of recently resolved domain names shared among all the resolver's users to improve the performance. However, this shared state exposes a side-channel by which a resolver user can figure out if another user has issued a query for a specific domain name. This side channel can be exploited by a process called DNS cache probing [79]. The researcher DNS cache probed for all domains they linked to NSO Group's infrastructure that were active and matching the fingerprints and identified 45 countries with suspected Pegasus spyware infections operated by at least 33 likely NSO customers (as shown in Figure 8.20). They also found that the cross-border surveillance with Pegasus was widespread.

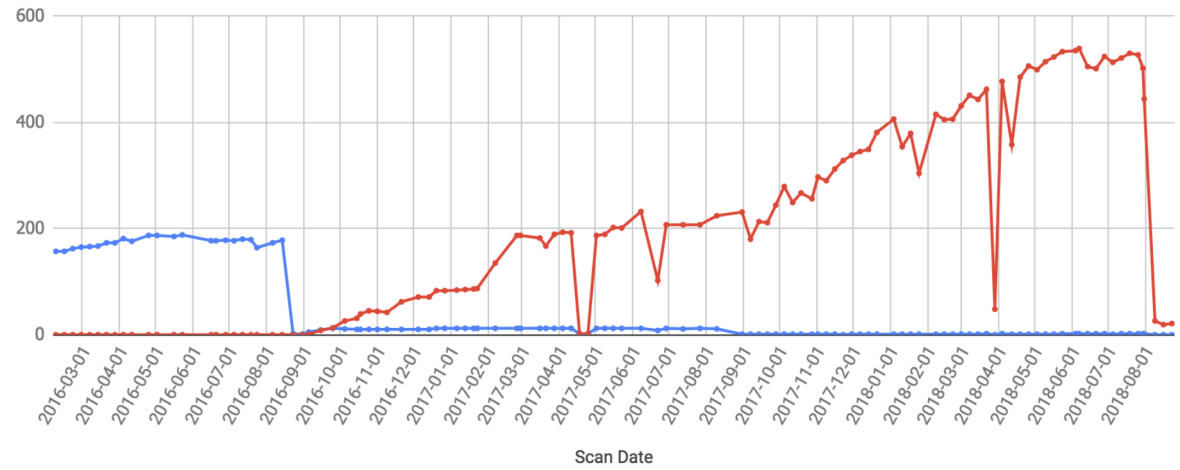

Figure 8.19: Pegasus servers available over time [21]

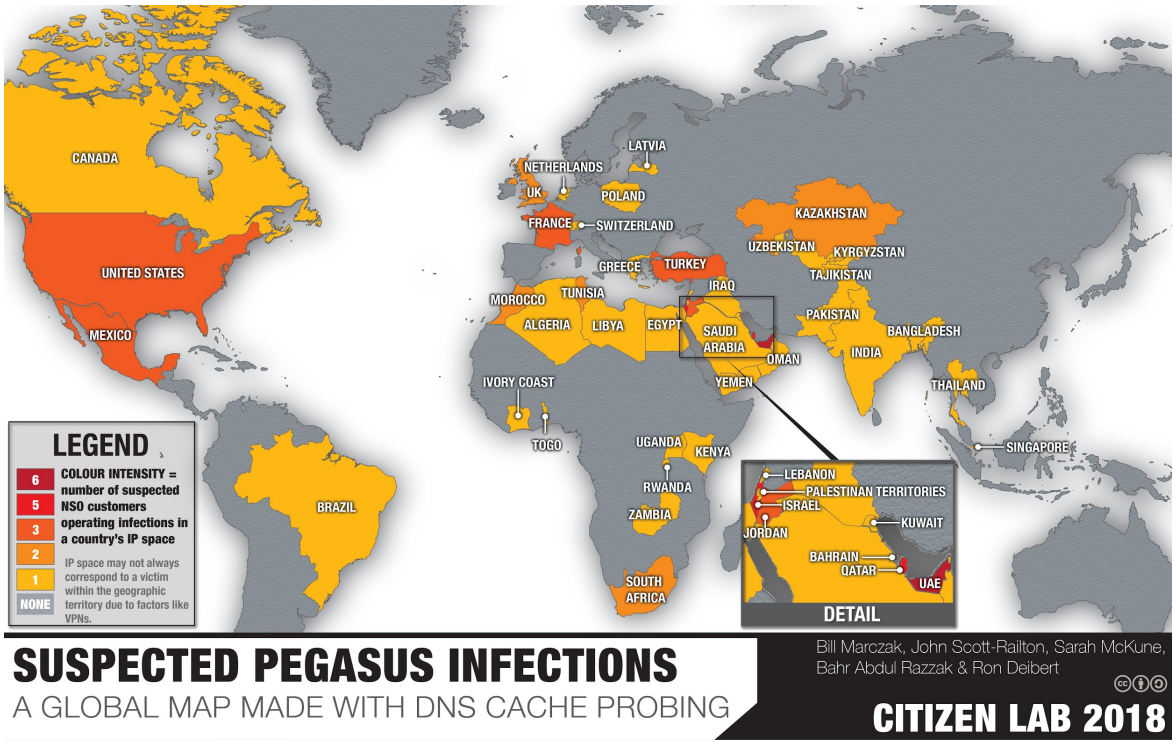

Figure 8.20: Global map of suspected NSO Pegasus infections [21]

8.6.3 EncroChat Hack

Crime organizations are using electronic devices such as smartphones for their communication. To prevent them from being bugged by the police, they rely on encrypted tools for sending messages or making phone calls. Some companies provide exactly such phones that are secure and anonymous.

EncroChat was that kind of a firm that advertised with end-to-end encryption and a guarantee of anonymity. As a result, messaging between two persons would be equivalent to their conversation in an empty room. Moreover, the servers in offshore datacentres would never decode or save encryption keys, messages, and user data. Interested customers could not simply purchase the devices in an electronic shop, but you had to have specific contacts. The phones are android devices (mostly "BQ Auqaris X2") modified through EncroChat by installing its encrypted communication applications that routed all the messages via their servers and removing the GPS, camera, and microphone hardware. A specific function enabled the user to delete all the data on the phone instantly. The phone, including a SIM card for worldwide usage for six months, cost EUR 2'500 [80].

Since 2017 the french police force, Gendarmerie Nationale, has been investigating against EncroChat because the phones provided by the company have been seized in more and more criminal cases. Even though EncroChat claimed that their servers are Offshore, the police forces have tracked them down in Lille, France. The investigators hacked themselves into these servers and sent an update to as many EncroPhones as possible. However, this update contained malware that infected nearly the whole EncroChat community. Likely, the encryption of the messages was not hacked, but the investigators received copies of the sent messages, usernames, and passwords. Additionally, they had them send a list of Wifi signals near the EncroPhone. As a result, the police forces were now able to locate the users. Later, the French Police forces shared this information with Europol. Investigators around Europe could now read live the chats of countless EncroChat users. The company realized to a later point what was going on and informed its users [80] [81]. Figure 8.21 shows a message sent by the EncroChat team to their users.

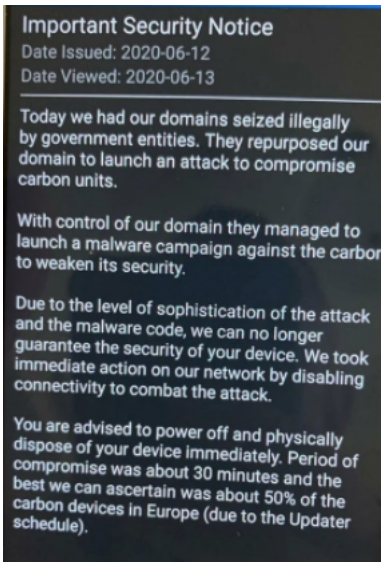

Figure 8.21: Message from EncroChat to its user [82]

This operation is seen as one of the biggest infiltrations by law enforcement agencies. In live-time, investigators analyzed over 100 million encrypted messages. The results were arrests in Great Britain, Norway, Sweden, France, and the Netherlands, with cases about international drug business, operating drug laboratories, murder, blackmailing, and hostage-takings.

8.7 Discussion

Technology Trends

The technological evolution comes with the advancement of cyber-crime, which continually creates new kinds of threats, attacks, tools, and techniques that allow offenders to penetrate more complicated or well-controlled environments, cause increased harm, and even remain untraceable. Due to this, various emerging fields, such as Internet-of-Things (IoT) forensics and cloud computing forensics, were researched and might be the trends in digital forensics.

IoT devices in critical infrastructures bring reliability and convenience to consumers and open a new world of opportunity for intruders and introduce a whole set of unique and complicated questions to digital forensics.

Wearable devices and technology fall under this IoT category, and the data stored on them, such as heart rate, geolocation data, and sleep data, can help investigators solve an investigation from the perspective of both the victim and the offender. MacDermott et al. tried to extract data from 3 fitness bands [83]. They found that data could be retrieved from paired Bluetooth devices (i.e., smartphones) with traditional digital forensics, but it was extremely challenging to retrieve data from wearable devices' flash memory. Williams et al. extended the research on the latest devices, but all the works were only on parent applications linked to the devices [84].

Although IoT data could be a rich source of evidence, forensics professionals cope with diverse problems, from the huge variety of IoT devices and non-standard formats to the multi-tenant cloud infrastructure and the resulting multi-jurisdictional litigations. A further challenge is end-to-end encryption which represents a trade-off between users' right to privacy and the success of the forensics investigation. [85]. Only in 2017, there was a 600 percent increase in attacks against IoT devices [86], which requires special attention on IoT forensics.

Cloud computing is another trend of digital forensics. It has become more prevalent in recent years and is being used to support multiple areas of human life. There are several benefits in switching to cloud infrastructures, such as reduced IT cost [87], scalability, and access to automatic updates. With the promise of unlimited, reliable, and always-available storage, many private and confidential data are now stored on different cloud platforms. Being such a gold mine of data, cloud platforms are among the most valuable targets for attackers. Therefore, many forensics investigators have tried to develop tools, tactics, and procedures to collect, preserve, analyze and report pieces of evidence of attackers' activities on different cloud platforms [88]. There is already an increasing focus on cloud forensics, but due to the architectural complexity of the cloud, cloud forensics is more complex than the typical computing [89].

Market Trends

From 2021 to 2026, the Digital Forensics Market is expected to grow at a CAGR of 10.97% with North America holding the largest share [72]. This rate is compared to the period (2017-2022) lower (see 8.5.1) but still on a high level. Further, we expect that the hype around cryptocurrencies will benefit the Market for Digital Forensics. Since more people are involved with such currencies, fraud cases will also rise. Hence, appropriate digital forensic tools will be required to restore essential data [67]. However, we think that for forensic tools and spyware products, the existing regulations will tighten, as seen in the US case, which will slow down the market's growth.

8.8 Summary and Conclusions

This report introduces the Digital Espionage and Forensics Market by viewing different perspectives as the technical and economic ones, followed by three application scenarios. Common digital forensic targets are introduced, such as storage devices, mobile devices, and networks. The digital forensic techniques applied to those targets are also included. Furthermore, the forensic techniques and tools of mobile devices are introduced and discussed, with the comparison of logical and physical acquisition and the comparison of logical and physical extraction. This report also presents some data extracted from smartphones and summarizes the recovered artifacts from several applications. Lastly, the Discussion Section explains the possible emerging new edges of digital forensics. In conclusion, even if digital evidence can be obtained using existing techniques and tools, investigators should keep up with technological advances.

We learned about the Digital Forensics Market's targets and tools and realized their difficulties. For example, the growing market for encrypted web pages is a challenge and pushes them to newer solutions. On the other hand, the Digital Espionage Market has been facing negative reviews for a long time since we assume that the activities of key players are seen more with a skeptical view in our society.

We are very curious about what the future of this market brings us. The regulations seem not to be strict enough to be effective, but we note a few changes regarding that. As an example, the US banned espionage products. The regulations for this market and the question of if or how the market could be regulated effectively will be worthy of another report.

Acknowledgements We are grateful to our supervisor Muriel Franco for his fruitful comments, corrections, and inspiration.

Bibliography

[1] Sule, D. (2021). *Forensic Readiness for Enhanced eDiscovery.* In *Handbook of Research on Cyber Crime and Information Privacy* (pp. 236-255). IGI Global.

[2] Reith, M., Carr, C., Gunsch, G. (2002). *An examination of digital forensic models.* International Journal of Digital Evidence, 1(3), 1-12.

[3] Pew Research Center. (2021, April 7). *Mobile Fact Sheet.* Pew Research Center: Internet, Science & Tech. Retrieved November 13, 2021, from `https://www.pewresearch.org/internet/fact-sheet/mobile/`

[4] Eichbaum, J. (2019, September 9). *Five continual challenges with smartphone forensics.* MSAB. Retrieved November 13, 2021, from `https://www.msab.com/blog/five-continual-challenges-with-smartphone-forensics/`

[5] Sinha, S. (2021, September 30). *State of IoT 2021: Number of connected IoT devices growing 9% to 12.3 billion globally, cellular IoT now surpassing 2 billion.* IoT Analytics. Retrieved November 17, 2021, from `https://iot-analytics.com/number-connected-iot-devices/`

[6] Norwich University. (2020, November 3). *How Nations Use Digital Espionage Against Each Other.* Norwich University Online. Retrieved November 13, 2021, from `https://online.norwich.edu/academic-programs/resources/how-nations-use-digital-espionage-against-each-other`

[7] The Economist. (2019, December 12). *Offering software for snooping to governments is a booming business.* The Economist. Retrieved November 13, 2021, from `https://www.economist.com/business/2019/12/12/offering-software-for-snooping-to-governments-is-a-booming-business`

[8] Casey, E. (2004). *Digital Evidence and Computer Crime, Second Edition.* Elsevier. ISBN 0-12-163104-4.

[9] Techopedia. (2017, January 12). *Digital Forensics.* Techopedia.Com. Retrieved November 13, 2021, from `https://www.techopedia.com/definition/27805/digital-forensics`

[10] Patterson, D. A., & Hennessy, J. L. (2012). *Computer organization and design: the hardware/software interface, (Rev. ed. of: Computer organization and design/John L. Hennessy, David A. Patterson. 1998.).*

[11] Geier, F. (2015). *The differences between SSD and HDD technology regarding forensic investigations.* p. i

[12] Moulton, S. (2006), *Hard Drive Recovery Part 3 at Toorcon* YouTube. Retrieved November 8, 2021, from `https://www.youtube.com/watch?v=Cayzw1iThjM`

[13] Moulton, S. (2011). *Solid State Drives Destroy Forensics & Data Recovery Jobs*. Las Vegas.

[14] Mutemwa, M., & Mouton, F. (2018, March). *Cyber security threats and mitigation techniques for multifunctional devices* In 2018 Conference on Information Communications Technology and Society (ICTAS) (pp. 1-6). IEEE.

[15] Al Sharif, S., Al Ali, M., Salem, N., Iqbal, F., El Barachi, M., & Alfandi, O. (2014, March). *An approach for the validation of file recovery functions in digital forensics' software tools*. In 2014 6th International Conference on New Technologies, Mobility and Security (NTMS) (pp. 1-6). IEEE. doi:10.4225/75/58a54cc3c64a2

[16] Gupta, K.P., & Nisbet, A. (2016). *Memory forensic data recovery utilising RAM cooling methods*. In Valli, C. (Ed.). (2016). The Proceedings of 14th Australian Digital Forensics Conference, 5-6 December 2016, Edith Cowan University, Perth, Australia. (pp. 11-16).

[17] Halderman, J. Alex, et al. *Lest we remember: cold-boot attacks on encryption keys*. Communications of the ACM 52.5 (2009): 91-98.

[18] Sira, R. (2003). *Network forensics analysis tools: an overview of an emerging technology*. GSEC, version, 1, 1-10.

[19] Sikos, L. F. (2020). *Packet analysis for network forensics: A comprehensive survey. Forensic Science International: Digital Investigation*, Volume 32, pp.200892, ISSN 2666-2817.

[20] Boot, M. (2018, December 5). *An Israeli tech firm is selling spy software to dictators, betraying the country's ideals*. Washington Post. Retrieved November 4, 2021, from `https://www.washingtonpost.com/opinions/2018/12/05/israel-is-selling-spy-software-dictators-betraying-its-own-ideals/`

[21] Marczak, B. (2020, May 8). *HIDE AND SEEK: Tracking NSO Group's Pegasus Spyware to Operations in 45 Countries*. The Citizen Lab. Retrieved November 6, 2021, from `https://bit.ly/3GtdYmv`

[22] Casey, E. (2004). *Digital Evidence and Computer Crime, Second Edition*. Elsevier. ISBN 978-0-12-163104-8.

[23] Jansen, W., Delaitre, A., & Moenner, L. (2008, January). *Overcoming impediments to cell phone forensics*. In Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008) (pp. 483-483). IEEE.

[24] Burnette, M. W. (2002). *Forensic Examination of a RIM (BlackBerry) Wireless Device* June, 2002.

[25] Grand, J. (2002, June). *pdd: memory imaging and forensic analysis of palm OS devices*. In Proceedings of the 14th Annual FIRST Conference on Computer Security Incident Handling and Response.

[26] Mellars, B. (2004). *Forensic examination of mobile phones*. Digital Investigation, 1(4), 266-272.

[27] Jansen, W., Delaitre, A., & Moenner, L. (2008, January). *Overcoming impediments to cell phone forensics*. In Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008) (pp. 483-483). IEEE.

[28] Statista. (2021, August). *Smartphone users worldwide 2016-2021*. Retrieved November 2021, from https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

[29] Perez, S. (2020, October). *Consumers spent a record $28 billion in apps in Q3, aided by pandemic*. TechCrunch. Retrieved December 2020, from https://techcrunch.com/2020/10/08/consumers-spent-record-28-billion-in-apps-in-q3-aided-by-pandemic/

[30] *Mobile Operating System Market Share Worldwide | Statcounter Global Stats.* StatCounter Global Stats. Retrieved November 2021, from https://gs.statcounter.com/os-market-share/mobile/worldwide

[31] Statcounter. (2018, September). *Mobile Operating System Market Share Worldwide*. StatCounter Global Stats. Retrieved November 2021, from https://gs.statcounter.com/os-market-share/mobile/worldwide

[32] Shimmi, S. S., Dorai, G., Karabiyik, U., & Aggarwal, S. (2020, June). *Analysis of iOS SQLite schema evolution for updating forensic data extraction tools*. In 2020 8th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-7). IEEE.

[33] Sammons, J. (2015). Chapter 10 - Mobile device forensics. In J. Sammons (Red), *The Basics of Digital Forensics (Second Edition)* (Second Edition, pp. 145-161). doi:10.1016/B978-0-12-801635-0.00010-3

[34] Daniel, L. E., & Daniel, L. E. (2012). Chapter 37 - Cell Phones. In L. E. Daniel & L. E. Daniel (Reds), *Digital Forensics for Legal Professionals (pp. 263-272)*. doi:10.1016/B978-1-59749-643-8.00037-7

[35] Vijayan, V. (2012). *Android Forensic Capability and Evaluation of Extraction Tools*.

[36] Ayers, R., Brothers, S., & Jansen, W. (2014, May 15). *Guidelines on Mobile Device Forensics*. NIST Special Publication, 800, 101. Retrieved November 9, 2021, from https://www.nist.gov/publications/guidelines-mobile-device-forensics

[37] UKEssays. (November 2018). *Challenges of Android and iOS Forensics*. Retrieved November 9, from https://www.ukessays.com/essays/information-technology/challenges-of-android-and-ios-forensics.php?vref=1

[38] Shaikh, H. (2017, July 21). *Practical android phone forensics*. Infosec Resources. Retrieved November 9, from https://resources.infosecinstitute.com/topic/practical-android-phone-forensics/#gref

[39] Mahalik, H., Tamma, R., & Bommisetty, S. (2016). *Practical mobile forensics*. Packt Publishing Ltd.

[40] Mikhaylov, I. (n.d.). *Android forensic analysis with Autopsy*, Digital Forensics Corp. Retrieved November 9, 2021, from https://www.digitalforensics.com/blog/android-forensic-analysis-with-autopsy

[41] Lwin, H. H., Aung, W. P., & Lin, K. K. (2020, February). Comparative analysis of Android mobile forensics tools. In 2020 IEEE Conference on Computer Applications (ICCA) (pp. 1-6). IEEE.

[42] Mikhaylov, I., Skulkin, O., Shorokhov, I. (n.d.). *MOBILE FORENSICS: UFED VS MAGNET ACQUIRE*, Digital Forensics Corp. Retrieved November 9, 2021, from https://www.digitalforensics.com/blog/mobile-forensics-ufed-vs-magnet-acquire/

[43] Belkasoft. (n.d.). *10 Reasons To Use Belkasoft Evidence Center* . Retrieved November 13, 2021, from https://belkasoft.com/10-reasons-to-use

[44] Salamh, F. E., Mirza, M. M., Hutchinson, S., Yoon, Y. H., & Karabiyik, U. (2021). *What's on the Horizon? An In-Depth Forensic Analysis of Android and iOS Applications*. IEEE Access, 9, 99421-99454.

[45] ODonnell, C., & Woodhams, S. (2021, May 12). *The Global Spyware Market Index*. TOP10VPN. Retrieved November 13, 2021 from https://www.top10vpn.com/research/global-spyware-market-index/

[46] NSO Group. (n.d.). *NSO Group - About us*. Nsogroup. Retrieved November 17, 2021 from https://www.nsogroup.com/about-us/

[47] Kirchgaessner, S. (2021, July 19). *Saudis behind NSO spyware attack on Jamal Khashoggis family, leak suggests.* The Guardian. Retrieved November 9, 2021 from https://bit.ly/3Dw2psV

[48] Hubbard, B. (2021, November 3). *I Was Hacked. The Spyware Used Against Me Makes Us All Vulnerable.* The New York Times. Retrieved November 8, 2021 from https://www.nytimes.com/2021/10/24/insider/hacking-nso-surveillance.html

[49] NSO Group. (2019, September). *Human Rights Policy.* Retrieved November 6, 2021 from https://www.nsogroup.com/wp-content/uploads/2019/09/NSO-Human-Rights-Policy_September19.pdf

[50] NSO Group. (2019d, September). *Transparency Statement of Principles.* Retrieved November 6, 2021 from https://www.nsogroup.com/wp-content/uploads/2019/09/Transparency-Statement-of-Principles_September19.pdf

[51] NSO Group. (2019b, September). *Internal Whistleblowing Policy.* Retrieved November 6, 2021 from https://www.nsogroup.com/wp-content/uploads/2019/09/Internal-Whistleblowing-Policy_September19.pdf

[52] NSO Group. (2019a, September). *External Whistleblowing Policy.* Retrieved November 6, 2021 from https://www.nsogroup.com/wp-content/uploads/2019/09/External-Whistleblowing-Policy_September19.pdf

[53] United Nations Human Rights. (2011). *Guiding Principles on Business and Human Rights. United Nations*. Retrieved November 6, 2021 from https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

[54] NSO Group. (2021b, June). *Transparency and Responsibility Report.* Retrieved November 6, 2021 from https://www.nsogroup.com/wp-content/uploads/2021/06/ReportBooklet.pdf

[55] Haskell-Dowland, P., & Musotto, R. (2021, July 20). *How does the Pegasus spyware work, and is my phone at risk?* The Conversation. Retrieved November 7, 2021 from https://theconversation.com/how-does-the-pegasus-spyware-work-and-is-my-phone-at-risk-164781

[56] ForbiddenStories. (2021, July 18). *Pegasus: The new global weapon for silencing journalists* Forbidden Stories. Retrieved November 6, 2021 from https://forbiddenstories.org/pegasus-the-new-global-weapon-for-silencing-journalists/

[57] BBC News. (2021, November 3). *NSO Group: Israeli spyware company added to US trade blacklist.* Retrieved November 6, 2021 from https://www.bbc.com/news/technology-59149651

[58] Cellebrite. (n.d.). *About.* Retrieved November 6, 2021 from https://www.cellebrite.com/en/about/

[59] Cellebrite Global HR: *Code of Business Conduct and Ethics*, Cellebrite, August 2021. https://investors.cellebrite.com/static-files/51fd1f5a-9ffc-402a-bedb-cf35308c0af3

[60] Cellebrite - The Board of Directors: *Corporate Governance Guidelines*, Cellebrite, August 2021. https://investors.cellebrite.com/static-files/15291c32-b352-48a2-883d-0304984c98b7

[61] Cellebrite *Whistleblower Policy*, Cellebrite, August 2021. https://investors.cellebrite.com/static-files/ec861a60-3cf4-42eb-a2d0-98a3fc511e42

[62] Cellebrite. (2021). Cellebrite. (2021). *Cellebrite UFED | Access and Collect Mobile Device Data.* Retrieved November 7, 2021 from https://www.cellebrite.com/en/ufed/

[63] Cellebrite. (2021). *Gain visibility to cloud-based evidence to solve your cases faster.* Retrieved November 7, 2021 from https://cf-media.cellebrite.com/wp-content/uploads/2020/05/ProductOverview_Cellebrite_UFEDCloud_web.pdf

[64] Cellebrite. (2021b). *Triage and acquire forensic images from Windows and macOS computers.* Retrieved November 7, 2021 from https://cf-media.cellebrite.com/wp-content/uploads/2021/01/ProductOverview_DigitalCollector_A4_2020_web-2.pdf

[65] Cellebrite. (2021, February 28). *Supporting the Prosecutors Office Is Their Priority: Mexicos Jalisco State Forensic Investigators Manage Nine-fold Increase in Digital Evidence With Solutions.* Retrieved November 11, 2021 from https://bit.ly/3rMaoQl

[66] The Global Spyware Market Index. (22.07.2021). [Dataset]. top10vpn. Retrieved November 11, 2021 from https://docs.google.com/spreadsheets/d/1FHIX71XHi4U5sX8_SqebekUkMrgSKmqKoQNbANFqMlc/edit#gid=0

[67] MarketsandMarkets. (n.d.). *Digital Forensics Market.* Retrieved November 11, 2021 from https://www.marketsandmarkets.com/Market-Reports/digital-forensics-market-230663168.html

[68] Borger, S. (2021, August 5). *Besitzer der Spionagesoftware Pegasus sind mit Selbstzerfleischung beschäftigt.* DER STANDARD. Retrieved November 11, 2021 from https://bit.ly/3rKFecj

[69] Seco, S. F. W. (2021, April 30). *Die Vereinbarung von Wassenaar (WA).* SECO admin. Retrieved November 11, 2021 from https://www.seco.admin.ch/seco/de/home/Aussenwirtschaftspolitik_Wirtschaftliche_Zusammenarbeit/

Wirtschaftsbeziehungen / exportkontrollen - und - sanktionen / exportkontrollpolitik/die-vereinbarung-von-wassenaar--wa-.html

[70] Roth, E. (2021, October 22). *New US rules on spyware exports try to limit surveillance tech like Pegasus.* The Verge. Retrieved November 12, 2021 from https://www.theverge.com/2021/10/22/22740155/commerce-departments-new-rule-hacking-tools-china-russia

[71] Statista. (2021, November 11). *Cyber espionage: most-targeted industries 2020.* Retrieved November 12, 2021 from https://www.statista.com/statistics/221293/cyber-crime-target-industries/

[72] Yahoo Finance. *Global Digital Forensics Market - Growth, Trends, COVID-19 Impact, and Forecasts (2021 - 2026).* (2021, November 3). Retrieved November 12, 2021 from https://finance.yahoo.com/news/global-digital-forensics-market-growth-125100253.html

[73] SKYbrary. (2021, September 25). *Flight Data Recorder (FDR).* SKYbrary Aviation Safety. Retrieved November 17, 2021, from https://skybrary.aero/articles/flight-data-recorder-fdr

[74] SKYbrary. (2021, May 26). *Cockpit Voice Recorder (CVR).* SKYbrary Aviation Safety. Retrieved November 17, 2021, from https://skybrary.aero/articles/cockpit-voice-recorder-cvr

[75] Chandler, N., & Bonsor, K. (2020, June 30). *How Black Boxes Work. HowStuffWorks.* Retrieved November 17, 2021, from https://science.howstuffworks.com/transport/flight/modern/black-box.htm

[76] George Cramoisi, Editor, & George Cramoisi, E. (2012). *AIR CRASH INVESTIGATIONS: MYSTERIOUS CRASH KILLS 25 The Crash of United Airlines Flight 585.* Lulu.com.

[77] Air Accidents Investigation Branch. (2018, October 8). *Guidance on CVR recording inspections.* GOV.UK. Retrieved November 17, 2021, from https://www.gov.uk/government/publications/guidance-on-cvr-recording-inspections

[78] Mozilla. (n.d.). *What is fingerprinting and why you should block it.* Retrieved November 17, 2021, from https://www.mozilla.org/en-US/firefox/features/block-fingerprinting/

[79] Niaki, A. (2021, June 22). *Cache me outside: A new look at DNS cache probing.* APNIC Blog. Retrieved November 17, 2021, from https://blog.apnic.net/2021/06/22/cache-me-outside-dns-cache-probing/

[80] Simplicissimus. (2021, February 17). *EncroChat: Das WhatsApp der Verbrecher.* YouTube. Retrieved November 17, 2021 from https://www.youtube.com/watch?v=pgyu4sY-61I&t=608s&ab_channel=Simplicissimus

[81] Cox, J. (2020, July 3). *Die Leute sind gefickt: Wie die Polizei heimlich ein Handynetzwerk für Drogengangs infiltrierte.* VICE. Retrieved November 17, 2021 from https://www.vice.com/de/article/3aza95/encrochat-hack-wie-die-polizei-ein-handynetzwerk-fur-drogengangs-infiltrierte

[82] Omerta. (2020, June 16). *EncroChat hacked, users exposed & arrests galore - the King is dead.* Omerta Digital Technologies. Retrieved November 17, 2021 from https://omertadigital.com/blogs/news/encrochat-hacked-users-exposed-arrests-galore-the-king-is-dead

[83] A. MacDermott, S. Lea, F. Iqbal, I. Idowu and B. Shah, *Forensic Analysis of Wearable Devices: Fitbit, Garmin and HETP Watches, 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, 2019, pp. 1-6, doi: 10.1109/NTMS.2019.8763834.

[84] J. Williams, A. MacDermott, K. Stamp and F. Iqbal, *Forensic Analysis of Fitbit Versa: Android vs iOS, 2021 IEEE Security and Privacy Workshops (SPW)*, 2021, pp. 318-326, doi: 10.1109/SPW53761.2021.00052.

[85] Stoyanova, M., Nikoloudakis, Y., Panagiotakis, S., Pallis, E., & Markakis, E. K. (2020). *A Survey on the Internet of Things (IoT) Forensics: Challenges, Approaches, and Open Issues.* IEEE Communications Surveys Tutorials, 22(2), 1191-1221. doi:10.1109/COMST.2019.2962586

[86] Symantec. (2018, March). *Symantec 2018 Internet Security Threat Report.* Retrieved November 2021, from https://docs.broadcom.com/doc/istr-23-2018-en

[87] Euractiv. (2014, June 2). *Cloud computing: Leveraging the digital economy.* Euractiv - Cloud Computing Report. Retrieved November 13, 2021, from https://www.euractiv.com/section/digital/linksdossier/cloud-computing-leveraging-the-digital-economy/

[88] Baldwin, J., Alhawi, O. M. K., Shaughnessy, S., Akinbi, A., & Dehghantanha, A. (2018). *Emerging from the Cloud: A Bibliometric Analysis of Cloud Forensics Studies.* In A. Dehghantanha, M. Conti, & T. Dargahi (Reds), Cyber Threat Intelligence (bll 311-331). doi:10.1007/978-3-319-73951-9_16

[89] Manral, B., Somani, G., Choo, K.-K. R., Conti, M., & Gaur, M. S. (2019). *A Systematic Survey on Cloud Forensics Challenges, Solutions, and Future Directions.* ACM Comput. Surv., 52(6). doi:10.1145/3361216

Chapter 9

Data Breaches and the GDPR: Direct Costs, Regulations, and Fines

Kai Zinnhardt, Guanda Zhao

This report focuses on data breaches and the General Data Protection Regulation (GDPR) and its economic effects. The concept of data breach is introduced on its causes and the protective measures of each cause is discussed. The GDPR is introduced in details on its definition and important terms and compared with previous data protection regulation on the differences and improvement. A particular focus of this paper was taken on how the companies experience the result of the data breaches and the GDPR. Five case studies related to data breaches and five case studies related to top five GDPR fines were discussed. It was found, that the costs of a data breach are comparatively more significant for small companies than for large organizations. Additionally, the size anomaly exists both for data breaches and the for the implementation costs of the GDPR. Furthermore, the GDPR is compared with the Brazilian General Personal Data Protection Law: Lei Geral de Proteção de Dados Pessoais (LGPD), it was noticed that the GDPR is stricter in several regards, such as in their Fines, the notification policy, and the legal bases for data processing.

Contents

9.1 Introduction

The amount of digital data stored and processed is increasing exponentially. Now, digital users create roughly 2.5 quintillion bytes of data everyday [2]. These data includes sensitive information, such as photos, chat history, emails, bank transfers, and confidential files. With the increasing amount of data, people care more about their privacy: whether the companies are legally using their data (e.g. Facebook selling the data to Cambridge Analytica [1]) and whether their data is protected, for example, from data breach.
Data breach poses a significant cybersecurity risk to companies with substantial financial impact (*e.g.,* fines) and individuals, as private data (*e.g.,* passwords, identity numbers, passport information) is compromised and leaked [8]. Such incidents have been increasing over the past years, in which companies reported an increasing number of data breaches, from January 2019 to April 2020, of 54% compared to the same period in 2018 [9]. The aggregated daily rate of breach notifications in Europe experienced double digit growth for the second year running with 331 notifications per day since 28 January 2020, a 19% increase compared to 278 breach notifications per day for the previous year [5].
Moreover, since the formal application of the European General Data Protection Regulation (GDPR) in 2018, more than 280,000 data breaches have been noticed, with a sum of over EUR 272.5 million in fines [5]. In 2021, data breach costs rose from $3.86 million to $4.24 million, which is the highest average total cost in the 17-year history [7]. These fines help companies to be more vigilant and invest in protecting the data of their clients and users. However, not only do fines directly impact the finances of a company after a data breach, but also its reputation is impacted, as these breaches and fines are tracked and detailed [3]. Therefore, investing the financial impact of such GDPR fines and how companies deal with such incidents presents an interesting research topic.
This paper presents and discusses the core GDPR regulations concerning data breaches, survey recent data breaches incidents, and detail which are the economic factors that such data breaches led (*e.g.,* direct downtime impact, reputation losses, and GDPR fines). Also, it explores the regulations regarding data breaches in other countries and compare those with GDPR.

9.2 Background

This section provides an overview of a what a data breach entitles in Section 9.2.1 and details the most relevant European General Data Protection Regulation (GDPR) articles to the paper's context in Section 9.2.2.

9.2.1 Data Breach

A data breach is an incident where information is stolen or taken from a system without the knowledge or authorization of the system's owner [10]. These data involves confidential information [11] of organizations, which refers to design materials, procedures, formulas, production processes, production methods, management know-how, customer lists, supply information, production and marketing strategy. They are not known to the public, and bring economic benefits to the organizations; hence, are protected by the organizations. In constrast privacy data of individuals, which refers to personally identifiable information [12] including name, ID number, contact information, address, and financial information including credit and debit card information, bank details, property status.
There are many causes of data breach, these include [6]

1. **Criminal hacking**: attackers hack into organization's system and database.

2. **Social engineering**: phishing attackers get access to customers' or employees' accounts.

3. **Weak security**: Malware passes through the firewall or through the application's vulnerabilities and injects into customers' or employees' accounts.

4. **Stolen devices**: customers' or employees' devices get stolen, attackers get direct access to customers' sensitive data or organization' s system.

5. **Weak credential**: Customers or employees using weak password that can be easily brute forced.

6. **Insider**: A "malicious" employee leaks the data.

Data breach incidents cause great consequences. Various customer information will be used illegally, leading to misappropriation, unauthorized charge of credit card and phishing. Further, the organization's reputation will be damaged: the organization's current customers are likely to stop using the services after the data breach. Potential new customers are likely to avoid the organization, because of data breach, the organization may receive negative reports. The organization's confidential information may be leaked to its competitors, thus affecting the future development of the organization.
Top 10 biggest data breaches in history record can be found in Table 9.1.

Table 9.1: Top 10 biggest data breaches in history record [4]

Company	Impact	Date
CAM4	10.88 billion records	March 2020
Yahoo	3 billion accounts	October 2017
Aadhaar	1.1 billion people	March 2018
First American Financial Corp.	885 million users	May 2019
Verifications.io	763 million users	February 2019
LinkedIn	700 million users	June 2021
Facebook	533 million users	April 2019
Yahoo	500 million accounts	2014
Starwood (Marriott)	500 million guests	November 2018
Adult Friend Finder	412.2 million accounts	October 2016

9.2.2 General Data Protection Regulation (GDPR)

The GDPR is the regulation regarding data protection from the European Union (EU) implemented in 2018 [13]. The GDPR aims to protect the personal data of EU citizens. Thereby, the regulations of the GDPR have a global reach, affecting all organizations that store, process, or otherwise use data from the EU citizens [14]. As this is an EU-Regulation, it applies to all countries throughout the EU-Zone, *e.g.,* Germany, France, and Spain. As an effect of this, the EU-Member states do not have to create their own regulations. Rather, they can implement the EU regulations. Therefore, the GDPR harmonizes the data protection regulations throughout the EU.
There are five main privacy and data protection requirements that the GDPR tries to enforce. Firstly, requiring the consent of subjects for data processing [15]. Therefore, if data was collected for one specific topic, this data can only be used for that specific cause. If the subject is fulfilled, the data cannot be used anymore. Secondly, the data has to be collected in an anonymous or pseudonymous way, to protect privacy [15]. This

should ensure that the data can not be referred to a natural person. Thirdly, the GDPR should provide notification on a data breach incident [15]. In this way, the EU tries to respond in a fast and efficient manner. Fourthly, to safely handle the transfer of data across borders so that data can not be intercepted [15]. Lastly, the GDPR could require certain companies to appoint a data protection officer to oversee GDPR compliance [15]. This ensures the privacy of the EU data in non-EU member states.
To have a full understanding of the GDPR regulations, five terms are defined:

1. **Personal Data**: is any information relating to an identified or identifiable natural person. An identifiable natural person can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, Art. 4 No. 1 GDPR. A directly identifiable person would be, when a natural person can be identified directly from the data. Whereas indirectly would mean using another source of data and identifying the person through specific characteristics.

2. **Processing**: any operation or set of operations that are performed on personal data or sets of personal data, whether or not by automated means, Art. 4 No. 2 GDPR. Therefore, any type of operation that includes the use of data can be considered as processing. For example, storing, structuring, or deleting data can be considered as processing.

3. **Controller**: is a natural or legal person, public authority, agency, or other body that, alone or jointly with others, determines the purposes and means of the processing of personal data, Art 4 No. 7 GDPR. This is the body that decides inside of the organization why and how personal data will be processed [16].

4. **Processor**: means a natural or legal person, public authority, agency, or other body which processes personal data on behalf of the controller, Art 4 No.8. The processor therefore only acts when the controller decides to process the data at an external organization [17].

5. **Personal data breach**: means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored, or otherwise processed, Art 4 No.12. This definition does not require intent for a data breach. Therefore it is not of importance how or why the data breach takes place [17]. As a consequence only the fact of the data breach is relevant.

The GDPR regulates how the controller should act in case of a data breach. Regarding Art. 33 sec. 1 of the GDPR: "*in the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority competent under Article 55, unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons.*" Hereby, the regulation refers to data breaches, where personal data is affected. Consequently, the notification is only mandatory if the personal data breached can be misused in such a way, that it harms the natural person. As quoted from this article the supervisory authority has to receive this notification 72 hours after the data breach was noticed by the organization. If this notification was not within 72 hours, Art. 33 sec. 1 further states, that "*it shall be accompanied by reasons for the delay.*"
Art. 33 sec. 1 only accounts for personal data breaches when notified by the controller. However, as the controller can also delegate the processing of personal data to a processor, personal data breaches can occur by the processor. Therefore, when personal data

breaches are noticed by the processor, the processor shall notify the controller without undue delay after becoming aware of a personal data breach, as stated in Art. 33 sec. 2. Additionally, Art. 33 sec. 3 formulates a set of minimum requirements for the notification of the data breach. The notification must contain the following:

1. Description of the nature of the personal data breach. If possible, notify the categories and the approximate number of data subjects and data records concerned.
2. Communicate the name and contact details of the data protection officer or other contact points.
3. Describe the likely consequences of the personal data breach.
4. Describe the measures taken or proposed to be taken by the controller to address the personal data breach.

The GDPR has two Tiers of Fines. The first tier is the **Standard Maximum Fine**, and on the second tier, is the **Higher Maximum Fine** for more severe cases. Both tiers correspond to the infringement of GDPR in the following cases, corresponding to Art. 83(2).
The Standard Maximum Fine, which corresponds to less severe infringements, can have fines of up to 10 million Euros, or a penalty of 2% of the company's worldwide annual revenue, whereby the higher figure will be taken, according to Article 83(4) of the GDPR. The infringements for this fine includes one of the following [18]:

- Integrating data protection 'by design and by default'.
- Recording or processing of activities.
- Co-operating with the supervising authority.
- The security in place for the processing of data.
- Communicating with supervisory authorities and data subjects where there is a personal data breach.
- Prior consultation with the appropriate authorities before processing commences.
- The appointment and tasks allocated to the Data Protection Officer.
- Certification completed ensuring GDPR compliance.

The Higher Maximum Fine is used, when infringements relate to the following principles: To the right of data privacy and the right to be forgotten. If this happens, then it is considered to have violated the fundamentals of the GDPR [18]. In this situation, penalties can increase to a maximum of 20 million Euros, or 4% of the previous financial year's worldwide annual revenue, whereby the higher of the two criteria will be taken. An overview of the mentioned Regulations can be found in Table 9.2.

9.3 Data Breaches and the GDPR

This section provides and overview of Incidents 9.3.1 where data breaches occured, with the corresponding Fines 9.3.2. Additionally, this section covers other incidents where the GDPR was applied 9.3.4.

Table 9.2: GDPR Regulations Overview

Article	Section	Main Point
4	1	Any information relating to an identified or identifiable natural person.
4	2	Any operation which is performed on personal data.
4	7	Determines the purposes and means of the processing of personal data.
4	8	Processes personal data on behalf of the controller.
4	12	A breach of security leadig to unathorised discolsure of personal data.
33	1	Controller has 72 hours to notify the data breach incident.
83	4	Fines of up to 10 million Euro or up to 2% of the total worldwide annual revenue.
83	5	Fines of up to 20 million Euro or up to 4% of the total worldwide annual revenue.

9.3.1 Incidents

Five incidents of data breaches were selected. For the five incident a wide spectrum of Organisations were chosen. Thereby, two multi-national cooperation's (from which one, Marriott International, suffered two data breaches), a national Agency, and a smaller national company were selected. For the five chosen cases, one can see the spectrum between breaches that could have been prevented easily and ones that are harder to prevent. On all of the stated examples, cyberattacks date back no further than 2014. This is due to the recent enforcement of the GDPR.

i British Airways

In 2018, one of the most known data breach attacks in recent European history was when attackers breached data from British Airways. From August 21st to September 5th, British Airways got hit by a Cyber-attack. During this Cyber-attack, the attackers could breach 380,000 records from the booking transaction site [19]. These records contained critical information about the customer's credit cards, thereby containing information of the credit card number, expiration date as well as the Card Verification Value (CVV) codes [19].
The attack was a web-based skimming attack [20]. Thereby, both the website and the mobile app for Android devices were affected [20]. The Android mobile app is based upon the website, whereby it shares the Javascript [20]. As the attackers only covered these two areas, and neither the databases nor servers were attacked, the RiskIQ suspects the Magecart group behind the attack [21]. The Magecart is a hacking organisation that has been active since 2015 [20]. Then, the Magecart web-based card skimmers were first detected in 2016 [21]. The affected script is connected to the British Airways baggage claim information page. The attackers modified the script to include code that is often used in clandestine manipulations. Thereby, the customer data was sent to an attacker-controlled server when a user clicked the submission button [20].

ii Marriott International

The Marriott Hotel group had two data breaches in recent history. The first one data breach was noticed in 2018 and the second in 2020. The first data breach started with a suspicious flag on September 8th in 2018. After an investigation in November, they found

that the breached data included 383 million guest records [26]. The breach also included 30 million EU residents. In this breach, personal data such as guests' names, addresses, passport numbers, and payment information was exposed [26].
In this data breach, the database was compromised. The attack originated in Starwood Group's reservations system in 2014. In 2016 Marriott acquired Starwood, and the compromise was not detected [26]. However, in 2018 Marriott had not migrated the two reservation systems yet, and therefore had to maintain two systems simultaneously [25]. The compromise was found after a security tool flagged, which is an unusual database query. Thereby the database query was made by a user, who, after analysis, could not have to build the query; hence another person took control over the account [25]. Additionally, they found a Remote Access Trojan and MimiKatz, a tool for sniffing out username/-password combinations in system memory, during the investigation. The Trojan has most likely been downloaded during a phishing attack [25].
The second data breach is believed to have started in mid-January 2020, and Marriott issued an incident notification on the 31st of March 2020. During this breach, an estimated 5.2 Million guests may have been exposed [23]. Thereby the data affected by the breach included contact detail, loyalty account information, additional personal details, partnerships and affiliations, and guests' preferences [23].
During this attack, it seems that the credentials of two employees at a franchise property were used [24]. Hence, it is not known if the credentials have been stolen in a phishing, social engineering scam or if the employees were actively involved in the attack [24].

iii Morele.net

Morele.net, which is all round online store with roots in Poland, got hit by a Cyber-Attack in November 2018. Morele.net first noticed the attack after customers reported suspicious SMS messages informing them that they needed to make an extra payment to complete the order. After investigation, it turned out that Morele.net got hit by a scam attack [27]. As investigations suggests, Morele.net could have prevented the attack from happening if penetration testing had been conducted [27]. Penetration Testing is designed to identify weaknesses in an organization's systems and exploiting them [27].
As a consequence of this scamming attack, 2.5 million data records were stolen. These records contained unique email addresses, phone numbers, names, and passwords stored as md5crypt hashes [28]. From which 35,000 customers further Personal Data got breached. Those records include payment installment information (including Personal ID number), education, source of income and net income, household maintenance costs, and marital status [27].

iv National Revenue Agency Bulgaria

Not only companies are in the focus of data breach attackers, government institutions such as the Bulgarian National Revenue Agency (KDBM) are also a exposed to attacks frequently. During this cyberattack, the Servers of the KDBM were hacked [29]. However, it is believed that the cyberattack could have been avoided, or at least reduced in severity, due to deficient security practices and technical infrastructure [29]. Furthermore, the KDBM violated European law by not informing affected people about which data had been compromised [29].
The data breach of the KDBM affected 5 million people. This included data about the names, contact details, and tax information [26].

9.3.2 Data Breach Costs

The data breach costs can vary from case to case. In this subsection, the fines of the data breach inducements discussed earlier will be examined. Such fines can vary due to the ranges described in Art. 83 sec. 2&4. The largest fine from the GDPR regarding data breaches was assigned to British Airways. British Airways received a penalty of 22 million euros after it got revised from 204.6 million euros. This reduction took place after other factors such as the COVID-19 situation were considered [22]. Additionally, this penalty also quotes Art 5 (1)f. This states: "*The personal data shall be processed in a manner that ensures appropriate security of the personal data*".
The Marriott Hotel group until now only got fined for the first data breach incident. This is as the second cyberattack is too recent, and therefore no fine was declared yet. The fine for the first incident accounted for 20 million euros [3]. The Morle.net fine is not as large, as the company has a smaller worldwide revenue stream as the previously mentioned co-operation. The fine amounts to 660 thousand euros [3]. The fourth case that was discussed is the Bulgarian National Revenue Agency. Thereby this institution got a fine of 2.6 million euros [3].

9.3.3 Summary of the incidents

An overview of the data breach incidents can be found in Table 9.3. All breaches occurred due to insufficient technical and organisational measures to ensure information security.

Table 9.3: Data Breaches and GDPR Fines [3]

Company	Country	Date of Decision	Fine [€]	Quoted GDPR
British Airways	UK	2020-10-16	22,046,000	Art. 32*
Marriott International	UK	2020-10-30	20,046,000	Art. 32
Morele.net	Poland	2019-09-10	660,000	Art. 32
National Revenue Agency	Bulgaria	2019-08-28	2,600,000	Art. 32

* Art. 5 (1)f)

9.3.4 GDPR and Other Fines

Top 5 GDPR fines were selected for discussion [3]. All these fines show that the data of ordinary users has been breached or exfiltrated.

i Amazon

On July 16 2021, the Luxembourg National Commission for Data Protection (CNDP) issued the biggest fine ever for the violation of the General Data Protection Regulation (GDPR) with a fine of €746 million ($888 million) to Amazon Europe Core S.a.r.l. for non-compliance with general data processing principles [30].
The fine was issued as a result of a complaint filed by 10,000 people against Amazon in May 2018, through a French privacy rights group that promotes and defends fundamental freedoms in the digital world - La Quadrature du Net [30]. The rights group said in a post that the "collective complaint" against Amazon was filed by 10,000 people and the ruling now indicates their stand that "the advertising targeting system imposed by Amazon is carried out without our free consent, in violation of the GDPR" [31].
The CNPD opened an investigation into how Amazon processes personal data of its customers and found infringements regarding Amazons' advertising targeting system that was carried out without proper consent [30].

ii WhatsApp

On September 2 2021, Ireland's data protection authority Data Privacy Commission (DPC) issued the second biggest fine to Facebook-owned instant messaging and Voice-over-IP (VoIP) service WhatsApp Ireland in the amount of €225 million ($267 million) after a three-year investigation [32] for insufficient fulfilment of information obligations which violates Art. 5 (1) a) GDPR, Art. 12 GDPR, Art. 13 GDPR and Art. 14 GDPR [3]. The fine relates to an investigation which began in 2018, about whether WhatsApp had been transparent enough about how it handles information. The issues involved were highly technical, including whether WhatsApp supplied enough information to users about how their data was processed and if its privacy policies were clear enough [33].

iii Google

On January 21 2019, the French National Commission on Informatics and Liberty (CNIL) issued the thrid biggest fine to Google LLC in the amount of €50 million for insufficient legal basis for data processing [36] which violates Art. 13 GDPR, Art. 14 GDPR, Art. 6 GDPR and Art. 5 GDPR [3].
Google failed to provide enough information to users about consent policies and did not give them enough control over how their personal data is processed and it was fined because of lack of transparency, inadequate information, lack of valid consent regarding the ads personalization [36].

iv H&M

On October 1 2020, the Hamburg Commissioner for Data Protection and Freedom of Information (BfDI) issued the fourth biggest fine to Swedish retail conglomerate Hennes & Mauritz - mostly known as H&M, registered in Hamburg, in the amount of €35.3 million ($41.5 million) for insufficient legal basis for data processing [34] which violates Art. 5 GDPR, Art. 6 GDPR [3].
After a technical error, the data on the company's network drive was accessible to everyone in the company for a few hours. With regards to the definition of a "personal data breach", with incidents can be regarded as an internal data breach. The press picked up the news making the Commissioner aware of the violation [36].
The H&M group currently has around 5,000 stores in 74 markets and according to their website 179,000 employees [34]. It collected sensitive personal data of their employees through whispering campaigns, gossip, and other sources to create profiles of employees and used that data in the employment process. The personal data included medical records including diagnoses and symptoms of the illness and private details about vacation and family affairs [36].

v TIM

On January 15, 2020, Italian Data Protection Authority (Garante) issued the fifth biggest GDPR fine to TIM (telecommunications operator) in the amount of €27,8 million for insufficient legal basis for data processing [35] which violates Art. 5 GDPR, Art. 6 GDPR, Art. 17 GDPR, Art. 21 GDPR and Art. 32 GDPR [3].
TIM was fined for aggressive marketing strategy, invalid collection of consents and excessive data retention period [35]. It failed to properly manage lists of data subjects who wanted to be excluded from commercial campaigns [35]. TIM has contacted non-customers multiple times (certain numbers over 150 times per month) without proper consent or other legal bases. Few million individuals were affected by their aggressive marketing strategy [36].

9.3.5 Summary of the incidents

An overview of the data breach incidents can be found in Table 9.4.

Table 9.4: Top 5 GDPR Fines [3]

Company	Country	Date of Decision	Fine [€]	Quoted GDPR	Type
Amazon Europe Core S.à.r.l.	Luxembourg	2021-07-16	746,000,000	Unknown	Non-compliance with general data processing principles
WhatsApp Ireland Ltd.	Ireland	2021-09-02	225,000,000	❶, ❸ , ❹, ❺	Insufficient fulfilment of information obligations
Google LLC	France	2019-01-21	50,000,000	❶, ❷, ❹, ❺	Insufficient legal basis for data processing
H&M Hennes & Mauritz	Germany	2020-10-01	35,258,708	❶, ❷	Insufficient legal basis for data processing
TIM	Italy	2020-01-15	27,800,000	❶, ❷, ❻, ❼, ❽	Insufficient legal basis for data processing

❶ Art. 5, ❷ Art. 6, ❸ Art. 12, ❹ Art. 13, ❺ Art. 14, ❻ Art. 17, ❼ Art. 21, ❽ Art. 32

9.4 Discussion

This section provides the discussion of why GDPR is important in Section 9.4.1, its impact on the companies in Section 9.4.2 and compare GDPR with the data protection regulation in Brazil: Lei Geral de Proteção de Dados (LGPD) in Section 9.4.3. Additionally, this section lists protective measures against the six causes of data breaches in Section 9.4.4.

9.4.1 Why GDPR is Important?

After four years of discussions in the European Union, the GDPR is officially effective since May 25, 2018. It is an important step to strengthen the security and protection of personal data. It is the world's most stringent law on data protection and privacy [40]. Further, the Data Protection Directive enacted in October 1995 by Europe Union [59], is being phases out and taken over by GDPR.
There are seven major differences between GDPR and previous data protection rules [41]:

1. Geographic reach and scope

   - Applicable to institutions established in the European Union (local EU companies and institutions).
   - Applicable to the processing of personal data of individuals in the European Union (for companies outside the European Union, their target users include users in the European Union).
   - Applicable to all walks of life, whether it is traditional industries such as banking, insurance, and aviation, or emerging fields such as e-commerce and social interaction, as long as it involves providing services to EU individuals and processing personal data [42].

   In other words, companies all over the world, no matter where the data is stored and processed, even if the business scope is not within the EU, as long as the company has any users from EU member states, they must comply with the GDPR.

   In addition, the supervisory authority of the country where the business owner is established will act as the leading supervisory authority to supervise all data activities of the enterprise, and its effectiveness will radiate throughout Europe.

2. Definition of personal data

   In terms of data protection, all parties in the data supply chain from the top to the bottom will be held accountable. In terms of obtaining and managing personal information, the GDPR puts forward new and stricter requirements, and gives individuals clear rights, which has brought a certain impact to the enterprise's user data management through manual, process, and technology. Personal data can be divided into two categories [43]:

   - Direct data: name, address, city, telephone, email, ID number and zip code.
   - Indirect data: cookies, IP address, Media Access Control (MAC) address and social account number.

   In addition, biological information such as iris, fingerprints, medical and health information, religious beliefs, political opinions, sexual orientation and sex life data, are all protected personal data.

3. Consent policies

   There must be a legitimate reason for processing the data. First, the user's permission must be obtained, and the user's permission must be specific and clear, and freely decided without knowing it. In the past, it was often adopted by enterprises: publishing a large number of obscure notices, and the practice of ticking users by default is no longer allowed. This legitimate reasoning can be seen in the Google case study (Section: iii). Secondly, under the premise of the user's permission, the data controller can use the user's personal data for marketing purposes, but the user can object at any time. Once the user objects, the data controller must immediately stop using it [44].

   Below is a list of data subjects' privacy rights [45]:

   1. The right to be informed
   2. The right of access
   3. The right to rectification
   4. The right to erasure
   5. The right to restrict processing
   6. The right to data portability
   7. The right to object
   8. Rights in relation to automated decision making and profiling.

   GDPR introduced two new types of rights: "the right to erasure" and "the right to data portability".

   The right to erasure: When the user withdraws his consent in accordance with the law or the data controller no longer has a legal reason to continue processing the data, the user has the right to request the deletion of the data. Moreover, the data controller shall not only delete its own data and copies, but also be responsible for its publicly disseminated data, and notify the third party to stop using and delete [46].

   The right to data portability: Users can transfer their personal data from one data controller to another without any problems. The data controller not only has no right to interfere, but also needs to cooperate with users to provide data text [47].

4. Data breach policies

   With the previous rules, businesses are under no obligation to report when data breaches occur, although they are encouraged to do so. This changes with the advent of GDPR, with any future breaches having to be reported to the relevant authorities within 72 hours of the incident [41].

5. Accountability

   Data Protection Officer: With expertise in data protection laws, this role must be independent, and its contact information must be published, reported to the supervisory authority, and directly reported to senior management [48].

   Document management: The data controller must fully document its data processing activities, including the purpose of data processing, the type of data, the type of data recipient, the data recipient transferred to a third party, the time of data storage, and the security measures taken and many more [48].

   The accountability mechanism of the above data controller can be understood as the data controller must take "sufficient measures" to ensure its data security.

6. Data protection governance

   The Data Protection Act does not stipulate how the governance of data security functions should be allocated, requiring only a basic commitment to the concept from management. GDPR will change this, as any company employing more than 250 people will be mandated to appoint a dedicated data protection officer, as will any firm processing more than 5,000 subject profiles annually [41].

7. Penalties and compensation [49]

   The less severe infringements could result in a fine of up to €10 million, or 2% of the firm's worldwide annual revenue from the preceding financial year, whichever amount is higher.

   The more serious infringements go against the very principles of the right to privacy and the right to be forgotten that are at the heart of the GDPR. These types of infringements could result in a fine of up to €20 million, or 4% of the firm's worldwide annual revenue from the preceding financial year, whichever amount is higher.

During the rapid development of the Internet in the past thirty years, personal information is being adopted on a large scale by various enterprises and organizations. Whether it is food delivery, online shopping, or public services, users' personal information has been collected on a large scale [50]. In recent years, there are countless cases of personal information leakage, so this problem is being faced by all countries in the world [4].
For example, there are over-precise advertising, people can see their favorite products on Amazon, and people have seen relevant recommendation on Facebook. In terms of the realization principle of advertisement matching, it is to write some cookies in the browser when the user visits the website. Amazon uses these cookies to display more accurate advertisements through its advertising platform.
Although in recent years people have become accustomed to this type of accurate advertising, but the operation behind it is illegal here in the GDPR.

9.4.2 Impact on the companies

This section is organized into two parts: Firstly, the direct and reputation costs will be discussed. In the second part, the cost of implementation to apply with the GDPR will get examined.

i Direct & Reputation Cost

The costs for data breaches will be discussed on a general notice, as no specific data to the data breach examples could be found.
Firstly, data breaches cost more for small businesses than large businesses when calculated on a per employee level. Data breaches for large organizations with more than 25,000 employees cost the enterprise €173 per employee. This amounts to a total of €4,33 million [37]. On the other side, the cost for small companies, with a size of 500-1,000 employees, has an average cost of €3,000 per employee. This calculates to a total cost of €2.24 million [37]. This indicates, that the impact for small companies is comparatively larger, therefore those organizations could have harder time to recover.
Whereas on average, in 2020, the cost for a data breach amounted to $3.89 million (approximately €3.3 million) and on an average cost per record level costing $146 (approximately €125) [51]. Consequently, the number of records per employee in a small business is higher than in a big business. Therefore the cost per employee is higher. Furthermore, this shows that the cost to protect against data breaches for smaller companies is

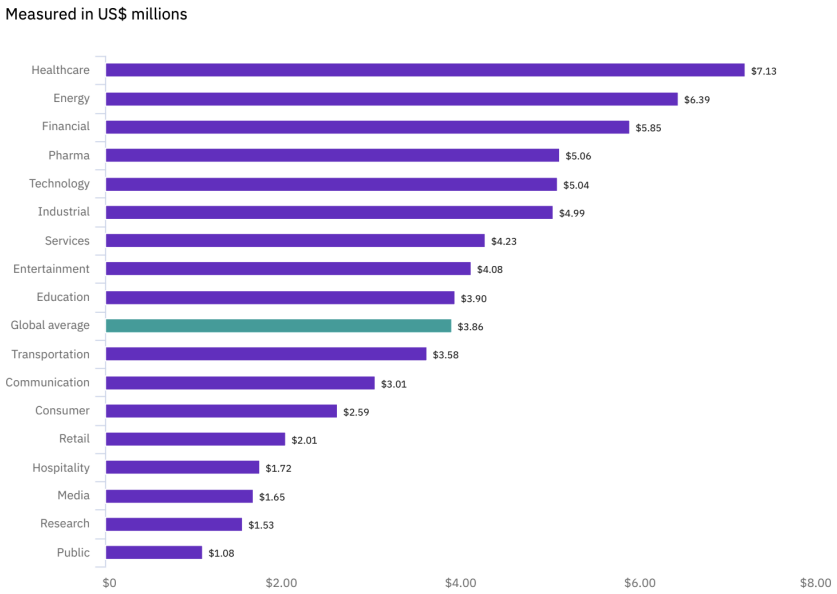

Figure 9.1: Average total cost of a data breach by industry [51]

comparatively larger. Small companies need a similar level of protection to increase the probability of not getting attacked in the first place.
Secondly, the cost of a data breach can spread over many years, as security researchers found out one-third of the costs span over more than one year. To be more precise, around 22% was incurred in the second year, while the remaining 11% accounted for more than two years [37]. Additionally, the researchers found that the rates increased for highly regulated sectors, such as the financial and healthcare services [37]. Consequently, due to the higher regulations and standards, the protection will get more expensive. Nevertheless, if a data breach occurs in one of these sectors, the overall cost increases. This correlation can also be seen in Figure 9.1 [51].
Figure 9.1 [51] shows that in sectors where companies work with sensitive data, hence high regulation standards (*e.g.,* healthcare, energy, or financial), the cost of a breach is higher. This can also be explained through a higher reputational loss and with it a higher reputation cost. Those sectors suffer a loss in customers of up to 1/3 [39]. Additionally, it is more challenging for organizations to acquire new customers, as it remains in the customers' heads for longer. This is as the uncertainty of the data is of higher priority in these sectors, where more sensitive data is being stored.
Thirdly, the findings of Obaydin, Xu, and Zurbruegg [38] suggest that the data breach Notification Laws, those that the GDPR also use, have unintended costs and side effects. As Obaydin & co. found out that the data breach notification laws can substantially affect managerial behavior and have a capital market effect [38]. The effect on the market could even go so far that it could lead to a crash for this particular stock. However, when analysing the above case for Marriott International, such a crash can not be seen. The Marriott stock react after the news, of a data breach came out, with a loss of 5.8%, suggesting a typical reaction to negative news. Therefore on a general notice, these notification rules can also contradict the intention. The regulations intend to provide a greater operational transparency [38]. This intention is fully filled through the notification. However, as the stock may crash as a consequence of the notification, it could happen that the company cannot get additional cash from the shareholders, which could be needed as a consequence of the breach. As a conclusion of knowing this, the paper [38] stated that the laws can incentivize managers to withhold other negative news, and therefore affecting the company financial disclosure practices [38].

ii Cost of Implementation

The cost of implementation is not just limited to the IT-security. Furthermore, the GDPR covers how the employees and the people in the organization act with the data an further security measures. Therefore, the GDPR requires employees to be trained so that they will be conscious about how they threaten the data and security standards [53].
In this section, the implementation costs from a size and sector perspective are described. For this analysis, the focus is on the companies listed in the FTSE 100 index (composition from 2020). Firstly, it will be focused on the size of the companies. For the factor of size, one will first look at a per-employee level. This can be seen in Figure 9.2 [52].

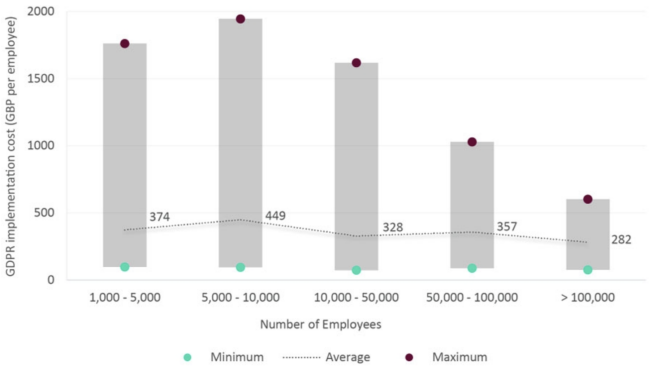

Figure 9.2: GDPR implementation cost by company size per employee [52]

Figure 9.2 shows that the size effect on a per employee level is such that smaller companies have to pay more than larger companies, on a general notice. This is due to the complexity of the Regulations, where one needs a specific basis before it can be scaled. However, if one looks at a general company perspective, one gets a second impression. This can be seen in Figure 9.3 [52].

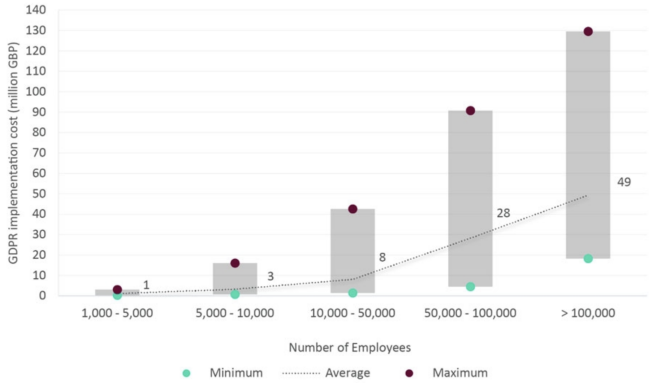

Figure 9.3: GDPR implementation cost by company size [52]

Figure 9.3 shows that the implementation for large companies is expensive on an absolute basis. As for cooperations, with more than 100,000 employees have to pay on average 49 million GBP (approximately €58 million). The combination of the views suggests that the scaling with the GDPR is limited. Even though IT security has a scaling potential, the cost to train employees cannot be neglected and is not vastly scaleable.
After focusing at the size factor for the cost of the implementation, one will now focus on the sectors. This industry viewpoint is essential for the implementation, as some sectors have to fulfill sector regulations more than others. Additionally, the sectors also define what type of data has to be stored, hence how secure it has to be. The implementation cost on a sector view can be seen in Figure 9.4 [52].
Figure 9.4 shows that banks have by far the highest cost of implementation; it is on average more than triple the amount of the second-highest sector. This is due to the high

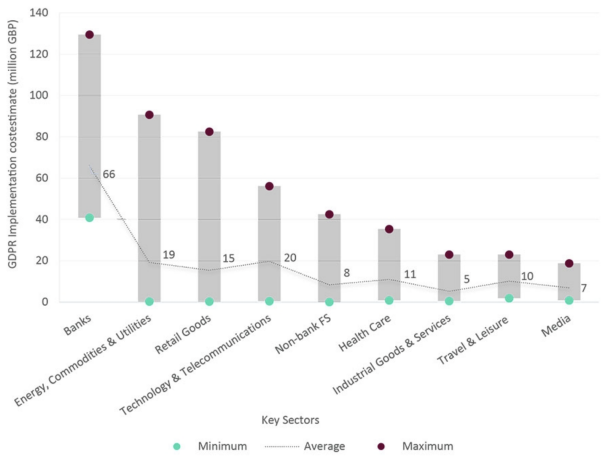

Figure 9.4: GDPR implementation cost per sector [52]

standards that the banks have to fulfill. As the banks have to store sensitive data at the heart of the system [52]. Additionally, linking this back to Figure 9.1, it can be seen that the same industries are at the top of the cost spectrum. Therefore, suggesting that companies that operate in these industries have to focus on implementing the GDPR and may consider to take additional actions to reduce the costs in the long run.
Looking at factors of size and industries suggests that the implementation of the GDPR is costly for all parties. This has benefits and disadvantages. A benefit of such a high implementation cost is that the probability is higher that the companies will invest more time and money to implement it correctly. Consequently, the standard of data protection will be higher, which is the intention of the European Union. However, these high implementation costs also have disadvantage. For example, companies may not try to implement the Regulations properly and consequently risk the fine if it is economically better. Nonetheless, this drawdown is limited due to the high fines that the GDPR proposes. Another disadvantage is that the implementation cost only fulfills the regulations. Only limited action can be considered during this implementation time to make the system more secure from a specific company viewpoint, as resources are bounded.

9.4.3 Compare with other countries

For this comparison, it was selected to compare the GDPR to the Lei Geral de Proteção de Dados (LGPD). The LGPD is the data protection regulation of Brazil that passed the Brazilian National Congress by August 14th, 2018. Lastly, the LGPD and was signed into national law on September 18th, 2020 [54]. The LGPD is based on the GDPR, and therefore these two regulations share many concepts [54]. This same regulation basis can be seen as something positive, as worldwide operating companies have a common basis to work on.
Firstly, the similarities between the LGPD and GDPR will be discussed. Most importantly, consent is vehemently protected and safeguarded by both laws. Additionally, the core parts are similar, and the details in the implementation do vary. The essential regulation of the LGPD are [55]:

- Information to holders of data on possible incidents,
- proof of consent,
- portability of data,
- indication and responsibility of the agents responsible for the operation of the data and the security rules for storage,

- transmission,
- and handling.

The points mentioned above are also one of the essential parts of the regulation for the GDPR.
The first difference between the regulations is the *territorial* scope. Both the GDPR and the LGPD follow an extraterritorial scope. Therefore, all companies offering goods or services to data subjects in the EU or Brazil, regardless of where the companies are located. However, the GDPR explicitly includes organizations that are not established inside of the European Union but monitor the behavior of its individuals [56]. This explicit statement is not included in the LGPD; therefore, it does not have such a strong *territorial* scope as the GDPR.
Another difference between the GDPR and the LGPD is the legal bases for data processing. Both the GDPR and the LGPD have the same essential six reasons. Namely, explicit consent, legitimate interest, contract performance, legal obligation, public task, and vital interest [57]. However, the LGPD has four additional reasons, which are the following: research, rights in legal proceedings, health protection, and credit protection [57]. As the LGPD has additional legal bases to process data, this means that under the LGPD, the processor is allowed to process a larger variety of data.
A third difference refers to when a Data Protection Officer (DPO) needs to be hired. In this case, the GDPR gives an outline description in what scenarios it should be done [58]. However, the LGPD states explicitly that a DPO is mandatory, but only for organizations that process data [58].
A further difference is when the company has to report a data breach. As mentioned in Table 9.2, the GDPR requires the controller to have 72 hours to notify a data breach incident. In contrast, the LGPD has no explicit time limit when a data breach has to be notified. Instead, it is stated that the notification has to happen "in a reasonable time period." Therefore, the LGPD regulation is a bit vague, as it leaves space for interpretation.
A final difference between the GDPR and the LGPD is the fines that can be given. As mentioned in Table 9.2, the highest possible fine is up to €20 million or 4% of annual global revenue, whereas the higher one should be depicted. However, the LGPD's maximum fine according to Article 52 is 2% of a private legal entity's group's, or conglomerate's revenue in Brazil, for the prior fiscal year, excluding taxes, is up to a maximum of 50 million BRL (approximately €11 millions). This shows that the fines in the EU are substantially stricter than those in Brazil.

9.4.4 Protective Measures Against Data Breaches

As mentioned in the Background section, there are 6 common causes of data breach, below are the protection strategies for each causes:

1. **Criminal hacking**: Deploy firewall in the database server and allow connections only from trusted hosts. Block all non-used ports and block all outbound connections. Companies can set exceptions for replication, and linked databases [60].

2. **Social engineering**: Teach the users how to avoid social engineering and show the following tips when they are using the system, *e.g.,* in the loading screen:

    - **Delete any request for personal information or passwords**: Nobody should be contacting the users for their personal information via email unsolicitedly [61], even the employees.

- **Reject requests for help or offers of help**: Social engineers can and will either request users' help with information or offer to help them (i.e., posing as technical support). If the user did not request any assistance from the sender, consider any requests or offers a scam. Research about the sender before committing to sending them anything. [61].
- **Set the spam filters to high**: Email software has spam filters. Check the settings, and set them to high to avoid risky messages flooding into the inbox. Just remember to check them periodically as it is possible legitimate messages could be trapped there from time to time [61].
- **Always be mindful of risks**: Double check, triple check any request the users get for the correct information. Look out for cybersecurity news to take swift actions if the users are affected by a recent breach [61].

3. **Weak security**:

   - **Be Updated with Security Patches**: Try to install patches as fast as possible. Database vulnerabilities are serious, sometimes the database server can be easily compromised with a simple query. Always test patches on non-production servers first and monitor for patch problems on mailing lists. Sometimes patches could open holes instead of fixing them [60].
   - **Timely Check for Object & System Permissions**: Always check views, stored procedures, tables of the database and their permissions. If any change in permission is found, then there may be high chances of compromise of misconfiguration [60].
   - **Use Encryption**: At the network level, companies must use SSL and database proprietary protocols. At file level, companies must encrypt file and file system timely backups should be done. On the other hand, database level column encryption should be done to encrypt all data. Companies can use APIs to provide this sort of encryption [60].

4. **Stolen devices**: Do the following steps to stop further access from the thief.

   - Change account passwords.
   - Clear auto-fill from browsers.
   - Deauthorize the device.

5. **Weak credential**: Set a good password policy:

   - **Must be at least 8 characters long.**
   - **Must contain at least one upper-case letter, number, and special character.**
   - **No personal information, includes date of birth, name and mother's name.**
     Reduce the risk of being social engineering attacks, i.e., pretexting.
   - **No dictionary words.**
     Ban common passwords and passwords that are exposed from previous data breaches to prevent dictionary attack.
   - **Must be changed after 3 months.**
     If the password is leaked in other databases, users' accounts can be recovered or mitigated from the risk on time.
   - **Warnings begin that a password should be changed 15 days prior to expiration.**

- **After 7 days with an expired password, an account is locked.**
  It might be difficult for users to notice someone else is using their accounts and leaking their data. So that forces users to change their password regularly will reduce the risk of this.

Use additional authentication measures:

- **Set a login expiration time and ask user to re-enter their password for every new session.**
  Users may login their accounts on public devices and forget to log off. After a certain period, the accounts will automatically log off. And if other people attempt to use their accounts, they will be authenticated. Therefore, prevents Time-of-Check to Time-of-Use (TOCTTOU) attack.
- **Enforce registration for multi-factor authentication.**
  For some systems that stores valuable and important data, force users to register for multi-factor authentication and keep them up to date. *E.g.,* bank apps.
- **Enable risk-based multi-factor authentication.**
  If the system detects suspicious activities, *e.g.,* different IP address from the previous login, different delivery address and too many login attempts, then forces users to multi-factor authentication.

Storing the password securely in the database:

- **Salting**
  Prepend a random salt to the password, then hash the password. This will make cracking multiple passwords harder and slower because each password is individually hashed. If one user's weak password is cracked, not all other users with that password will be cracked because the attacker will need to try all combinations of the salts and the password to crack another one. And it will prevent rainbow table attacks because there are so many different combinations that it cannot be pre-computed and stored.
- **Use complex hash function**
  The more complex hash function, the slower the hash speed, therefore password cracking is slower and harder. *E.g.,* PBKDF2: iterates HMAC to increase complexity; bcrypt: memory intensive and does not work well on GPU.

6. **Insider [62]**:

- Perform enterprise-wide risk assessments.
- Clearly document and consistently enforce policies and controls.
- Monitor and control remote access from all endpoints, including mobile devices.
- Recycle old hardware and documentation properly.
- Use a log correlation engine or security information and event management system (SIEM) to log, monitor and audit employee actions.

9.5 Summary and Conclusion

This report focussed on the aspects of GDPR and its implications on data breaches. Firstly, five different examples of data breaches were covered: British Airways, Marriot

International, who got successfully attacked twice, Morle.net, and the National Revenue Agency Bulgaria. Additionally, comparing these GDPR Fines with the five most significant fines given under the GDPR to know where to rank the fines. Afterwards, we discussed four different aspects of the GDPR:

1. Discussing the importance of the GDPR. The importance was examined by comparing it to the old regulation, which the GDPR replaced in 2018.

2. The impact on the companies was examined by focusing on the direct and reputation costs and examining the implementation cost. As a result of this, it was discovered that the GDPR is a burden for smaller companies. Additionally, the cost for companies in the healthcare and financial sectors are the highest.

3. A comparison was conducted between the GDPR and the LGPD, based on the GDPR. Thereby finding that the GDPR is, in general, the stricter regulation.

4. We examined protective measured against getting attacked. Thereby focussing on the six common causes of data breaches. Mainly: criminal hacking, social engineering, weak security, stolen devices, weak credentials, and Insiders.

After examining these different view point, an overview of the different implications and effects of GDPR can be gained.
Although GDPR is the world's most stringent law on data protection, but it still requires the data controller to take adequate measures to ensure its data security. There are thousands of data breaches annually in the recent years, billions of customer information are leaked around the world. GDPR stipulates that if data is leaked, users have the right to obtain relevant information quickly. This requires companies to report data breaches to data protection agencies within 72 hours. But currently only few service providers can notify customers of data security incidents within 24 hours.
GDPR is a good opportunity to allow companies to review their privacy protection policies. This process requires the services and technical support of the data company, which will increase customer costs, but at the same time it can also increase the value of the company. According to a survey reported by the business analytics leader SAS, with the arrival of GDPR in 2018, only 7% of the companies have completed compliance, and nearly half of the interviewed companies said that this new regulation will have a significant impact on the company's Artificial Intelligence (AI) related projects. Less than half of the global companies said they can meet GDPR compliance requirements on time. Many small companies are still on the sidelines due to lack of resources and guidance [63].
The fines for violating the GDPR are very high and have a certain deterrent. Enterprises have only three options, either to strive to comply with regulations; or to withdraw from the European market; the other is to bear the risk of being fined. The last two options are not the options that any company is willing to choose, so the only option is to comply with GDPR.
In fact, GDPR has a tremendous impact on the Internet. It is related to the future development of every Internet company, it is related to the major adjustment of the current Internet basic business model, and more importantly, it is related to each of our customers.

Bibliography

[1] National Public Radio, "Facebook Pays $643,000 Fine For Role In Cambridge Analytica Scandal", October 2019. https://www.npr.org/2019/10/30/774749376/facebook-pays-643-000-fine-for-role-in-cambridge-analytica-scandal?t=1633695530813, Last visit November 11, 2021

[2] Jacquelyn Bulao, "How Much Data Is Created Every Day in 2021?", November 2021. https://techjury.net/blog/how-much-data-is-created-every-day/#gref, Last visit November 11, 2021

[3] CMS - Law, Tax, Future. "GDPR Enforcement Tracker", 2021. https://www.enforcementtracker.com/, Last visit August 16, 2021

[4] Abi Tyas Tunggal, "The 60 Biggest Data Breaches (Updated for November 2021)", November 2021. https://www.upguard.com/blog/biggest-data-breaches, Last visit August 16, 2021

[5] DLA Piper, "DLA Piper GDPR Fines and Data Breach Survey: January 2021", 2021. https://www.dlapiper.com/en/uk/insights/publications/2021/01/dla-piper-gdpr-fines-and-data-breach-survey-2021/, Last visit August 16, 2021

[6] Trend Micro, "Data Breach". https://www.trendmicro.com/vinfo/gb/security/definition/data-breach, Last visit August 16, 2021

[7] IBM, "Cost of a Data Breach Report 2021 explores ways to help mitigate risk", 2021. https://www.ibm.com/security/data-breach, Last visit August 16, 2021

[8] Z. Fang, M. Xu, S.Xu, and T. Hu. A Framework for Predicting Data Breach Risk: Leveraging Dependence to Cope With Sparsity. volume 16, pages 2186-2291, 2021

[9] ENISA. DataBreach ENISA Threat Landscape, 2020. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2020-data-breach, Last visit August 16, 2021

[10] Trend Micro, "Data Breach". https://www.trendmicro.com/vinfo/gb/security/definition/data-breach, Last visit August 16, 2021

[11] Schwegman Lundberg Woessner, "Protecting & Handling Confidential Information", August 2019. https://www.slwip.com/resources/protecting-handling-confidential-information/, Last visit August 16, 2021

[12] U.S. DEPARTMENT OF LABOR, "Guidance on the Protection of Personal Identifiable Information". https://www.dol.gov/general/ppii, Last visit August 16, 2021

[13] Local Government Association, "General Data Protection Regulation (GDPR)", 2021. https://www.local.gov.uk/our-support/guidance-and-resources/general-data-protection-regulation-gdpr, Last visit October 6, 2021

[14] Antoni Gobeo; Connor Fowler; William J. Buchanan, "1 The GDPR Fundamentals," in GDPR and Cyber Security for Business Information Systems , River Publishers, 2020, pp.1-36.

[15] Juliana De Groot, "A DEFINITION OF GDPR", 2020. https://digitalguardian.com/blog/what-gdpr-general-data-protection-regulation-understanding-and-complying-gdpr-data-protection, Last visit October 6, 2021

[16] Ben Wolford, "What is GDPR, the EU's new data protection law?", 2021. https://gdpr.eu/what-is-gdpr/, Last visit November 6, 2021

[17] P. Voigt and A. Bussche, *The EU General Data Protection Regulation (GDPR) A Practical Guide*, Cham Switzerland, 2017, DOI:10.1007/978-3-319-57959-7

[18] GDPREU, "GDPR Fines & Data Breach Penalties", 2021. https://www.gdpreu.org/compliance/fines-and-penalties/#How_are_GDPR_Fines_Calculated, Last visit November 6, 2021

[19] Kate O'Flaherty, "How the British Airways breach will reveal the true cost of GDPR", 2018. https://www.forbes.com/sites/kateoflahertyuk/2018/09/20/how-the-british-airways-breach-will-reveal-the-true-cost-of-gdpr/?sh=ba48dad3edff, Last visit November 6, 2021

[20] Lily Hay Newman, "How Hackers Slipped by British Airways' Defenses", 2018. https://www.wired.com/story/british-airways-hack-details/, Last visit November 6, 2021

[21] Yonathan Klijnsma, "Inside the Megacart Breach of British Airways: How 22 Lines of Code Claimed 380,000 Victims", 2018. https://www.riskiq.com/blog/external-threat-management/magecart-british-airways-breach/, Last visit November 6, 2021

[22] Data Privacy Manager, "British Airways fine for 2018 data breach reduced to 20 million British Pound", 2020. https://dataprivacymanager.net/ico-reduces-british-airways-gdpr-fine-to-20-million-for-2018-data-breach/, Last visit November 6, 2021

[23] Data Privacy Manager, "New Marriott breach - what is going on?", 2020. https://dataprivacymanager.net/new-marriott-breach-2020-what-is-going-on/?hsCtaTracking=7588fcc1-7d1e-448d-8a8d-b3124c48ab46%7Cbe9ec45a-9d23-4ea3-81ce-87709e992bd6, Last visit November 6, 2021

[24] Graeme Messina, "Lessons not learned? Another Marriott data breach", 2020. https://resources.infosecinstitute.com/topic/lessons-not-learned-another-marriott-data-breach/, Last visit November 6, 2021

[25] Josh Fruhlinger, "Marriott data breach FAQ: How did it happen and what was the impact", 2020. https://www.csoonline.com/article/3441220/marriott-data-breach-faq-how-did-it-happen-and-what-was-the-impact.html, Last visit November 6, 2021

[26] Tessian, "20 Biggest GDPR Fines of 2019, 2020, and 2021 (So Far)", 2021. https://www.tessian.com/blog/biggest-gdpr-fines-2020/, Last visit November 6, 2021

[27] Alice Baker, "Polish data protection authority issues €645,000 fine to online retailer", 2019. https://www.itgovernance.eu/blog/en/polish-dpo-issues-fine-to-online-retailer-for-data-breach, Last visit November 6, 2021

[28] Data breaches, "Morele.net", 2019. https://databreach.es/companies/morelenet/recDQ7x0PeOq9YM2t, Last visit November 6, 2021

[29] Digital Freedom Fund, "Mass data breach by the Bulgaria National Revenue Agency", 2021. https://digitalfreedomfund.org/mass-data-breach-by-the-bulgaria-national-revenue-agency/, Last visit November 6, 2021

[30] Data Privacy Manager, "Luxembourg DPA issues €746 Million GDPR Fine to Amazon", July 2021. https://dataprivacymanager.net/luxembourg-dpa-issues-e746-million-gdpr-fine-to-amazon/?hsCtaTracking=0e8e88ee-aef4-4c83-be29-8ea99797c411%7C46ee48bb-39fd-44dd-b159-64af4a04c98c, Last visit November 6, 2021

[31] News 18, "EXPLAINED: Why Amazon Was Fined $887M, And How EU's Data Law Keeps Tripping Up Tech Giants", August 2021. https://www.news18.com/news/explainers/explained-why-amazon-was-fined-887m-and-how-eus-data-law-keeps-tripping-up-tech-giants-4033355.html, Last visit November 6, 2021

[32] Data Privacy Manager, "GDPR fine: WhatsApp faces €225 million for transparency violation", August 2021. https://dataprivacymanager.net/gdpr-fine-whatsapp-faces-e225-million-for-transparency-violation/?hsCtaTracking=931b036b-7ed1-4666-a9de-9a2e81b7e7ee%7Ca713759d-456e-4257-969b-9b4017772042, Last visit November 6, 2021

[33] BBC news, "WhatsApp issued second-largest GDPR fine of €225m", September 2021. https://www.bbc.com/news/technology-58422465, Last visit November 6, 2021

[34] Data Privacy Manager, "H&M fined €35,3 Million for violation of the GDPR", October 2020. https://dataprivacymanager.net/german-dpa-issued-e353-million-gdpr-fine-to-hm-for-violation-of-the-general-data-protection-regulation/?hsCtaTracking=2ff6d63d-d289-492c-8d1d-12d887afdf5a%7C606030f5-4b1f-4d11-b8f2-75490fb5e45f, Last visit November 6, 2021

[35] Data Privacy Manager, "€27,8 million GDPR fine for Italian Telecom - TIM", Feburary 2020. https://dataprivacymanager.net/e278-million-gdpr-fine-for-italian-telecom-tim/?hsCtaTracking=6680ce94-947d-4fb2-9f28-7d6aa4b9f485%7C76022d76-9c5e-4c0b-a5c6-b1039731b829, Last visit November 6, 2021

[36] Data Privacy Manager, "20 biggest GDPR fines so far [2019, 2020 & 2021]", October 2021. https://dataprivacymanager.net/5-biggest-gdpr-fines-so-far-2020/, Last visit November 6, 2021

[37] M. Lourenco and L. Marinos, "Data Breach ENISA Threat Landscape", ENISA, Attiki, Greece, DOI: 10.2824/552242

[38] I. Obaydin, L. Xu, and R. Zurbruegg, "The Unintended Cost of Data Breach Notification Laws: Evidence from Mangerial Bad News Hoarding", Adelaide, Australia, April 2021

[39] Todd Feinman, "Companies Need to Take Responsibility for Protecting Sensitive User Data", 2015. https://www.entrepreneur.com/article/242355, Last visit November 6, 2021

[40] Muskaan Prasad, "GDPR: What to Know About the World's Most Strict Law on Data Protection", March 2021. https://legex.in/blog/gdpr-what-to-know-about-the-worlds-most-strict-law-on-data-protection, Last visit November 6, 2021

[41] Virtual College, "What are the main differences between GDPR and the Data Protection Act?", January 2018. https://www.virtual-college.co.uk/resources/the-differences-between-gdpr-and-data-protection, Last visit November 6, 2021

[42] GDPR, "Does the GDPR apply to companies outside of the EU?". https://gdpr.eu/companies-outside-of-europe/, Last visit November 6, 2021

[43] GDPR, "What is considered personal data under the EU GDPR?". https://gdpr.eu/eu-gdpr-personal-data/, Last visit November 6, 2021

[44] GDPR, "What are the GDPR consent requirements?". https://gdpr.eu/gdpr-consent-requirements/, Last visit November 6, 2021

[45] GDPR, "What is GDPR, the EU's new data protection law?". https://gdpr.eu/what-is-gdpr/, Last visit November 6, 2021

[46] GDPR, "Art. 17 GDPR: Right to erasure (right to be forgotten)". https://gdpr.eu/article-17-right-to-be-forgotten/, Last visit November 6, 2021

[47] GDPR, "Art. 20 GDPR: Right to data portability". https://gdpr.eu/article-20-right-to-data-portability/, Last visit November 6, 2021

[48] GDPR, "Everything you need to know about the GPDR Data Protection Officer (DPO)". https://gdpr.eu/data-protection-officer/, Last visit November 6, 2021

[49] GDPR, "What are the GDPR Fines?". https://gdpr.eu/fines/, Last visit November 6, 2021

[50] Business News Daily, "How Businesses Are Collecting Data (And What They're Doing With It)". https://www.businessnewsdaily.com/10625-businesses-collecting-data.html, Last visit November 6, 2021

[51] IBM, "Cost of a Data Breach Report 2020", United States of America, July 2020

[52] Consultancy.uk, "GDPR compliance to cost FTSE100 firms 15 million British Pounds, banks face largest bill", 2017. https://www.consultancy.uk/news/15101/gdpr-compliance-to-cost-ftse100-firms-15-million-banks-face-largest-bill, Last visit November 6, 2021

[53] Luke Irwin, "How much does GDPR compliance cost in 2021", 2021. https://www.itgovernance.eu/blog/en/how-much-does-gdpr-compliance-cost-in-2020, Last visit November 6, 2021

[54] Securiti, "What is Lei Geral de Proteção de Dados (LGPD)", 2021. https://securiti.ai/what-is-lgpd/, Last visit November 6, 2021

[55] Leonardo Neri, "Brief Comparison Between the LGPD and the GDPR Regarding the use of Personal Data", 2021. https://www.mazzuccoemello.com/en/brief-comparison-between-the-lgpd-and-the-gdpr-regarding-the-use-of-personal-data/, Last visit November 6, 2021

[56] Andrada Coos, "LGPD vs. GDPR: The Biggest Differences", 2019. https://www.endpointprotector.com/blog/lgpd-vs-gdpr-the-biggest-differences/, Last visit November 6, 2021

[57] Email Marketing Journal, "Towards and Beyond - Legal Bases for Email Marketing", 2019. https://www.emailmarketingjournal.com/2019/11/towards-lgpd-and-beyond-legal-bases-for-email-marketing/, Last visit November 6, 2021

[58] Richie Koch, "What is the LGPD? Brazil's version of the GDPR", 2021. https://gdpr.eu/gdpr-vs-lgpd/, Last visit November 6, 2021

[59] Wikipedia, "Data Protection Directive", October 2021. https://en.wikipedia.org/wiki/Data_Protection_Directive, last visit November 11, 2021.

[60] The CyberSecurity Place, "Database Hacking & Its Prevention", April 2019. https://thecybersecurityplace.com/database-hacking-its-prevention/, last visit November 11, 2021.

[61] Jen Trang Nguyen, "Five Ways to Prevent Social Engineering Attacks", October 2018. https://www.mdsny.com/5-ways-to-prevent-social-engineering-attacks/, last visit November 11, 2021.

[62] Netweix, "Insider Threat Prevention Best Practices". https://www.netwrix.com/Insider_Threat_Prevention_Best_Practices.html, last visit November 11, 2021

[63] PR Newswire, "Survey: Only 7 percent of businesses GDPR-compliant as deadline looms, data privacy gains prominence", April 2018. https://www.prnewswire.com/news-releases/survey-only-7-percent-of-businesses-gdpr-compliant-as-deadline-looms-data-privacy-gains-prominence-300635209.html, last visit November 11, 2021.

Chapter 10

Impact of the Remote Working Economy on Cybersecurity

Yao Song, Denys Trieskunov

This report presents the impact of the remote working economy on cybersecurity. The first problems that we face due to the pandemic are stated. Then we list and describe the major cybercriminal types and attacker techniques during the pandemic. Countermeasures are also discussed, from two aspects of security awareness management and purely technique levels.

Contents

10.1 Introduction

Data protection and cybersecurity were always prevalent questions for IT companies. With the switch to remote work, it became harder for companies to prevent cyberattackers and protect data privacy. All the existing vulnerabilities have not disappeared, while the ones that were not an issue before have emerged. Employees working online opened the door for attackers because either the necessary countermeasures are not implemented on their home office equipment or employees neglect to use them. Also, additional sources of vulnerabilities are the applications the company uses for remote communication. To prevent attacks, companies implemented countermeasures such as using Virtual Private Networks (VPN) and other means for secure connection to company networks, and drifted away from external applications, security of which was proven to be not reliable or questionable. Also, the internal security policies for employees were set to protect themselves and the company from cyberattacks. These measures helped to prevent the majority of attacks. In the following sections, we will talk about specifics of the changes that happened, cyberattacks in general, and the ones that got more widespread after changing to the home office. Also, we will talk about organizational security management and policies and introduce some typical security models and newly proposed network architectures.

10.2 Problem Statement

With the Covid-19 outbreak, companies were forced to switch to doing home office. Since this switch was planned by little to no companies and the adaptation had to go fast, many were unprepared to undergo this change. Companies had to sacrifice security measures at the start to allow their workers to start working from home faster, which opened the door for cyberattacks as well as other issues related to this switch. Many companies had the infrastructure set up for the office environment: internal networks, firewalls, and limited access machines. As an example of limited access machines, some companies store some of their databases on the local machines that can be accessed by employees in the office, but are not connected to the internet so nobody can access them from outside the office. The companies that had means for office employees to work at home had arguably easier times applying this measures for employees' personal machines, but still, it is a huge amount of work that cannot be done instantly and without any testing. Therefore the technical aspect of the shift is one of the targets of cybercriminals. The other aspect that, despite existing before the switch, became much more prevalent, is the human aspect of the switch. With switching to a home office work environment, naturally, all communication got transferred to emails and messages in various apps. This made it significantly harder to differentiate between a legitimate message and a carefully crafted malicious one.

10.2.1 Shift to Home Office

With the shift to home office companies had to undergo a lot of changes. This was both equipment changes to provide employees with everything necessary, activities to support employees, and policy changes to maintain a productive and secure work routine. [6] have devised two metrics for remote work potential: the maximum potential, including all activities that theoretically can be performed remotely, and a lower bound for the effective potential for remote work, which excludes activities that have a clear benefit from being done in person.
To determine the overall potential for remote work for jobs and sectors, McKinsey used the time spent on different activities within occupations. They found that remote work potential is concentrated in a few sectors. Finance and insurance have the highest poten-

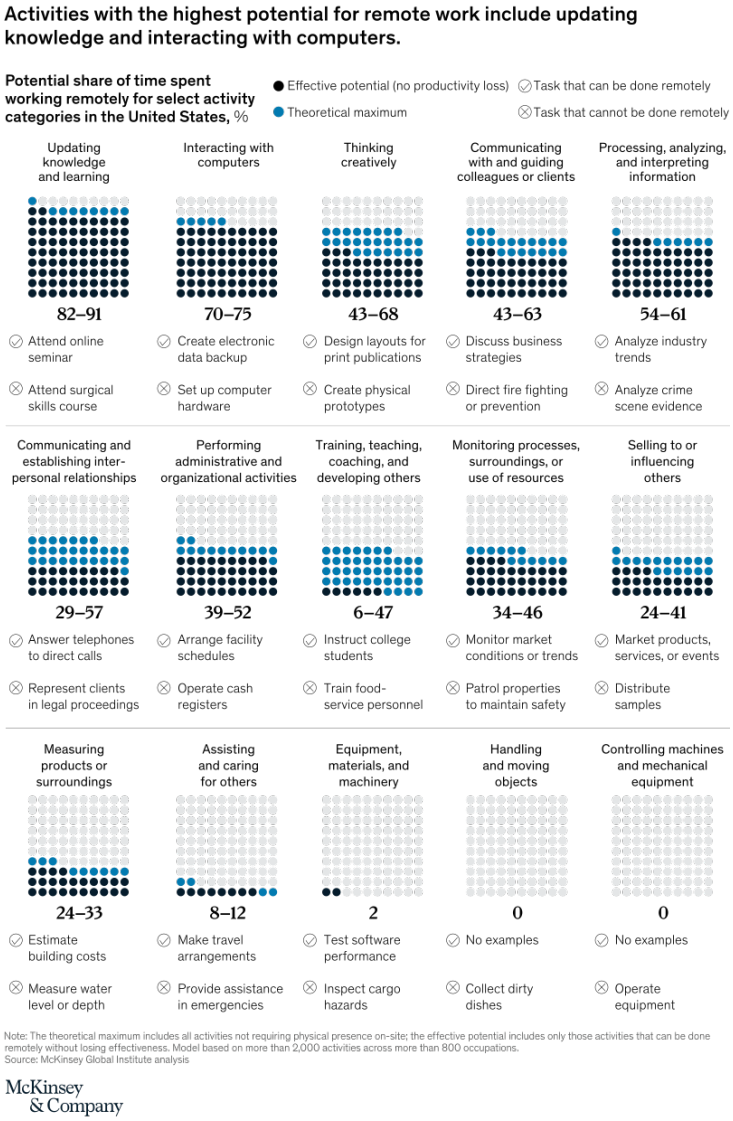

Figure 10.1: Activities with the Highest Potential for Remote Work

tial, with three-quarters of time spent on activities that can be done remotely without a loss of productivity. Management, business services, and information technology have the next highest potential, all with more than half of employee time spent on activities that could effectively be done remotely. These sectors are characterized by a high share of workers with college degrees or higher.

All companies should develop home office policies for the employees to ensure cybersecurity, business data protection, and clients' data protection. These policies should formulate the security measures and ensure that the employees are fully aware and complicit with them. The security measures are both technical and behavioral. Behavioral security measures are implemented to secure the employee from social engineering attacks and other forms of stealing data using social means. [7] SKW Schwarz article provides the following examples:

- Employees should stay away from using personal email accounts. Using personal email may give an access to confidential information if the connection to the email provider is unsecured or not secured well enough.
- Employees should be careful when using messages. If the message is used it should have end-to-end encryption.

- Personal work space needs to be protected from other people. Nobody in the household should have an access to any protected information. It includes paper documents, digital information, phone calls, etc.
- Employees should be instructed what to do in case of data loss if it occurs. In the case of data loss the employee must know how to act immediately to allow the best possible damage control.

As for technical changes, it is recommended to provide the employees with company issued devices with pre-installed security software. This allows the company to control that the work devices that employees use to comply with all security regulations and prevents potential mistakes employees might make a installing security software. Additionally, the connection to all company services should be secured using VPN tunnel or encrypted remote connection. For video conferences only secure video conference providers have to be used, the same goes for the phone calls. [**?**] There are three main categories of cyberattacks: criminal, political, and personal. Since the latter two are not related to remote work and in essence are very similar to criminal ones, we will focus on the criminal category. The main goal of the attackers is to get financial benefit from either money theft, data theft, or intervention into business processes. Another target for the attackers is the company's private information that they can use to get a competitive advantage over the competitor or steal intellectual or trade secrets.
Typical cybercriminals can be either private individuals or organized group of people.
[2] One way to classify cyberattack risks is by outsider versus insider threats.
External cyberthreats include:

- Organized criminals or criminal groups
- Professional hackers, like state-sponsored actors
- Amateur hackers, like hacktivists

Insider threats are users who have authorized and legitimate access to a company's assets and abuse them either deliberately or accidentally. They include:

- Employees careless of security policies and procedures
- Disgruntled current or former employees
- Business partners, clients, contractors or suppliers with system access

More specifically, targets for cyberattacks are usually one of the following (or similar):

- Financial data of the business
- List of customers
- Financial data of the company's customers
- Information protected by GDPR
- Login credentials of the users
- Intellectual and trade secrets
- Disrupting services of the company

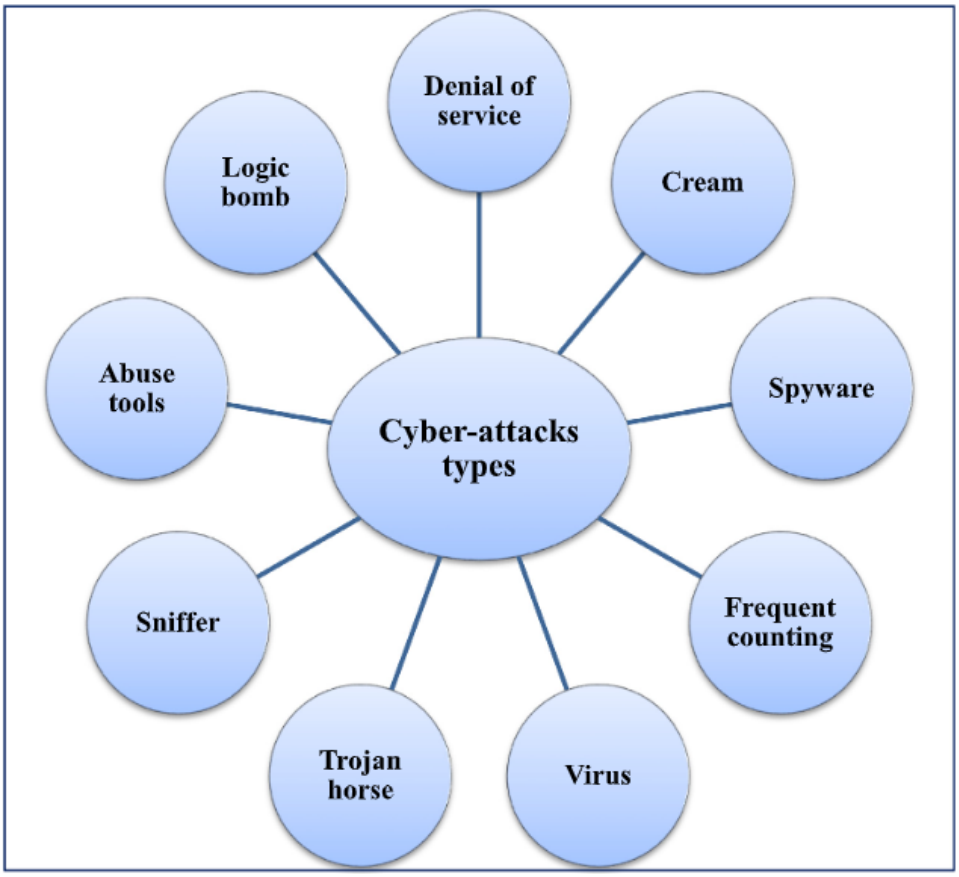

Figure 10.2: Main Cyberattack Types [3]

There are several popular types of cyberattack methods: denial of service, virus, botnet, abuse tools, logical bomb, sniffer, trojan horse, worm.
The denial of service attack aims to deny the users access to the various services of the system or the system as a whole. The main goal of it is to disrupt the inner processes of the company. It can be performed in different ways, one of the popular ones is a coordinated attack from a large number of distributed systems, it is usually referred to as a DDoS attack [3]. This is often done by using worms and multiplying them on multiple computers to attack the target. Abuse tools are available to the public that can detect and enter vulnerabilities in networks with different skill levels. A logic bomb is a type of attack when the malicious activity automatically triggers after a specific event in the system. Sniffer gets into the package stream and looks for important information like passwords, financial information, etc. Trojan horse is a malicious program that conceals itself into the other legit program. After infiltrating the target it can act like other types of attacks. The virus is the attack when the malicious program inserts itself into other programs in the system ”infecting” them. It can also contain other means of attacks like bots(to add the computer into the botnet, which will be explained further). Botnets are usually secretly installed on the target computer, allowing the unauthorized user to remotely control the target system to achieve their malicious goals.
All the upper mention cyberattack types are software related, however, there is another huge cluster of attacks - social engineering attacks. While software-related attacks aim to find and exploit weaknesses in software, social engineering attacks aim to exploit human behaviour. [4] In essence, social engineering refers to the design and application of deceitful techniques to deliberately manipulate human targets. In a cybersecurity context, it is primarily used to induce victims towards disclosing confidential data, or to perform actions that breach security protocols, unknowingly infecting systems or releasing classified information. The basis of a social engineering attack is to avoid cybersecurity systems

through deceit, exploiting the weakest link, the people involved. Throughout the interaction, victims are unaware of the destructive nature of their actions. The social engineer exploits innocent instincts, not criminals. Explicit methods such as threats or bribery do not fall within the scope of social engineering.
Further, we'll describe main types of the social attacks.
One of the most popular social engineering attacks is phishing. This attack is usually conducted by concealing malicious software in a seemingly legitimate email or website. In some cases, these attacks apply psychological tricks on the target to create a sense of urgency or criticalness of the said email or website to affect the target's judgment and make him operate quickly and without double checking the legitimacy of the email/website. On the figure 2.2 the example of such email can be seen.

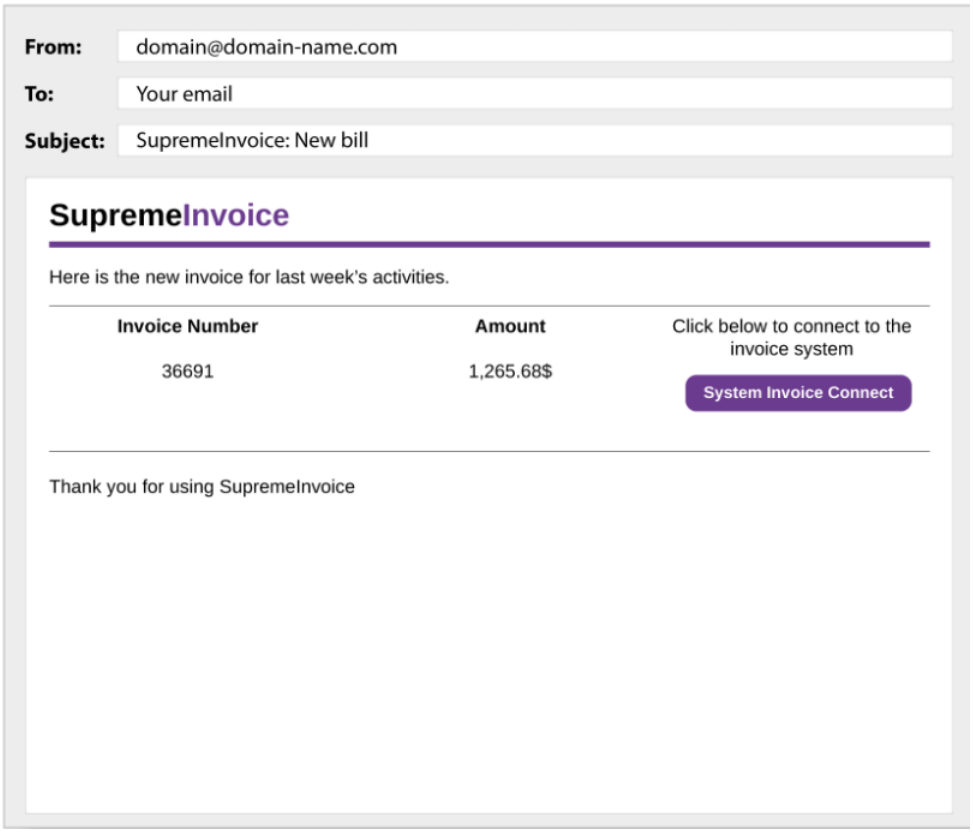

Figure 10.3: Example of the Phishing Email

Spear-fishing is a modification of phishing, where the attacker first researches the target's personal information to make the message more personal and therefore less likely to be identified as an attack.
Another popular attack type is watering whole. This attack involves the attacker researching legitimate websites that the target often uses and infecting the websites themselves with malware. After that, the attacker just waits for the target to enter the website. This type of attack, though, requires good technical knowledge.

10.2.2 Cyberattacks/Threats during the Pandemic

Most workplaces already have systems set up to deflect the most popular cyberattacks. The home office employees, however, are more susceptible to them. Despite the companies doing their best to equip the home office employees with the proper security tools, it is significantly harder to do that in the remote work environment. Additionally, because these tools create certain complications with using the system and the employees might

be unfamiliar with them, employees might willingly or unwillingly neglect using the tools or misuse them, becoming the potential target for the attack. [5] SailPoint study showed that forty-eight percent of total U.S. respondents said they have experienced targeted phishing emails calls or texts in a personal or professional capacity during the first six months of remote work. Similarly, over half of EMEA and ANZ respondents (51 percent) experienced a phishing attack since the pandemic began, with one in ten (10 percent) reporting a phishing attack targeted them once a week. Working from home lends itself to sharing sensitive company data and communicating important information via virtual platforms; it is a digital gold mine for hackers who own all the tools necessary to infiltrate a company when key defenses are down.

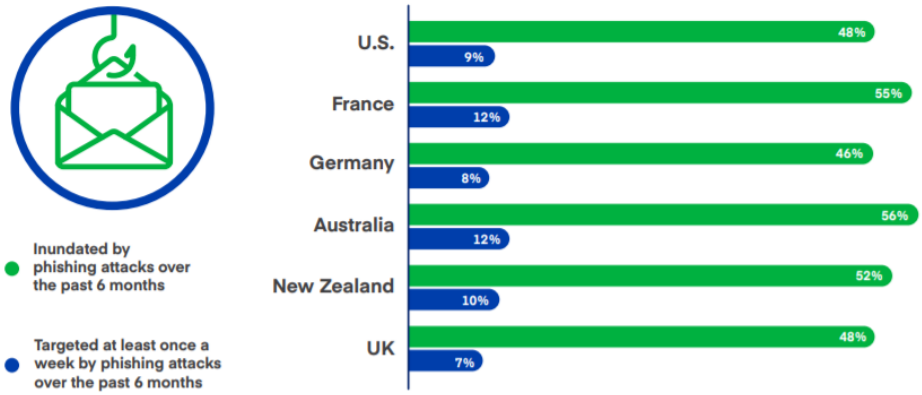

Figure 10.4: Phishing Attacks

Another outlined problem that often occurs during the home office is password sharing. SalePoint study shows that one in four respondents intentionally shared work passwords with a 3rd party, including passwords, roommates, or friends. Over half of the total the U.S respondents have not changed their password for the last six months, 32 percent haven't changed the password in the previous six to twelve months or longer, which is considered malpractice. The figures 2.3 and 2.4 shows the statistics regarding password related problems.

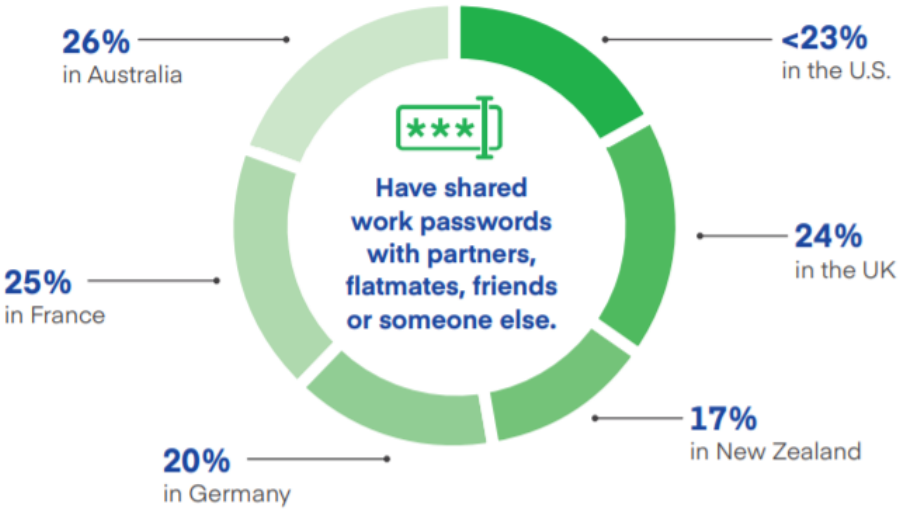

Figure 10.5: Password Sharing Across the World

	Changed password within 1 month:	Haven't changed password in over 6 months:	Computer isn't password protected in the first place:
U.S.	20%	14%	18%
EMEA	23%	44%	3%
France	23%	49%	4%
Germany	24%	39%	2%
Australia	21%	42%	3%
New Zealand	22%	45%	4%

Figure 10.6: Password Malpractices Across the World

Another set of problems arises when the employees use the work devices for personal needs. It may include opening non work related links, online shopping, messengers, social media, etc. This makes a device susceptible to the attacks mentioned in the previous article, such as trojan horses, viruses, etc. Figure 10.7 shows what percentage of employees use company issued devices for personal needs.

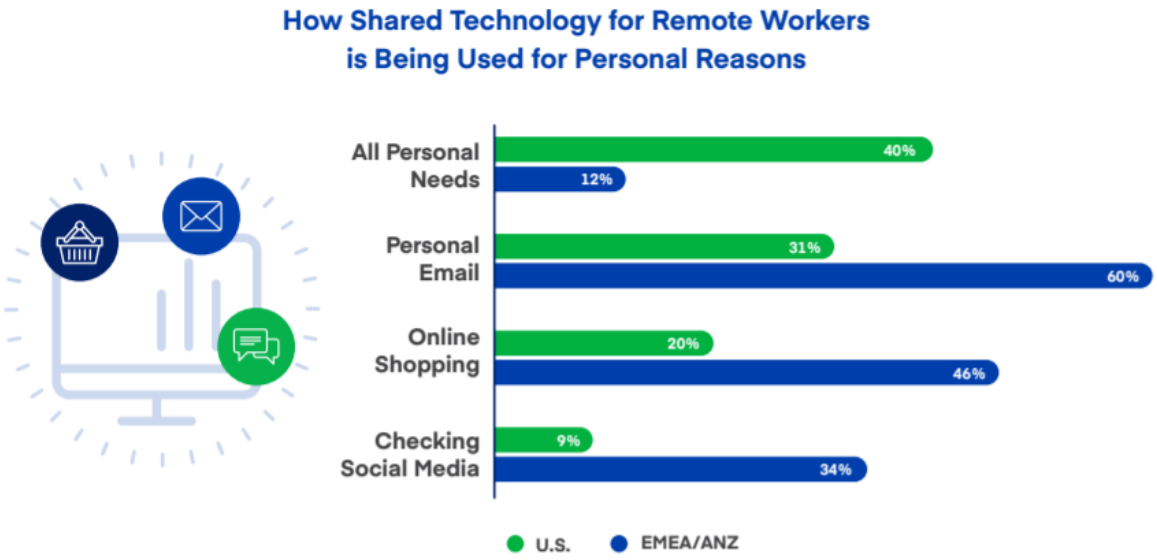

Figure 10.7: Using Company Devices for Personal Needs

Additionally, the attackers during covid started to exploit the covid related information and infect the covid-related websites or send phishing or spam emails with the fake information. [11] BBC has tracked five of the spam campaigns:

- Click here for a cure
- Covid-19 tax refund
- Little measure that saves
- The virus is now airborne
- Donate here to help the fight

Related to the first scope of the campaign, the attackers send messages indicating that they have information related to the development of the vaccine.
Attackers in web-phishing use URL-personalization through of include words related with covid or coronavirus, such as the following URLs

- covid19mobile.app
- covid19-stats.co.za
- coronavirusnotalone.com
- sars-cov2numbers.com
- limpiezascovid.com
- coronademic.net
- coronavirusalerts.com
- coronavirus.technology
- coronavirusmedicine.com
- covidrule.com

Attackers also targeted unsecured applications that companies were forced to use because of the Covid-19 epidemic. As an example, "VPN Proton" was a VPN solution used by several companies. Attackers created a fake "VPN Proton" and infected it with the malware called "AzorUlt", which is a trojan horse that focuses on taking control over users' private information. Another example is Zoom-bombing. Zoom bombers hijacked Zoom conferences taking control over them. Sometimes the attackers used it to behave offensively, posting hate speech or disrupting processes in other ways. However, the attackers can also get monetary benefits by getting access to the private information discussed on the meeting.
Despite the regular cyberattacks are still happening, most successful ones either target employees directly or involve malpractice on the employee side. These malpractices can be either one of the upper mentioned ones or ignoring security policies the company has set up. This may include not using a secure connection, anti-virus software, being reckless with the private data, etc.

10.3 Countermeasures

10.3.1 Cybersecurity Awareness

Based on previous analysis of kinds of cyberattacks, certain responses from the victim (target) are inevitable. Responses include clicking a link enclosed in an email, providing personal information like SSN or bank account, and so forth.

Due to the quarantine policy and social distancing guidelines, people shift to work at home with much more access to the internet, which increases the possibility of cyberattacks. Meanwhile, with less physical connection with others, attackers can take advantage of coronavirus and people's anxieties and scariness towards COVID-19. Failures in communication to official media, fear, uncertainty, and desire to learn about the information about the virus, health situation, and labour market of the country, will result in a blind belief in the online news and more risk of being attacked.
Apart from the disturbed mental condition of people, people working at home usually will not strictly comply with the regulations such as not sending personal email on the working laptop, using anti-virus software, and separating network at home to conduct work. In addition, risk perception plays an important role in protection. If the person barely believes that they will be the target of malicious emails using the identity of an organization or authority, not to mention the existence of kinds of cyberthreats they do not know, their awareness of protection will be minimal.
Thus to put a stop to future cyberattacks, the focus will be on continuous safety awareness training and preparing all employees for cyber risks, which can build on a well-grounded security layer to the system.
Here we discussed some possible policies [16].

- Qualify multi-factor authentication as much as possible to add another layer of security to any application you may use. It can effectively prevent a safety leak.
- Apply password manager, helping avoid potential malpractice including saving or sharing credentials.
- Try a VPN solution. Transmission through an encrypted connection, workers can securely access IT resources within the organization and elsewhere on the Internet.
- Organizations should regularly update cybersecurity policies that are applicable to both remote working and office working. Due to a shift to remote working, the maximum number of people outside the office could be might be not easily estimated. Thus, make sure it enables adequate online access. In addition, policies should consider the use of personal devices. For example, organization data breaches on a personal laptop, the organisation, as well as the employee, may not respond immediately and effectively. And such policies should be adapted according to the currently updated data privacy policy.
- Communication between employees and employers should be conducted on IT equipment provided by their employers. Such as replying to company emails via a private email address should never happen.

10.3.2 Cybersecurity Techniques

Cybersecurity techniques can be used to ensure system availability, integrity, authenticity, confidentiality, and non-repudiation. Cybersecurity can be used to ensure that user privacy is respected. Cybersecurity techniques can be used to establish the user's trustworthiness [8]. Figure 10.8 and Figure 10.9 illustrate the major cybersecurity techniques.

i End-to-End Communicate Network

Here, an End-to-end communications security model is introduced to have an overview that how layered network structure could facilitate cybersecurity. This network architecture can be generally used to solve the security issues in the range of wireless, optical and wired voice, data, and converged networks. It is applicable to service providers,

Techniques	Category	Technology	Purpose
Cryptography	Certificate and public key architecture	Digital signatures	Used to enable the issuance and maintenance of certificates to be used in digital communications
		Encryption	Used encryption of data during transmission or storage
		Key exchange	Establish either a session key or a transaction key to be used to secure a connection
	Assurance	Encryption	Insures data authenticity
Access control	Perimeter protection	Firewalls	Control access to and from a network
		Content management	Monitors traffic for non-compliant information
	Authentication	Single factor	A system that uses user ID/password combinations to verify an identifier
		Two factor	A system that requires two components in order to grant a user system access, such as the possession of a physical token plus the knowledge of a secret
		Three factor	Adds another identification factor such as a biometric or measurement of a human body characteristic
		Smart tokens	Establish trusted identifiers for users through a specific circuitry in a device, such as a smart-card
	Authorization	Role based	Authorization mechanisms that control user access to appropriate system resources based on its assigned role
		Rule based	Authorization mechanisms that control user access to appropriate system resources based on specific rules associated with each user independent of their role within an organization

Figure 10.8: Cybersecurity Technologies Part I[8]

enterprises, and consumers. Besides, it takes security issues of network infrastructure, services and application management, control management into consideration separately. The core idea is to divide a complex set of end-to-end network security-related features into separate architectural components [8]. This layered separation makes wide use of a novel systematic end-to-end scheme that can address new potential security challenges, as well as the existing possible. Another important characteristic of this model is its independence of the underlying structure of a network. To illustrate its layered structure, we will develop it based on three steps: the dimensions, the layer, and the plane.
A security dimension is a set of security measures designed to address a particular aspect of network security [8]. There are eight security dimensions defined to fight against cyberattacks, which are:

- Access control
- Authentication
- Non-repudiation
- Data confidentiality
- Communication security
- Data integrity
- Availability
- Privacy

Those eight dimensions are not only applicable to network security, but also applications and user port. By identifying which dimension is attacked, we can effectively confirm the attacks' type, evaluate the damage and find potential solutions.

Techniques	Category	Technology	Purpose
System integrity	Antivirus	Signature methods	Protect against malicious computer code, such as viruses, worms, and Trojan horses using their code signatures
		Behaviour methods	Checks running programs for unauthorized behaviour
	Integrity	Intrusion detection	Can be used to warn network administrators of the possibility of a security incident, such as files on a server are compromised
Audit and Monitoring	Detection	Intrusion detection	Compare network traffic and host log entries to match data signatures that are indicative of hackers
	Prevention	Intrusion prevention	Detect attacks on a network and take actions as specified by the organization to mitigate the attacks. Suspicious activities trigger administrator alarms and other configurable responses
	Logging	Logging tools	Monitor and compare network traffic and host log entries to match data signatures and host address profiles indicative of hackers
Management	Network management	Configuration management	Allows for the control and configuration of networks, and fault management
		Patch management	Install latest updates, fixes to network devices
	Policy	Enforcement	Allow administrators to monitoring and enforce security policies

Figure 10.9: Cybersecurity Technologies Part II[8]

To construct a hierarchy network architecture, security dimensions are grouped into three different security layers to offer end-to-end security solutions. They are defined as following:

- the infrastructure security layer
- the services security layer
- the applications security layer

Separation of layers can provide sequential layers service with a minimal influence on each other. This separation of these three layers allow for a sequential view of network architecture which aims to identify and solve security issues in products. Figure 10.10 depicts the structure of eight security dimensions and three grouped layers. Furthermore,

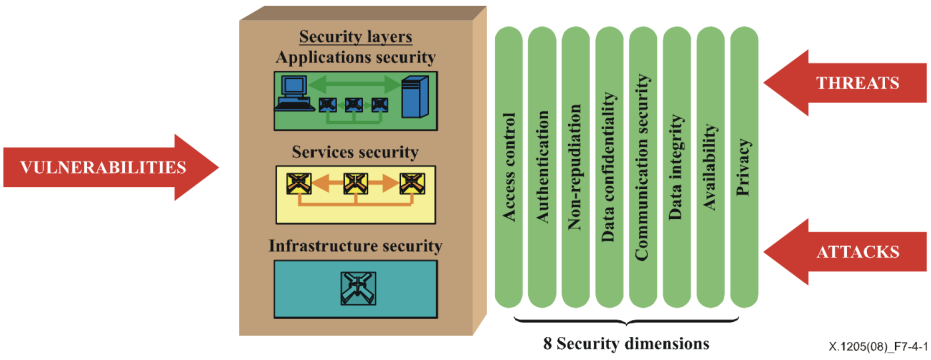

Figure 10.10: Applying Security Dimensions to Security Layers[8]

a certain kind of network activity safeguarded by security dimensions forms a security plane. Three security planes are defined as:

- the management plane

- the control plane
- the end-user plane

Each plane is isolated and will not be affected by the others when performing network security activities. The conceptualization of plane separation allows the distinction between specific security issues related to these activities and the capability to solve them unassisted and independently. Figure 10.11 provides an overview of the whole architecture of dimensions, layers and planes.

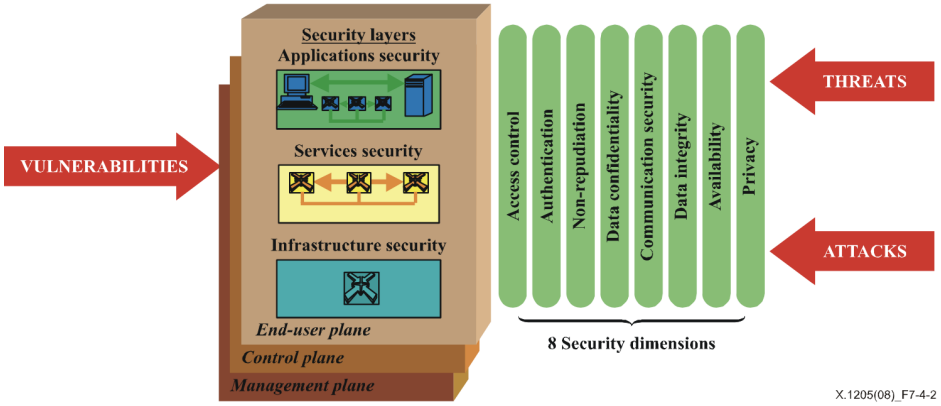

Figure 10.11: Security Planes Reflect the Different Types of Network Activities[8]

This decomposition enables the higher-level to define its own security requirements at that specific layer and enable it to use lower-level security services [8].
The layered security approach allows the development of flexible and scalable security solutions across the network level, application level, and management level for all organizations.

ii Possible Network Strategies

- Software-Defined Networks

  As the relentless pace of worldwide cities transforming into smart cities, magnificent opportunities to use IoT devices emerge, which allow themselves to communicate to each other without human intervention and to produce, process, and understand data by AI methods. Since there are countless devices working at the same time in the system and the probability of each device being attacked is greatly increased, data leakage has become a serious problem. The privacy issue of IoT becomes a top priority. To cope with it, an IoT-based smart city with a Software Defined Networking paradigm (SDN) is proposed. It is a new innovative network architecture proposed by Clean-Slate from Stanford University, which is a way to realize network virtualization. The core technology, called as OpenFlow, is to separate the control plane from the data plane of network devices, thereby realizing flexible control of network traffic, making the network more intelligent as a pipeline, and providing a good platform for core network and application innovation. And mount an efficient privacy-preserving method on top of it that manages flowing data packets of split IoT device's data.[9]. By splitting data into different packets, data can be well protected by being operated separately and sent through different routes and VPN.

  – SDN [18]
    Figure 10.12 depicts the core idea of SDN technology.

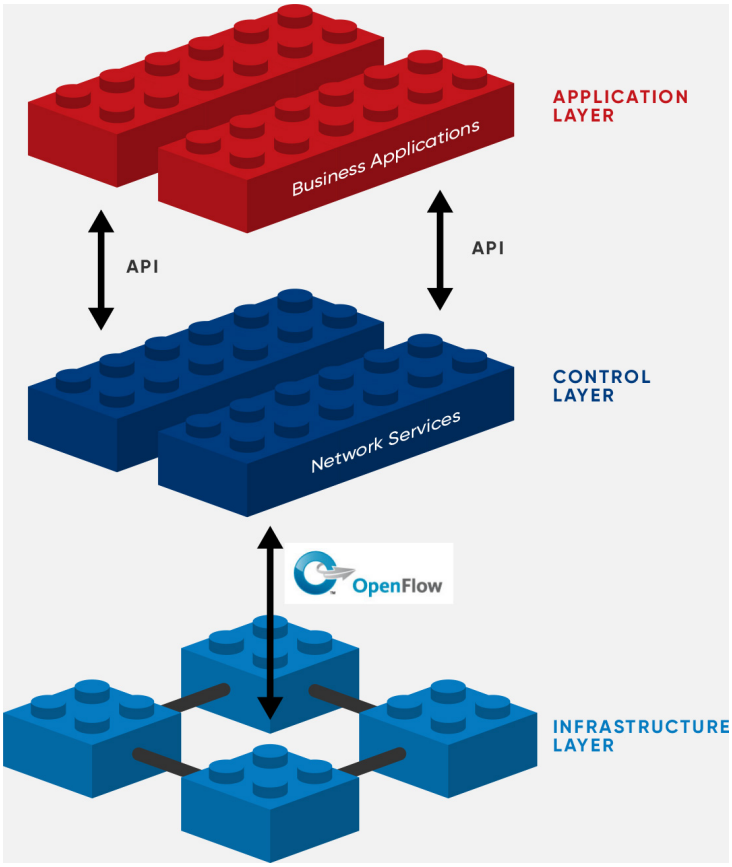

Figure 10.12: SDN

In the past 30 years, traditional IP networks have always been fully distributed. However, SDN is proposed for better and faster fulfillment of user needs in the future. The essence of SDN is the softwareization of the network, and the improvement of the programmability of the network is a reconstruction of the network architecture, rather than a new feature or function. SDN will implement various functions and features better, faster, and simpler than the traditional network architecture.

Traditional networks have three limitations. First, the flexible adjustment capability of the flow path is insufficient. Second, the network protocol is complex to implement and difficult to operate and maintain. Third, the speed of upgrading new services on the network is slow. Besides, the adjustment of the flow path needs to be realized by configuring the flow strategy on the network element. However, adjusting the flow of a large network is not only cumbersome but also prone to failure. Of course, the flow adjustment can also be achieved by deploying TE tunnels, but because of TE tunnels The complexity of the maintenance personnel is very demanding. Additionally, Traditional network protocols are very complex, including IGP, BGP, MPLS, etc.

In the traditional architecture, switches and routers have to implement the intelligence of the entire network under the control of operating 6000 distributed protocols. This means that even if only one network element adds a new protocol, all other network elements need to make corresponding structural changes. In fact, it often takes several years to add a new protocol to the network before it can finally complete the process from standardization to actual deployment. SDN makes the network programmable, which makes the network more flexible in meeting the needs of users.

As for its architecture, it separates the control function from the network switching equipment and moves it into a logically independent control environment-the network control system. The system can run on a general-purpose server, and any user can directly program the control function at any time. The control function is no longer limited to the router. The control system provides a set of APIs through which users can monitor, manage, and maintain the control system. Control plane is the element that makes forwarding decisions in the data network, such as routing protocols, routing strategies, and software and hardware resources for running these protocols on network devices. The data plane is the part that specifies the forwarding decision of the control plane, including data encapsulation and decapsulation technology, network protocol notification forwarding chip, etc.

SDN accelerates the introduction of new services. Network operators can deploy related functions through controllable software, instead of waiting for a certain equipment provider to add corresponding solutions to proprietary equipment as before. And it helps realize network virtualization. Since long term manual configuration through the command line interface has been preventing the network from moving towards virtualization.

– VPN

A Virtual Private Network (VPN), processes, encrypts and, secures all internet connections between the client and the server, but VPN uses a public link on the Internet, which is why it is virtually private. It can block specific information for example your location, the data being sent. In addition, since it straightly encrypts all the traffic between the server and client, any main-in-the-middle attack is out of possible. This can significantly improve security especially connected to a shared WLAN network when working at home[13]. An VPN model is described in Figure 10.13.

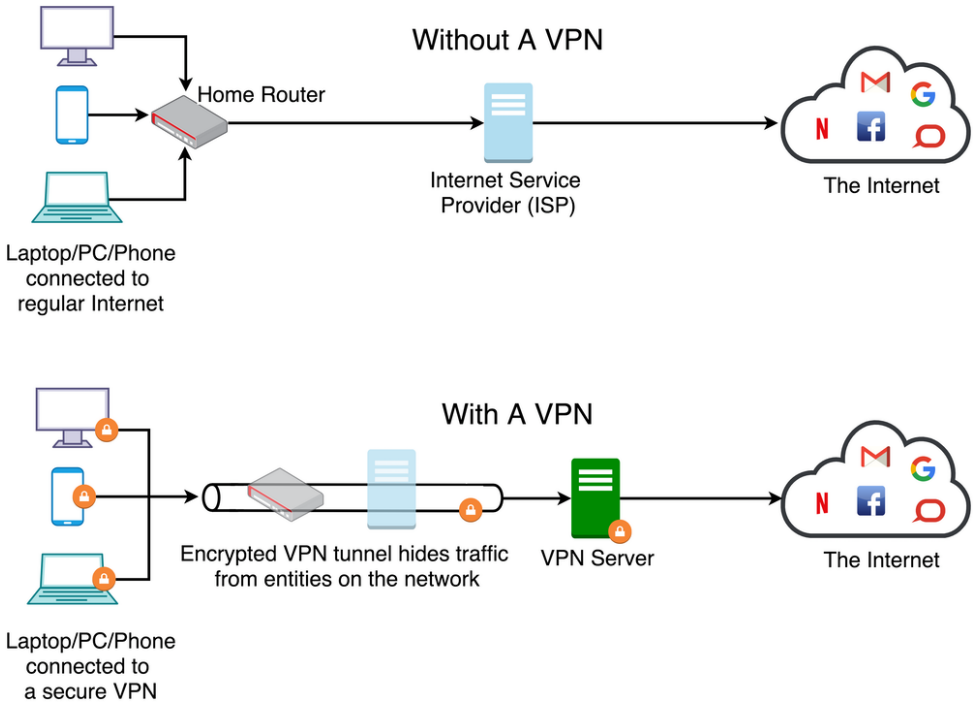

Figure 10.13: VPN

iii Artificial Intelligence and Blockchain

Artificial Intelligence (AI) and Blockchain are two different technologies and are even applied in opposite working domains. Both of them have quite a few flaws, AI suffers from privacy and credibility while Blockchain has flaws in scalability and security. Merging these two technologies has been instrumental for smart city infrastructure, which is expected to create more job opportunities and expand the labor market to improve the worldwide economic upswing after the pandemic hit. Besides, cooperation between AI and Blockchain can smooth the path to being applied to more smart city sectors such as futuristic transportation systems and communication systems.

- DistBlockNet With the increase in the diversity and number of smart devices in IoT, the issues of flexibility, efficiency, usability, security and scalability of IoT are becoming more and more obvious. In accordance with the design principles of high adaptability, availability, fault tolerance, performance, reliability, scalability and security, construct a high-performance architecture for a securely distributed IoT network. In order to satisfy current requirements, a single, dependable, distributed network architecture,along with robust connection, simultaneous safety protection, and high-speed output is motivated for authorities. Therefore, a state-of-art distributed Blockchain-based network architecture is proposed, called DistBlockNet.

  There are two major contributions. First, use Blockchain technology to provide a distributed SDN security architecture for IoT. Second, a technology to update the flow rule table in the architecture is proposed, which can safely verify and download the latest flow rule table of the IoT forwarding device [17].

  It is presumed to solve the current and ongoing cybersecurity problems by upgrading system speed, accuracy, and capacity. With the function of threat detection and data protection, it is able to fight against network threat and attacks like spoofing and Distributed Denial of Service (DDoS) attacks. It achieves this by reducing the attack time window by allowing quick IoT forwarding and downloading the latest table of flow rules. The system adapts dynamically to the incoming threat without including the administration and avoiding manual processing and approvals [14].

  Figure 10.14 provided an overview of the DistBlockNet structure. In this architecture, the SDN controller is connected by Blockchain. At the same time, each IoT network contains OrchAPP, controller, and Shelter modules. The OrchAPP and Shelter modules deal with different levels of attacks: OrchAPP functions the management and application, and Shelter functions at data level, at different part in the Figure 10.14. Through the cooperation of the two, it is possible to implement active or passive prevention based on recurring threat situations.

  More specifically, OrchAPP uses host-based software, which is deployed to the appropriate application layer execution point and divides security issues into 3 types, each with a corresponding security strategy, which are access control, data protection, and threat intelligence. Utilizing automated threat reaction control that works with management, OrchAPP can automatically protect against the identified attacks.

  On the other hand, Shelter contains the following two components: a flow control analyzer and pocket migration components. The analyzer mainly maintains the infrastructure in the network when a saturation attack occurs. It is deployed as a control application in the control layer. In contrast, the packet sends a benign network flow to the OpenFlow controller without causing overload. In addition, the migration agent of the migration component is applied to the controller application between the elements of the control plane, the data plane and the cache data plane.

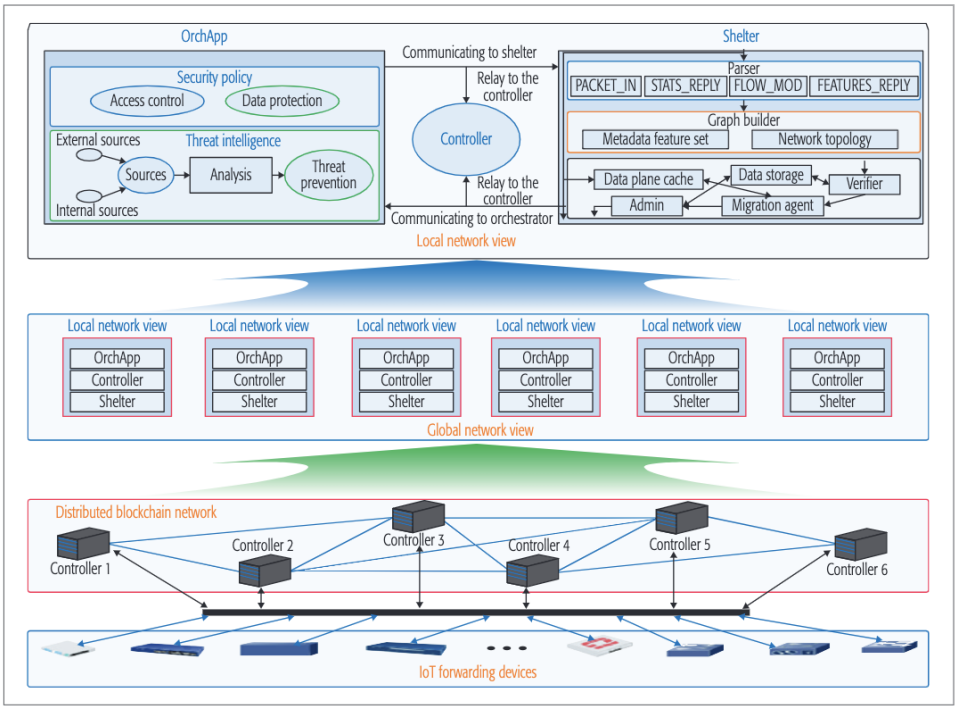

Figure 10.14: Overview of DistBlockNet.
[17]

Shelter has 3 workflows. First, build or change the network view by monitoring and parsing interaction between data packets. Second, analyze the topology metadata and status in the Open Flow message, and establish the flow incremental graph in the network. In the last step, collect metadata flow in real time. When a new flow behavior is detected, Shelter will not issue an alarm signal. Or, when the recipient detects an untrusted entity calling a modifier to the existing traffic behavior, or when the stream rejects any rules or security rules specified by the administrator, the recipient will prompt an alarm signal. Moreover, since the FLOW_MOD message from the trusted controller is generated, the Shelter will not trigger any alarms in the flow rerouting, which greatly reduces the alarms that may occur if each new behavior signal is recognized.

Figure 10.15 shows how flow rule tables updating: a) a distributed blockchain network; b) flowchart of flow rules table updating.

iv Data-Driven Security for Smart Cities

Data-driven security is considered an application of data-driven decision-making. It can be a combination of data mining, data analysis, and data visualization. By collecting, producing and processing data based on pre-defined metrics, facts, patterns, correlations, insights and knowledge can be learnt from massive data. In return, these insights can advance and adapt the systems, activities and strategies [10]. Which is a relatively low cost method to advance the cybersecurity under the bulky data information nowadays.

10.4 Conclusions

Due to the pandemic, people shift to work at home relying more on the internet working. Therefore, cybersecurity becomes an urgent problem to cope with. Covid-19 epidemic not only creates new possibilities for attackers but also uncovers the existing vulnerabilities. However, it also spikes the development of cybersecurity systems in return. On

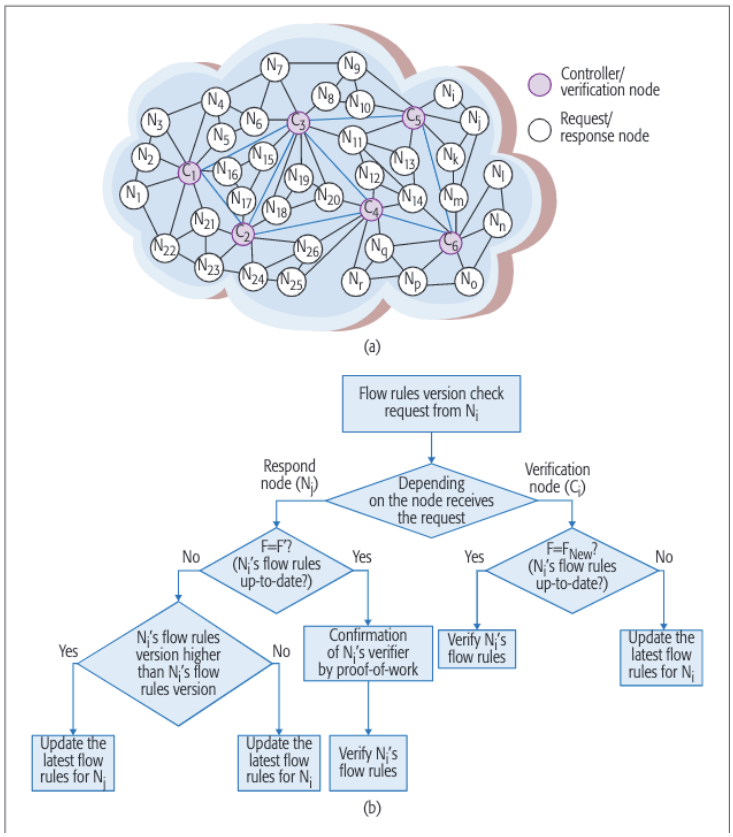

Figure 10.15: Updating Flow Rules Table.
[17]

the top of the list, it is cybersecurity awareness for organizations and individuals as well. Organizational risk management plays a fundamental role in increasing awareness of cyberprotection, which will be a substantial policy in the future. Additionally, under the current network framework, a more separated, flexible and robust architecture with higher output is motivated to fulfill the need for smart city construction. IoT and Blockchain will be the direction of future life and DistBlockNest is proposed to address the rapid increase of smart devices.

Bibliography

[1] Martin Waldburger, Patrick Poullie, Burkhard Stiller: *Guideline for Seminar Reports*, Communication Systems Group, Department of Infromatics, University of Zurich, January 2013. http://www.csg.uzh.ch/teaching/guideline-seminar-report-v05.pdf.

[2] *What is a cyberattack?* https://www.ibm.com/topics/cyberattack

[3] Yuchong Li, Qinghui Liu: *A comprehensive review study of cyberattacks and cybersecurity; Emerging trends and recent developments.* https://www.sciencedirect.com/science/article/pii/S2352484721007289

[4] F. Breda, H. Barbosa, and T. Morais, "Social engineering and cybersecurity," in *International Technology, Education and Development Conference*, vol. 3, no. 3, 2017, pp. 106–108.

[5] SailPoint: The Cybersecurity Pandora Box of Remote Work https://www.sailpoint.com/identity-library/the-cybersecurity-pandoras-box-of-remote-work.

[6] What's next for remote work: An analysis of 2,000 tasks, 800 jobs, and nine countries https://www.mckinsey.com/featured-insights/future-of-work/whats-next-for-remote-work-an-analysis-of-2000-tasks-800-jobs-and-nine-countries

[7] B.-A. B.-F. A. S.-L. Oliver M.Bühr, Oliver Hornung, "Design your home office in compliance with the law: How to comply with data protection, data security, and employment law," 2020. [Online]. Available: https://www.skwschwarz.de/en/details/design-your-home-office-in-compliance-with-the-law-how-to-comply-with-data-protection-data-security-and-employment-law

[8] International Telecommunication Union: Overview of Cybersecurity: Recommendation ITU-T X.1205 https://www.itu.int/rec/T-REC-X.1205-200804-I/en.

[9] M. Gheisari, G. Wang, W. Z. Khan, and C. Fernández-Campusano, "A context-aware privacy-preserving method for iot-based smart city using software defined networking," *Computers and Security*, vol. 87, p. 101470, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404818313336

[10] N. Mohamed, J. Al-Jaroodi, I. Jawhar, and N. Kesserwan, "Data-driven security for smart city systems: Carving a trail," *IEEE Access*, vol. PP, pp. 1–1, 08 2020.

[11] Roberto O. Andrade, Ivan Ortiz-Garces, Maria Cazares: Cybersecurity attacks on Smart Home during Covid-19 Pandemic https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9210363

[12] R. O. Andrade, I. Ortiz-Garcés, and M. Cazares, "Cybersecurity attacks on smart home during covid-19 pandemic," in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4).* IEEE, 2020, pp. 398–404.

[13] S. Larsson, M. Svensson, M. d. Kaminski, K. Roenkkoe and J. A. Olsson, "Law, norms, piracy and online anonymity: Practices of de-identification in the global file sharing community", *Journal of Research in Interative Marketing*, 2012, vol. 6, pp. 260-280

[14] T. Himdi, M. Ishaque, and J. Ahmed, "Cybersecurity challenges during pandemic in smart cities," in *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom).* IEEE, 2021, pp. 445–449.

[15] A. Al Shammari, R. R. Maiti, and B. Hammer, "Organizational security policy and management during covid-19," in *SoutheastCon 2021.* IEEE, 2021, pp. 1–4.

[16] Ahmad, Tabrez, Corona Virus (COVID-19) Pandemic and Work from Home: Challenges of Cybercrimes and Cybersecurity (April 5, 2020). `https://ssrn.com/abstract=3568830`

[17] P. K. Sharma, S. Singh, Y. Jeong and J. H. Park, "DistBlockNet: A Distributed Blockchains-Based Secure SDN Architecture for IoT Networks," in *I IEEE Communications Magazine.* vol. 55, no. 9, pp. 78-85, Sept. 2017. `10.1109/MCOM.2017.1700041`.

[18] Q. Yan, F. R. Yu, Q. Gong and J. Li, "Software-Defined Networking (SDN) and Distributed Denial of Service (DDoS) Attacks in *Cloud Computing Environments: A Survey, Some Research Issues, and Challenges," in IEEE Communications Surveys & Tutorials* vol. 18, no. 1, pp. 602-622, Firstquarter 2016, `doi:10.1109/COMST.2015.2487361`.

Chapter 11

Public Surveillance: A Benefit to Society or a Privacy Breach?

Neha Arora & Luca Guenin

As instances of public surveillance become quotidian, the debate between the benefits it brings to society and the potential breach of privacy it inherently entails intensifies. This paper will critically examine two aspects of public surveillance that we frequently encounter in the twenty-first century, namely surveillance for detection of Child sexual Abuse Material (CSAM) and video surveillance in its various avatars. To this effect, we will assess the pros and cons, the associated economics and the enabling technology for specific use cases in each of the above-mentioned two areas of public surveillance. We showed that the applications of these technologies to public surveillance purposes could be a double-edged sword for society. CSAM and second-generation video surveillance systems could help fight a just cause, yet it could also turn into a handicap when the implementation flawed and nontransparent. We therefore recommend a public debate that discusses profound purposes, operations, and regulations of the systems preceding their implementation.

11.1 Introduction

As instances of public surveillance become usual in different spheres (i.e. virtual and physical), the debate between the benefits it brings to society and the potential breach of privacy it inherently entails intensifies. This report critically examines two specific aspects of public surveillance, namely surveillance for detection of Child Sexual Abuse Material (CSAM) and camera surveillance in its various avatars. To this effect, we assess the pros and cons of public surveillance and the enabling technology for specific use cases in each of the above-mentioned areas.

The rapid growth of the Internet over the decades has produced many benefits for people worldwide by facilitating user interaction and information exchange over various platforms. However, the internet has also enabled individuals to abuse its pseudo-anonymity to distribute CSAM in the form of images, videos, and even live-streaming. In fact, experts have observed a significant increase in the distribution of CSAM online during the Covid-19 pandemic [12]. Once CSAM is uploaded on the internet, it continues to circulate online as it is not straightforward to remove it completely, causing a perpetual cycle of abuse by re-traumatising victims each time it is viewed [10]. Indeed, research indicates that victims of child sexual abuse often suffer from lifelong physical, psychological, and emotional trauma that negatively impacts their future [10].

Statistics reveal a harrowing degree of the pervasiveness of CSAM online. In its annual report in 2017, the Internet Watch Foundation (IWF) states that the number of processed reports containing CSAM rose 35% compared to the previous year [8]. 55% of the children depicted in the incidents appeared to be 10 years or younger in age and 33% of the detected content showed sexual activity between adults and children with extreme violence [8]. In 2019, almost 9 of 10 URLs containing CSAM were hosted in Europe with 71% of the reported content being hosted in the Netherlands. The year 2020 saw a 77% increase in self-generated images to a total of 68,000 of which 80% were contributed to by girls aged 11 to 13 [8]. According to a study conducted in Switzerland in 2018, which has a high digital inclusion: 99% of youngsters aged 12 to 19 have a smartphone, 86% of children aged 6 to 13 use the internet, 30% of adolescents have been approached online by a stranger with unwanted sexual intent [15]. In the same year, the FBI reported 9,000 cases of CSAM to Switzerland [15].

Historically, public surveillance was based on paper and composed of a limited number of records. Technological advances have disrupted this model and significantly expanded the boundaries of public surveillance. Initially, closed-circuit television was praised as an efficient and effective instrument to prevent the most severe crimes and facilitate their investigations. However, multiple studies show that these systems did not accomplish the desired objective: Piza [20] found evidence of a 13% reduction in crimes in areas equipped with CCTV compared to regions without camera surveillance. In addition, the effect measured depends on the type of crimes: CCTV is more effective at tackling premeditated crimes than emotional ones. Most strikingly, only active monitoring systems are more effective at reducing crimes than passively monitored CCTV. The observed trend towards computer-augmented cameras allowing for automated analysis of image flows should therefore come as no surprise. Our research highlighted the risk of this evolution, primarily due to the possibility of promptly identifying individuals captured on an image and the possibility to combine these data with other databases for monitoring purposes.

This report is organized as follows: in Section 2.2, we define some key terms to which the reader can refer to during reading. In section 2.3, we critically assess the case for and against the detection of CSAM, the international legal situation relating to CSAM, the online distribution methods for CSAM, and various technological tools and approaches available to detect and delete online CSAM. We then critically evaluate the available technology solutions to conclude that an approach that combines various technological

methods may be the most effective against online CSAM detection. The case for and against the deloyment of video surveillance technologies is critically presented in section 2.5, and followed by a discussion of good practices in their implementation.

11.2 Key Definitions

Child Sexual Exploitation (CSE): The umbrella term Child Sexual Exploitation (CSE) encompasses child exploitation in all its forms including Child Sexual Abuse Material (CSAM) or Child Sexual Abuse Imagery (CSAI).
Child Sexual Abuse Material (CSAM): The term "child pornography" - representing visual depiction of sexually explicit conduct involving a child - is used synonymously to Child Sexual Abuse Material (CSAM) in the legal context. However, it is argued in literature, that CSAM more appropriately reflects the extent of sexual abuse and exploitation of children because it is never consensual [9].
International Centre for Missing and Exploited Children (ICMEC): Established in 1999, the International Centre for Missing and Exploited Children works in over 120 countries to power global searches for missing children and to disrupt the economics and mechanics of Child Sexual Exploitation (CSE) [7].
Internet Watch Foundation (IWF): Established in 1996 and funded by the European Union and member companies from the online industry, the Internet Watch Foundation (IWC) works in partnership with the internet industry, police forces, governments, and charities across the world to minimize online CSAM [8].
National Center for Missing and Exploited Children (NCMEC): Established in 1984 in the United States of America by the Congress, the National Center for Missing and Exploited children is the nation's non-profit clearinghouse and comprehensive reporting center for all issues related to the prevention of recovery from child victimization. NCMEC leads the fight against abduction, abuse, and exploitation" [11].
Closed-Circuit Television (CCTV): It is a system that transmits television signals to displays and is usually deployed in shops and public spaces to avert crime [35].
Facial Recognition Technology: It is a biometric technology that uses images to identify a person [22].

11.3 Detection of Online Child Sexual Abuse Material (CSAM)

11.3.1 Critical Assessment

As indicated by the above-mentioned 77% increase in self-generated images in the 2020 IWF annual report, the expansive use of social media allows not only adults, but also minors to share their self-produced content online. Although the hugely popular music-based social media platform TikTok states that users under the age of 13 are not allowed to use it, its age-verification system can easily be bypassed by entering a false birthdate [14]. As a result, it features many videos of children performing to sexually explicit songs that have been uploaded by child users who do not make their accounts private or disallow contact from strangers which, in turn, can lead to online grooming [14]. While journalists have exposed active communities of TikTok users who appear to be soliciting nude images from children, minor users have complained about repeated solicitation for sexualized images [14]. In such cases, however, public outcry against CSAM, often devolves into calls for more parental responsibility and internet safety programs for children [14].
Social media not only makes CSAM easily available to millions of users, but also enables them to view CSAM on the internet without directly downloading content to a personal computer thereby leaving no trail of their viewership. For example, despite stated

policies against child abuse, many users have uploaded videos that qualify as CSAM on YouTube. In 2019, YouTube user Matt Watson exposed how the YouTube recommendation system-"the machine learning process that suggests and curates videos for users" [14]-linked together self-created videos of young children engaged in various physical activities [14]. Thus, once the system detected a user's preference for videos with young children, it would re -contextualise videos- which were neither illegal nor CSAM content- to generate a playlist of similar content thereby forming what Watson called a "softcore paedophile ring" [14]. As a result of this scandal, it became paramount for YouTube to identify viewers who leave disturbing comments on such videos in order to delete inappropriate comments and/or block the related user accounts [14]. Technological use cases for CSAM detection/deletion are discussed in Section 3.4.

Critics of online CSAM detection (and removal) base their claims on libertarian views that internet regulation is impossible, unworkable, and most of all, unwanted. Defined as a "distinct political stance and moral psychology whose guiding principle is the freedom of individuals, in particular from interference by the state", libertarianism prioritizes concern for individual liberty from control and regulation over altruistic moral values and social responsibility [14]. According to Salter et al., emerging from the socio-political foment of the 1960s and the 1970s, libertarianism and new technologies played a formative role in the culture and practices of the Silicon Valley [14].Therefore, promoting the view that the products of internet and technology companies embody personal and collective freedoms, online CSAM is regarded as a natural artefact beyond the control of any company or government by the critics of CSAM detection.

Moreover, CSAM detection is often regarded by critics as an attack on end-to-end encryption, an expensive endeavour, a violation of free speech, a breach of privacy, and as a limitation on seeking access to information. Technology companies are concerned that they will have to weigh the risk associated with offering end-to-end encryption against victims of child-abuse since privacy is a highly desired feature on all platforms [5]. Small and medium sized companies fear that their reputations will be tarnished if news of CSAM content on their platforms goes public [5]. Furthermore, CSAM detection (and removal) requires a considerable financial investment that small and medium sized companies shy away from [5]. Platforms like Twitter, described by a senior executive as the "free speech wing of the free speech party", have been built upon the idea of free speech which is inherently counterintuitive to measures such as CSAM detection [14]. The 2020 announcement of Apple's Chief Privacy Officer that the company was scanning images on iCloud in order to identify CSAM, led to widespread breach of privacy concerns [5]. Using filtering and blocking techniques against CSAM has also been criticised as wide censorship and suppression of free speech as well as for limiting access to information [5].

11.3.2 Legislation: International Comparisons

As countries worldwide struggle to control the online dissemination of CSAM, ICMEC emphasizes that technologies to detect online CSAM are more effective if the legal framework allows law enforcement agencies to penalize perpetrators engaging in the production, distribution, or possession of CSAM [7], [10]. Due to the geographical illimitability of the internet, such legislation has to be internationally implemented [7], [10]. In their 2018 Annual Report, the ICMEC evaluated 196 countries according to five criteria: a) existence of national legislation regarding CSAM b) definition of CSAM c) criminalization of technology-facilitated CSAM offenses d) knowing possession regardless of the intent to distribute e) Internet Service Providers' (ISPs') responsibility to report suspected CSAM to law enforcement or another mandated agency [7], [10]. In 2018, 118 out of 196 countries fulfilled at least one of the five criteria, but only 21 fulfilled all five criteria. 16 out of 196 countries do not have any legislation at all that specifically addresses CSAM and 62

lack in an adequate coverage of criteria. Despite the gaps, the situation has improved by 45% since 2006 when only 27 of 184 countries had sufficient legislation [7], [10]. Merely viewing CSAM online is not an offense in many countries which require possession of CSAM for prosecution [10]. Countries like the United States of America, Canada and the United Kingdom have CSAM clearing houses in the form of NCMEC, Canadian Centre for Child Protection, and the IWF respectively, that technology companies can work together with to curb online CSAM. However, most countries do not hold technology companies legally liable for the dissemination of CSAM on their platforms, but leave it instead to the companies to develop their own safety standards [10].
In Switzerland, the Swiss Foundation for the Protection of Children is a national organisation that aims to protect children from violence and sexual abuse. According to their report, "adults involved in illicit pornography, including child sexual abuse material", could be sentenced to five years in prison, although the conviction rate remains dismally low [15]. Moreover, minors over the age of 16 go unpunished if they produce, possess or consume mutually consensual pornography, while minors under the age of 16, who record themselves in sexual acts and thus produce illegal child sexual abuse material, are criminally liable [15]. Thus, while Swiss law does address CSAM, and contains paragraphs for prosecuting adults for procurement, possession or dissemination of CSAM, minors over the age of 16 are not liable to any penalty for producing CSAM with mutual consent, thereby introducing a legal channel for distribution of online CSAM [15].

11.3.3 Distribution Methods

Online CSAM exists in several formats-images, videos, live-streaming of child sexual abuse-and is disseminated in the internet through a multitude of channels. This section provides a summary of such platforms listed by Hue-Eun Lee et al. in their article [10]. The next section focuses upon detecting and deleting online CSAM distributed through these platforms.
Peer-to-Peer Networks (P2P) [10]: Vast global file-sharing systems with billions of users around the world, Peer-to-Peer (P2P) are typically known for acquiring music, films, books, and other digital material for free. Users join a P2P network by installing software that connects them to computers of other users (peers) in the network so that they can exchange files. Accessible publicly and free of cost, P2P networks provide an expansive platform to exchange CSAM. Furthermore, P2P networks are more anonymous as they do not require servers and can transmit CSAM without any oversight from electronic service providers.
Darknet [10]: Used to refer collectively to covert networks that allow encryption and anonymity for users' activities, darknets are not indexed by search engines and are accessible only via authorization or special software. The anonymity and security of the darknet makes it very attractive for CSAM traffickers.
Websites and Search Engines [10]: First-time CSAM users use search engines to scan for CSAM content. Law enforcement agencies in the US differ about the effectiveness of search engines in regard to CSAM as some argue that such content is easy to find [1], while others [13] argue that monitoring by the ISPs and law enforcement agencies reduces the available of CSAM on the normal web. Consumers access CSAM not by directly contacting the supplier, but by accessing CSAM uploaded onto a website. Deterrence efforts by popular search engines such as Google and Microsoft, which applied technical controls on their platforms that included the removal of CSAM from their indices enhanced filtering of queries that were exclusively CSAM-related, and deterrence messaging to users when their queries were strongly associated with CSAM, were found to be very successful.
Mobile Devices [10]: Both CSAM production and consumption has increased with the increase in the use of devices such as smartphones, tablets, digital cameras, and laptops.

Social Media [10]: As discussed in section 3.1 Critical Assessment above, social media enables widespread dissemination of CSAM.
Other Distribution Methods [10]: Although such methods do not account for a large number of CSAM being circulated or consumed online, CSAM may be uploaded, viewed, or distributed via channels such as e-mail, instant messages, newsgroups, bulletin boards, or chat rooms.

11.3.4 CSAM Detection and Removal Approaches and Tools

This section reviews selected major technological solutions, with specific use cases for each, that can be used for detection (and removal) of online CSAM.

i Image Hash Database

Generally used for encryption and verification of sensitive data, image hashing is the primary technology used for detecting CSAM, for example, in THORN's Safer, Apple's CSAM Detection, and Microsoft's PhotoDNA. In this method, each of the images or videos previously identified as CSAM are assigned a unique hash value, derived from a mathematical algorithm, that serves as a cryptographic fingerprint for the underlying image and are then stored in a database [10]. Suspected CSAM images can be then verified against the database of known CSAM images for CSAM detection (and removal) [10].
THORN's Safer: In order to help small and medium sized technology companies-that lack financial resources and expertise for CSAM detection-to detect and remove CSAM from their platforms, THORN, a non-profit organisation, has developed a commercial product called Safer [5]. Used by technology companies such as Flickr, VSCO, Slack, Medium, Vimeo, and GoDaddy (to name a few), Safer identifies known and unknown CSAM at the point of upload with perpetual hashing and machine learning algorithms [16]. It then queues flagged content for manual review with content moderation tools that have been designed to safeguard employee wellness [16]. Verified offensive content is then reported to NCMEC in accordance with regulatory requirements in the United States of America [16]. The hashes of new, verified CSAM are then added to the Safer community databases [16].
Apple's CSAM Detection: According to the company website, Apple is introducing child safety features in three areas: first, new communication tools will enable parents to play a more informed role in navigating communication online through Apple's Messages app; second, iOS and iPadOS will use new applications based upon cryptographic image hashing technologies to detect CSAM in images stored in iCloud; and lastly, updates to Siri and Search will provide users expanded information and help if they encounter unsafe situations and will also intervene when users search for CSAM-related topics [1]. The Messaging app, for example, will warn children and their parents while sending and receiving sexually explicit photos. While receiving such content, the photo will be blurred and the child will be warned, presented with helpful resources, and reassured that it is okay to not view the content should it choose not to. If the child does choose to view the content, the child can be told that its parents will be informed to ensure its safety. A similar procedure will be followed if the child tries to send a sexually explicit photo [1].
Apple's CSAM detection system is based upon an on-device (that is, not on iCloud) matching using a database of known CSAM images provided by NCMEC and other child-safety organisations [1]. This database is further transformed by Apple into an unreadable set of hashes and secured safely on the user device. The hashing technology, called NeuralHash, generates a unique cryptographic fingerprint of the image such that either only an identical image or the same image with different size or quality parameters will have the

same NeuralHash value [1]. Before an image is uploaded to iCloud, it is matched against the on-device database of CSAM photos using a cryptographic technology called private set intersection, which determines if there is a match without revealing the result [1]. The device creates a cryptographic safety voucher that encodes the match result [1]. It also encrypts the image's NeuralHash and a visual derivative [1]. This voucher is uploaded to iCloud Photos along with the image [1].
Using another technology called threshold secret sharing, the system ensures that the contents of the safety vouchers cannot be interpreted by Apple unless the iCloud Photos account crosses a threshold of known CSAM content [1]. Only when the threshold is exceeded does the cryptographic technology allow Apple to interpret the contents of the safety vouchers associated with the matching CSAM images [1]. The threshold is selected to provide an extremely low (1 in 1 trillion) probability of incorrectly flagging a given account [1]. This is further mitigated by a manual review process wherein Apple reviews each report to confirm there is a match, disables the user's account, and sends a report to NCMEC [1]. If a user feels their account has been mistakenly flagged they can file an appeal to have their account reinstated [1].
Microsoft's PhotoDNA: Developed in 2009, Microsoft's PhotoDNA is used by many organisations worldwide to detect and report CSAM [10]. Building upon the image hash technology, PhotoDNA creates a unique digital signature for an image to enable the detection of copies of the same image [10]. The original application has been extended to Cloud and has been further developed in partnership with IWF to also identify known CSAM videos [10]. The fundamental technology in video identification same as in image identification: key frames of the video containing CSAM are selected and receive a unique hash through PhotoDNA; these can then be compared against frames from other suspected CSAM videos [10]. The review process of PhotoDNA includes reporting, followed by an account suspension, content removal, and creation of the hash [10].

ii Web-crawler

Web-crawlers, also known as search bots, are designed to automatically browse websites and collect data about them based upon a set of pre-defined criteria [10]. By using specific characteristics of websites hosting CSAM, website content can be downloaded and indexed into a database [10]. They also exploit the online communities of criminal networks through which many websites are interconnected [10]. Examples of proactive crawler-based search for CSAM include IWF and Project Arachnid.

- IWF: Using a combination of intelligent crawler with Microsoft's PhotoDNA, IWF can crawl several million webpages per day searching for CSAM, scans images and videos while matching them against the IWF database [10]. The objective of the crawler is to protect victims from revictimization through images being distributed across the internet [10]. The crawler is often used in collaboration with analysts, law enforcement agencies worldwide, and sister hotlines [10].

- Project Arachnid: Built by the Canadian Centre for Child Protection, this web-crawler uses the same combination of crawler and PhotoDNA as the IWF web-crawler [10].

iii Detection Based Upon Filename and Metadata

Researchers in the area have proposed CSAM detection solutions based upon filename and metadata. For example, Pereira et al. propose statistical models that compute the likelihood that a file path is associated by a CSAM file [12]. By training and comparing several machine-learning based models that analyse metadata from file storage systems

and determine a probability that a give file contains CSAM, analysing the robustness of their models against other techniques used for CSM detection, and using a real-world dataset, they propose a system for CSAM identification based solely upon file paths [12]. Pancheko et al. propose a system that is able to detect CSAM on P2P networks encompassing a machine learning technique for language processing that can recognize queries of offenders or filenames in the data of P2P networks that contain CSAM [10]. To distinguish between files containing CSAM and regular porn files, features from metadata of known CSAM files are extracted by segmenting and normalizing the text. Then a text classification method uses statistical machine learning classifiers or a Regularized Logistic Regression to separate regular files form those containing CSAM [10].

iv Visual Detection

Visual detection methods are able to identify previously unknown CSAM and are crucial in recognizing newly emerging content online in, for example, P2P networks. In order to detect CSAM and distinguish it from other content like legal pornography or normal content featuring children, specific features of CSAM are extracted [10]. Researchers have developed several methods of visual detection of CSAM [10]. For example, the CSAM detection application proposed by Sae-Bae et al. applies an updated set of facial features and a skin tone filter that improves upon existing skin tone filters in terms of robustness and speed [10]. In this model, low-level and high-level facial features as well as feature of skin tone are extracted and used to design classification systems that can distinguish between CSAM from non-CSAM content with an accuracy rate of 74.19Deep learning architectures apply different methods like local descriptors and Convolutional Neural Networks to analyse facial features in images for classification [10]. As an example, Google's AI Implementation for detecting CSAM uses deep neural networks to detect CSAM photos and videos [10]. The classifier can identify previously unknown CSAM content too [10]. Developed in partnership with IWF, this implementation prioritizes the most likely CSAM content for review [10].

11.4 Discussion: CSAM

By reviewing literature on CSAM detection, this report critically assesses the case for and against the detection of CSAM, the legal framework, the online distribution methods for CSAM, and various technological use cases for CSAM detection and/or removal. Technologies based upon image hash databases-such as THORN's Safer, Apple's CSAM Detection, and Microsoft's PhotoDNA-represent the most widely used methods for detecting online CSAM. Characterized by a low false detection rate, these systems have a proven track record for detection (and removal) of online CSAM. However, they are largely focused upon images or photos. The detection of CSAM videos is a relatively new feature in, for example, Microsoft's PhotoDNA, while live-streaming of CSAM is largely outside the purview of these technologies. CSAM detection using this method is limited by the databases of images/videos of known CSAM content made available by NCMEC or other child-safety organisations. Detection of previously unknown CSAM content using this method still remains a manual activity to be undertaken by humans who have to be trained to identify CSAM content and deal with the psychological impact of viewing CSAM content.

Moreover, Apple has to overcome significant legal hurdles concerning data privacy for its CSAM Detection system especially because of its on-device scanning of photos. In an article on the prospects for this system in Europe, Cobbe highlights that EU law in its current form would require that Apple obtain an opt-in, rather than opt-out, consent of individual iPhone users for on-device scanning [4]. Furthermore, there would have to be

a real option for users to refuse consent without being denied access to iCloud services [4]. However, these barriers maybe removed by future EU or Member State legislation just as similar potential barriers for automated CSAM detection by certain messaging services have already been removed [4]. Nonetheless, EU's data protection (GDPR) and ePrivacy laws highlight people's concerns about Apple's CSAM Detection system on the fundamental right to privacy.

Web-crawlers can be made more efficient in detecting online CSAM by incorporating keywords that are known to be associated with online CSAM [10]. This is contingent upon understanding the file names used or the codes used to identify people seeking CSAM [10]. While incorporating keywords clearly has the advantage of finding content that has not been identified before, and even of protecting children who are currently being abused, a disadvantage can be that of false positives especially involving legal adult pornography [10]. Techniques based upon filename and metadata detection also face a similar challenge-that of the ability to correctly distinguish CSAM material from other legal material. An advantage, however, of these techniques is that they do not require maintaining a database of CSAM images for comparison thus providing a medium agnostic classifier, whereas machine learning based visual detection techniques for CSAM detection can be constrained by legal restrictions on maintaining databases containing CSAM.

In a nutshell, CSAM detection applications are likely to yield best results when multiple approaches are combined together, for example, in IWF's use of the web-crawler with Microsoft PhotoDNA. Furthermore, deep learning techniques are more likely to detect previously unknown CSAM. Due to the legal restrictions on maintaining CSAM databases in most countries, researchers in the area have to work together with law enforcement agencies to make progress.

11.5 The Rise and Evolution of Public Video Surveillance

11.5.1 Origin of Video Surveillance

Embracing novel technologies to structure, harmonize and control human interactions on the market and social circle has always constituted an integral part of the evolution of society [18]. Historically, public surveillance was print-based and was made up of records and reports [17]. Paper-based surveillance is scalable, but its nature ultimately confined this extension [17]. The information was manually gathered, stored, updated, and employees had to define links between information sets, which was cumbersome and prone to errors [17]. The emergence of closed-circuit television (CCTV) symbolized a historical transition from paper to visual surveyance and the permanent expansion of public surveillance networks. Nowadays, electronic security and safety tools are the cornerstones of more complex technological infrastructures, of which public-space video surveillance systems are the most widespread yet contentious situational crime prevention (SCP) strategy [18][20].

Officially, the primary rationale of the closed-circuit television system has been crime reduction and prevention, anti-terrorism, and public safety [17][18]. Policymakers and police authorities have often praised these systems for facilitating decisions on deploying patrols, identifying suspects, and deterring crimes and socially undesirable conduct [17]. Williams and Johnstone [19] listed four types of crime or socially problematic behaviors that CCTV can reasonably help eliminate:

- Serious felonies.
- Behaviors preceding criminality like disorderly confrontations.
- Minor crimes, including nuisance offenses

- Unpleasant yet not a violation of the law per se, for instance, drunkenness.

Another scholar [21] argues that the emergence of CCTV in the UK stems from the "surface plausibility" of the instruments and of the political benefits politicians expected from "being seen to be doing something visible to widespread concerns over crime". Others [17] assimilated the widening adoption of video surveillance to both fashion and desperation. Irrespective of the underlying motives and beyond the public and political perception of CCTV effectiveness, these surveillance system have flourished around the globe in both autocratic and semi-autocratic countries [22].
There are conflicting dates of the initial appearance of public video surveillance systems, but it is usually traced back to the 1960s in England [17]. This buoyant market has since been growing resiliently. By one estimate, 9.9 million professional surveillance cameras were sold globally in 2006, increasing to 106.4 million in 2016 ([23]). In addition, China has morphed into the world's largest domestic market for video surveillance equipment, accounting for 41% of world revenue in 2016 [23]. Countries like the USA, the UK, Japan, and much of Western Europe exhibit signs of a matured market with an extensive network of cameras, meaning that demand is typically for replacement rather than expansion [23]. In contrast, demand is growing faster in other regions like India and Latin America, where the penetration rate is much lower [23]. As a summary, Figure 11.1 depicts global professional video surveillance market in 2016.

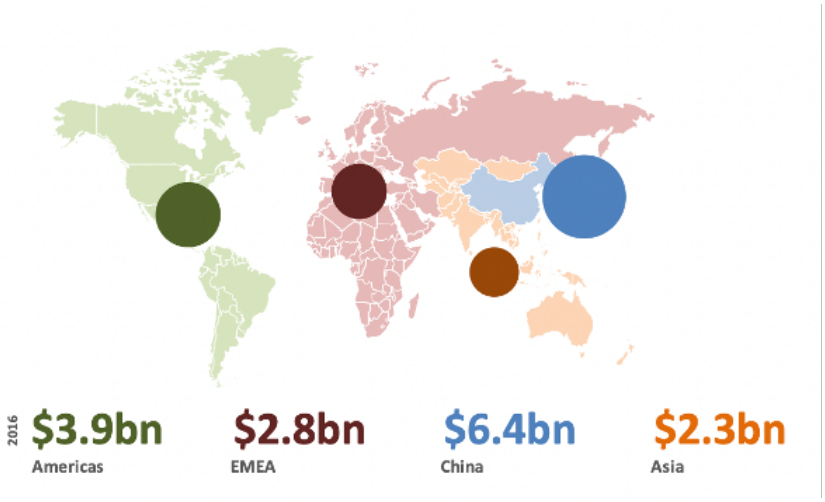

Figure 11.1: Global professional video surveillance market. Retrieved from [23].

11.5.2 Video surveillance: Much Ado about nothing?

Despite the craze around CCTV, scholars have repeatedly pointed out the discrepancy between the public perception of their effectiveness against crime and the results of empirical studies.
In the early years of a new century, Williams and Johnston [19] considered CCTV as "an academically ignored topic". When it attracted more attention from researchers, a majority of studies quantifying the crime prevention effects of CCTV still suffered from flawed methodology [20], thus inhibiting any causal inference. The usage of rigorous designs, including quasi-experimental design, randomized field trials, and matching, has gained momentum with time [20], providing precious insights on the effectiveness of these systems.
In the meta-analysis, Piza et al. [20] considered only research design that involved at least a before-after measure of crime in treatment and comparable control areas. The results of the analysis show that CCTV is associated with small yet significant prevention

in crime [20]. Crime drops by roughly 13% in areas equipped with surveillance cameras compared to control areas [20]. Convergent to previous evidence, the authors highlighted that CCTV is not effective against all types of crimes. CCTV was not associated with a significant reduction in violent crime or disorder, but it had a large effect on drug crime with a reduction of approximately 20% [20]. One theory to rationalize this evidence suggests that by increasing the costs of crimes and socially offensive behaviors, public video surveillance could deter more deliberate felony including thefts, but may have limited influence on emotional crimes [17].
For their part, Williams and Johnstone [19] clinched that CCTV is most useful in the control of minor crimes and unpleasant behaviors such as littering, public urinating, but far less for serious criminal activities such as terrorism. In Switzerland, the experience of Crans-Montana, a pioneer in the Swiss landscape, also supports this view: with a network of more than 100 cameras, the cantonal police has majorly used the system to identify individuals causing minor offenses, of which the frequent smashing of flowerpots. There are conflicting interpretations of this phenomenon. On one side, Skogan [24] presents a positive interpretation: the commitment to monitor and combat minor crimes causing the upheaval of community life may result in improved order maintenance, in turn yielding a decline in serious crimes. On the other side, Gates [25] is more critical and defines the phenomenon "function creep", where the deployment of public video surveillance for minor offenses purposes phases off over time and substitutes for mass surveillance intentions. This topic is explored in more detail in the following sections.
Another key empirical observation of [20] was the differential effects of actively and passively monitored CCTV systems. Indeed, the former showed evidence of a significant reduction in crime, while passively "managed" schemes did not. Nevertheless, a passive system, as set in Crans-Montana, may prove particularly useful for generating forensic images assisting authorities in their investigation. A major reason favoring the implementation of passive monitoring is the greater commitment to resources to watch the video streams live. In fact, in urban areas, home to thousands of cameras, the cost of this strategy is a point of contention [24].

11.5.3 Computer-Augmented CCTV

In light of dumb first-generation CCTV's ineffectiveness under certain scenarios to detect crimes and surreptitious behaviors, combined with a cultural preference for technical answers to social problems, rationalized the desire to decrease CCTV's passivity[24]. The idea of a transition to smart second-generation CCTV is not novel with early research already conducted in the 1990s, yet without being subject to public debate about the inherent threats of these systems [17]. The benefits of computer-augmented video surveillance systems have, however, been recognized a long time ago [17]. The underlying idea is that live or archived video streams are automatically analyzed by the software to identify, sort, and track predefined objects and behavior patterns [23]. Technical constraints have long limited their development, but at least two trends have reshaped the landscape:
Deep learning : Deep learning (DL) is a well-known method for the application of machine learning, which originates in the 1980s. A salient feature that differentiates DL from traditional computer vision technologies resides in the usage of artificial neural networks to mimic the human brain. The usage of convolutional neural network addresses the long-held criticism levied against conventional video analytic products, namely the algorithm's inability to classify objects and identify human behaviors. The capacity of deep learning algorithms to approach intuitively a case, in a similar fashion as human being would, has contributed to greater accuracy and marked a step forward for the research community, and the society in general, as video analytic capabilities have progressively delivered on some of the lofty claims made in the past [23]. As such deep learning video analytics have

offered enhanced capability to recognize shoplifter, someone smashing a car-window, or the outbreak of a brawl [24].
Another advantage stemming from deep learning techniques is the speed of execution, a feature that can make the balance tips towards success or failure in public safety policy. Increases in computing power have been central to the sound democratization of deep learning alright, as have advances in big data. The dependence of deep learning algorithms on computational power is a source of concern for future development. As stressed by Thompson et al.[26]:

> Deep learning progress will be constrained by its computational requirements and that the machine learning community will be pushed to either dramatically increase the efficiency of deep learning or to move to more computationally-efficient machine learning techniques.

Network equipment : IP cameras, also known as network cameras, have gradually replaced analog cameras in recent years[1]. By one estimate, while network cameras accounted for only 9% of the professional video surveillance camera in 2006, this share rose to about 59% in 2016 [23]. Both systems transfer image streams to the desired location, but they do so slightly differently. Analog cameras transform the video signal into a format readable by televisions, or Video Cassette Recording, while IP cameras transform the video signal into IP packets transferred through the internet to a network storage device. An advantage of IP cameras is their higher resolution than their analog peers, resulting in a larger field of view and offer sharper, more clearly defined pictures [23]. Such benefit is non-negligible when it comes to facial recognition and video analytic. There is also an increasing proportion of network cameras being embedded with sophisticated video analytic and wide dynamic range [23].
The second-generation camera development addresses three technological goals: identifying individuals of interest, behavior recognition, and classifying objects such as weapons or abandoned parcels [17]. The ultimate technological purpose is the creation of intelligent centralized or decentralized camera systems with algorithms directly embedded on the camera or a central device, respectively. Both options present advantages and disadvantages, notably in computing power, but have shown strong growth in recent times. For example, cameras with video analytics capabilities will account for 53% of the market by 2022 compared to 23% in 2017. Similarly, recorders with video analytics is expected to rise by a factor of 5 and top to 10% of the market [23]. These findings are summarized in Figure 11.2. This might support Surette's [17] opinion regarding the prevalence of decentralized camera systems, "where in additional of being part of a larger network, each camera will have independent analysis capabilities".
The next section will review two interesting applications of video analytic that may serve the public good or threaten it. First, a brief introduction is given regarding facial recognition. Then a first case study explores its development in the field of public surveillance, the benefits that may be harvest from it, and potential concerns. A second case study explore the not so discussed driver monitoring systems (DMS) that will soon be available on new cars sold in the European Union.

i Facial Recognition Technologies: A primer

Unlike traditional CCTV systems, which have been a pillar of polices forces in the last three decades, video surveillance systems with facial recognition capabilities are more intrusive [22]. The basic idea is to utilizes cameras to cross-reference recorded or live

[1]There are basically two types of video surveillance: CCTV and Network Camera. Theoretically, one should abstain from referring to CCTV when IP cameras are used, but such distinction is rarely made in researches.

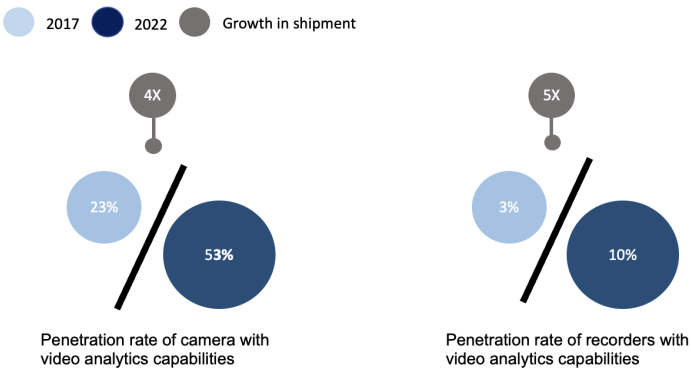

Figure 11.2: Global forecast for Video Analytic Market. Adapted from [23].

stream images of subjects to footage from a database [22][28][31]. The usage of biometric technologies is, however, not restricted to identification purposes, but also to analyze aggregate demographic trends and analyze sentiment through mass scanning [22]. Further, there is even the use of Wireless signals such as IEEE 802.11 family protocols [39], light-reflection technologies such as LiDAR (Light Detection and Ranging) [41], or Bluetooth Low Energy (BLE) [40] to track mobile device signals.

They can extract distinctive features to produce a precise biometric mapping of human faces without requiring active participation from subjects. Therefore, it offers a wide range of applications from static "mug-shut" to dynamic, uncontrolled face identification in encumbered background [sasan]. In public safety and law enforcement, facial recognition technologies have been implemented in surveillance systems, trailing criminals, and detecting fugitives [22][31]. In these non-consent scenarios, the accuracy rate of FRT is hampered by the real-life environment's conditions. For instance, researchers have identified pose problems, illumination changes, different facial expressions, and age factors as one of the most salient challenges of these systems [Face reco, holistic]. In business, facial recognition has established itself as a reliable mode of payment, facilitated check-in at hotels, or contributed to enhancing security in shops and minimizing shoplifting [29][33]. In the transportation sector, facial recognition has contributed to more efficient check-in processes at airports, tracking unlicensed drivers and jaywalkers [30], and potentially identifying drunk drivers [32]. Scholars have also pointed the use of facial recognition in education to improve school attendance and students' attention or campus security [30].

For applications in the field of security, FRS should often incorporate the following features: the system must be able to process in real-time both images and videos, show evidence of high accuracy under different illumination and weather conditions, to be independent of the subject gender, ethnicity, hair or beard, and finally, it must work with different head orientation [45]. Although many approaches have been adopted to address these challenges, existing face recognition systems all rely on three basic steps: (1) face detection & normalization, (2) feature extraction, (3) face recognition [45].

Face Detection & Normalization: FRS begins the analytical process by detecting and locating the human face in the input image. Under certain circumstances, illumination changes, weather conditions, or facial expressions can pose significant challenges in the face detection step. However, pre-processing operations have shown outstanding usefulness in tackling these challenges.

Feature Extraction: The purpose of this step is the extraction of facial features, including the nose, mouth, or eyes, with the geometric distribution and represent them as a set of features vector. A subject face is set apart by its structure, shape, and size, so existing systems exploit the resulting variations in size and distance to uniquely identify a person. Various approaches exist for the extraction of features dependn

Face Recognition: In the step, the features extracted in the previous stage are compared to faces stored in a database.

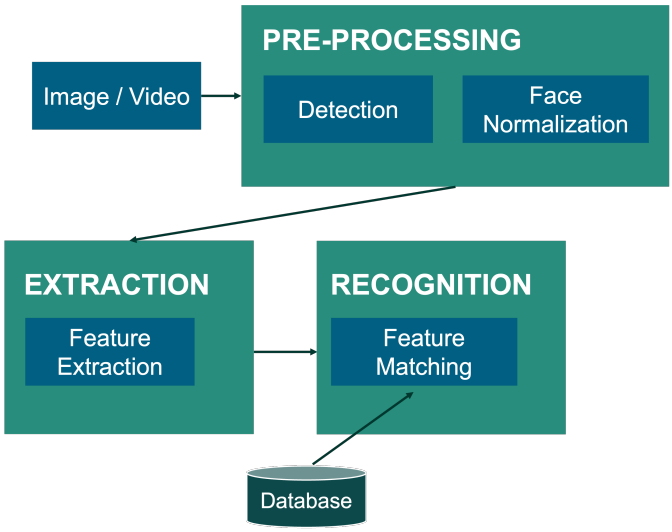

Figure 11.3: Face Recognition Block Diagram. Adapted from [45].

ii Facial Recognition Technologies: Concerns

Video surveillance and facial recognition technologies promise to improve lives and allow for more efficient operations in society. Still, it comes at the costs of controversy and concerns for privacy and freedom, security, accuracy, and bias [17], [31], . Threats to privacy and liberty are inherited from dumb first-generation CCTV, although probably amplified [17], while concerns regarding algorithms' accuracy and bias are unique to second-generation CCTV.

Privacy and freedom:

Public camera surveillance shapes our ability to define ourselves through complex channels. Tatslitz [36] observed that privacy is an integral part of personal development because it ensures individuals are free to define themselves and present to others only the information about themselves that we are willing to disclose. That, in turn, grants individuals the opportunity to expose different aspects of their nature under different contexts: one can wear a certain mask at the job, another at home and disclose a different aspect of our personality at social events [36].

Scholars noted how public camera surveillance and recording could regulate individuals' behavior in the desired direction. For example, Richard Wasserstrom [38] stressed:

> No matter how innocent one's intentions and actions at any given moment... persons would think more carefully before they did things that would become part of the record. Life would to this degree become less spontaneous and more measured

The pervasive effects of video surveillance may go beyond restraining our spontaneity. For example, the practice of some habits is not unlawful but may demand some form of anonymity to protect the perpetrator from being judged by its peers [37]. It suggests that self-definition stems partly through our daily routine occurring outside our home, and the constant monitoring of these moments could amount almost to physical harm of the self [37],[36].

Apart from challenges to liberties, the proliferation of computer-enhanced CCTV challenges privacy through security breaches on their systems. Given the sensitive nature of the databases containing citizens' personal information and biometric information, there

are prone to attacks by criminal organizations, which could whistle data, erase, or append data [31]. Any breaches could have dramatic implications for victims. Facial recognition enhanced camera surveillance, when deployed with deficiencies, could fail to safeguard public safety and threaten citizens' freedom while leaving the door open for abuses [31]. Therefore, authorities must strike the right balance between public safety and privacy-infringing surveillance systems.

Accuracy: A significant source of concern of FRS is their accuracy at identifying individual under uncontrolled scenarios [42],[31],[17],[46]. For instance, London's Metropolitan Police tested live facial-recognition technology between 2016 and 2019, and a report of the trials has raised multiple red flags [42]. The FRS flagged 46 people throughout the test phase, with four unsatisfactory cases for analysis [42]. Out of the remaining 42 matches, the operators dismissed 16 matches as 'non-credible' and did not attempt to verify the identity of the individuals [42]. Four potential matches were lost in the crowd, while 22 could be arrested [42]. Yet only eight of them turned out to be correct identification; hence, from all computed-generated red flags (42 matches), the systems correctly identified the individuals in only eight instances [42].

Bias: Advocates of augmented-video surveillance praised facial recognition systems for their neutrality and the opportunity they represent in mitigating police discrimination and racial prejudice. Historical racism and white male privilege have, however, cropped up at every step of the evolution of facial recognition systems: through the manner human bias have slithered in model design to how dataset fed to machine learning have been inclined to represent excessively a subset of the population and marginalized other communities [46]. In other words, certain groups of people are disproportionally matched to authority's database compared to their counterparts [31]. Buolamwini and Gebru [47] exposed the bias embedded in automated facial analysis algorithms and evaluated the accuracy of 3 commercial gender classification systems. They first found that commonly used datasets are overwhelmingly composed of lighter-skinned subjects, highlighting the need to create a balanced dataset to train algorithms [47]. Despite an overall error rate of 6.3%, they indeed showed that Microsoft's FaceDetect model suffered from concerning bias: lighter-skinned males were correctly identified without residual error while the application exhibits an error rate of 20.8% dark-skinned females [47].

Checklist: A salient point raised in the review of the London Metropolitan Police Services' Trial of live facial-recognition technologies was the ambiguity regarding the criteria employed for constructing the watchlist [**?**]. It included people 'wanted by the courts' and 'wanted by the police' across a wide range of potential offenses. Therefore, it is imperative to create clear guidelines defining what types of crimes subjects can appear on the list in order to avoid all kinds of abuse of these technologies. Also, these guidelines must define the reasons and time-horizon under which an individual shall be removed from the list. An independent body should certainly review these procedures to ensure these alignments to the public interests.

It is crucial to distinguish between threats fundamentally related to privacy and those stemming from algorithms' flaws because they entail a very different questioning process. Algorithmic bias is subtle because the deployment of such a system would discriminate against a part of the population, accentuating social conflicts within the society. As an illustration, the state of Oakland recently instituted the ban of facial recognition in public spaces, because of its inaccuracy and bias [43]. As for accuracy rate, technological advances have allowed for significant improvements and will continue in the coming years. One area of particular interest is the creation of fake faces for producing more variations in the training data-set [44]. Such practice could help mitigate algorithmic bias, but a prerequisite is the reconsideration of the social values that have created these biases in the first place. Concerns regarding privacy and freedom may pose even more significant challenges. It belongs to governments to conduct a public debate involving relevant stake-

holders to define the objectives of these systems, their limits, and the appropriate legal framework to minimize the risk of governmental abuses. Such an issue is explored in further detail in the discussion.

11.5.4 Applications of Facial Recognition Technologies

i Computer-Enhanced Video Surveillance in China

The Chinese government has drawn wide attention for its ascent in artificial intelligence (AI) and facial recognition [51]. The vast network of AI-augmented video cameras and the perpetual introduction of new facial recognition technologies, such as systems that can ID's mask wearers, combined with a regulatory ecosystem covering biometric information with only a low degree of granularity, makes China an exciting case study for our purposes. One of the driving forces supporting the country's tech industry, from established giants to audacious start-ups, is the immense pool of data and few constraints about mining it from its citizens [52]. Moreover, it has allowed firms to create innovative and accurate algorithms for corporations and states with little consideration of empowering a modern surveillance state [52], [53]. Western law enforcement agencies have also deployed facial recognition systems to track criminals. Still, with few exceptions, they have had the intention to track problematic behavior, track social activists, or control ethnic groups [53]. As a result, the scale and ambitions have so far set the country apart.
Borg [27] conceptual framework of social monitoring can be applied to facilitate the analysis. He discussed five dimensions to assess quantitative and qualitative differences in social monitoring [27], based on which the deployment of intelligent systems in China likely increases the intrusiveness of the social monitoring. Only the most relevant are discussed in this section.
The first dimension inspects the extent to which systems produce standardized information, and Borg [27] explains, "The more mechanical the process, the more standardized the information, and the more surveillance-like, or formal, the monitoring". By substituting human capital by technological capital, the Chinese government has greatly increased its capability to monitor citizens' life en Masse. In addition, Borg's [27] second dimension analyses the extent to which information is centralized. Underlying this criterion is the belief that a holistic file recording distinct aspects of behavior could be more of a threat than multiple, unaggregated dossiers describing only single dimensions [27]. Second generation system with analytic capabilities will increase the centralization of the collected data, at least in terms of ownership. This dimension is crucial in the case of China because of the existence of multiple private and public databases containing a wide range of personal information. With the standardization and the connection of these databases, the Chinese government would therefore enjoy a powerful instrument of mass surveillance.
Borg's [27] fourth dimension is the quantity of information produced, which dramatically increases under systems with video analytic capabilities. Indeed, Rule defined the system size as the number of individuals contained in the database and "corresponds to the amount of a subject's life depicted there". It is therefore evident that more information can be collected thanks to facial recognition, but also that these images are much more interesting than in the first systems.
Hence, with the combination of video analytic capabilities and previous database such as the Police Cloud, the Chinese government has the instruments to build a model of people's behavior gradually: it becomes simple to systematically identify and trace a subject of interest to investigate connections, habits, and everyday life [54]. Officially, the objective is to "track and predict the activities of those considered threatening to the regime" [55]. Individuals considered as a threat include criminals, activists, or terrorists, but as stressed by Yang [56], the definition of a criminal is fluid:

> "Hong Kong's own Personal Data (Privacy) Ordinance, which allows people to request their data, does not do the job, as it has broad exemptions for "the prevention or detection of crime". The problem is Hong Kongers already disagree with their government over what is criminal, and the category of "prevention" is broad and vague"

Despite this risk toward mass surveillance and potentially misusage of these technologies for discrimination or repression against minorities [53], the legal framework was considerably reshaped in 2020 [57]. On the one side, individuals now have the right to refuse to be identified via facial recognition technologies in private places such as shops [57]. On the other side, a new regulation stipulates that facial recognition in public spaces can be used for specific purposes and only when necessary [57]. The vague language prevents any stringent applications of the new law, yet it clearly shows an increased interest in regulating this hot topic.

11.5.5 Driver Monitoring Systems

In 2019, distracted driving caused over 3,000 deaths in the United States, and more than 10,000 had lost their lives in drunken driving crashes [48]. Driver monitoring systems (DMS) are set to be game-changer. These systems encompass advanced safety features using various kinds of sensors to detect driver drowsiness or distraction [49]. The inclusion of these systems in new vehicles is expected to become standard practice soon [50]. The European Union has taken a step in this direction and requires DMS to be installed in all new vehicle models from 2024.

Corporations have designed different forms of DMS, but the fundamental process remains the same. A sophisticated inboard software gathers data points from the images stream and processes these features to characterize the driver's attentive state. It can subsequently assert that if the driver is abnormally blinking, disproportionally closes his / her eyes and tilts his / her head at an unusual angle [50]. Any suspect variations from the baseline state are signaled to the driver by sounds and messages on the dashboard [50]. This software can considerably curb traffic offenses such as texting while driving and prevent traffic accidents. However, the benefits harvested do not necessarily conflict with individuals' privacy and freedom. For instance, DriveFocus, designed by Subaru, does not record video and audio.

The effectiveness of these systems in reducing accidents remains to be studied. It is easy to envisage that users to be annoyed by a frequent number of false signals, without the latter being able to interact with the system to inform it of our state. Yet, implementing this feature could also be counter-productive from a safety perspective if users mis-perceive their state of fatigue.

11.6 Discussion

The previous sections have shown that the proliferation of camera surveillance, and facial recognition, can be a double-edged sword for society. CCTV's advocates praised its capacity to restore social life in the public space by reducing crimes and socially offensive behaviors. Cameras provide public agencies with the possibility to re-establish or maintain a sense of security in the public sphere and thus create a city. However, people must also be able to appropriate the space and wear a mask of choice. By standardizing individuals' behavior in the desired direction, camera surveillance could also transform cities into shallow areas, where interactions are homogenized and eccentricity erased.

With second-generation CCTV, human rights such as freedom of expression, association, and peaceful assembly are at risk, accentuating society's normalization. In particular,

freedom of peaceful assembly protects individuals' right to articulate their opinion as part of a collective. Assemblies are potent means to share views, express discord, propose solutions, or raise awareness on important issues. Therefore, this right is essential to the foundation of a democratic civil society that considers its environment holistically and stimulates the inclusion of minorities in the decision-making process. Malevolent governments can use camera surveillance, in particular facial recognition, in their quest to identify and track down protesters [56]. Anti-government protests, which have rocked Hong Kong for months, showed clear examples of (mis)-usage of facial recognition. As a reaction, protesters wore facemasks, glasses and used umbrellas to prevent identification using automated facial-recognition technology [56]. At the heart of this issue is the lack of transparency and public debate before deploying these technologies and, in due course, the lack of controls to ensure utilization according to their original purposes.
The technological society as we know it might favor the gradual acceptance of a narrowing right to privacy. Individuals have been accustomed to the interference from governments and corporations: from the collections of personal data on the internet, smart meters for electricity and water, or DNA testing services to unveil one's family history. But, as the limit of the possibilities is constantly being pushed back, the consequences of constant sharing personal data with the states or corporations attract only periodic scrutiny and consideration. In order to prevent any abuse of power, misuse, and maintain public support, the introduction of powerful tools such as facial recognition should be preceded by a detailed discussion on the "purposes, operations, and regulations of the systems" [58]. The benefits of technological advances, whether through facial recognition or alcohol detection, cannot be achieved when these prerequisites are violated. Society should then be allowed to define under "what circumstances it is legitimate and necessary then decide to sacrifice privacy and other fundamental rights to a certain degree, in the interest of security" [58].
However, a ban on facial recognition is not the solution to its current problems. The example of DMS showed the society could harvest significant gains from the implementation of FR technologies, from efficiency gains to road safety. Therefore, a policy of defining a clear legal framework to prevent abuses is preferabl

11.7 Conclusion

This report attempts to discuss key technological developments in the domain of online CSAM detection in the background of arguments in favour of and against such detection, concerns about privacy, the prevailing legal framework in countries across the world, and distribution methods for online CSAM. Throughout this research, we have weighed the advantages and disadvantages of adopting these new technologies and asserted that a one-size-fits-all solution is implausible. Threats to privacy, freedom of expression, and protest must be assessed in the specific context of the technological application. As an illustration, we have shown that facial recognition could help reduce traffic accidents but could also serve as a tool for mass surveillance.
In light of these findings, it is essential to adopt these systems with infinite transparency as to their purpose, use, and limits, while developing an adequate legal framework. This is the only way to ensure that the goals of the tool are aligned with the demands of the consortium and the citizens. In this sense, this work should be learned by paying greater attention to the legal framework regulating CSAM and its second in the European Union, as well as the changes needed to ensure that the privacy of users is not violated.
We would like to conclude this work with the following quote from a victim:

> "I have to live with the knowledge that my abuse will never end, and that every second of every day, someone could be—almost certainly is—watching

my torture and abuse. Even once I'm dead, my degradation will continue. I will never be able to escape it. This trauma is infinite." [2]

Bibliography

[1] Apple.com: *Expanded protections for children*; Website, 2021, https://www.apple.com/child-safety/.

[2] Apple Inc.: *Child Sexual Abuse Material (CSAM) Detection*; Technical Summary, 2021, https://www.apple.com/childsafety/pdf/CSAM_Detection_Technical_Summary.pdf.

[3] Ronald Clarke: *Burglary of Retail Establishments*; Report, US Department of Justice, Office of Community Oriented Policing Services, 2002, https://popcenter.asu.edu/content/burglary-retail-establishments-0.

[4] Jennifer Cobbe: *Data Protection, ePrivacy, and the Prospects for Apple's On-Device CSAM Detection System in Europe*; SocArXiv, Vol. 11, 2021, https://doi.org/10.31235/osf.io/rhw8c.

[5] MaryJane Gurriell: *Born into Porn But Rescued by Thorn: The Demand for Tech Companies to Scan and Search For Child Sexual Abuse Images*; Family Court Review, 2021, https://doi.org/10.1111/fcre.12613.

[6] Gail Hornor: *Child sexual abuse: consequences and implications*; J. Pediatr. Health Care, Vol. 24, No. 6, 2010, pp. 358-364, https://doi.org/10.1016/j.pedhc.2009.07.003.

[7] icmec.org: *International Centre for Missing and Exploited Children (ICMEC)*; Website, 2021, https://www.icmec.org.

[8] iwf.org: *Internet Watch Foundation (IWC)*; Website, 2021, https://www.iwf.org.uk.

[9] Danielle Kettleborough: *What's wrong with "child pornography"? The impact of terminology*; 2015, https://www.researchgate.net/publication/286202306_What's_wrong_with_child_pornography_The_impact_of_terminology_weblog_pos

[10] Hee-Eun Lee, Tatiana Ermakov, Vasilis Ververis, Benjamin Fabian: *Detecting Child Sexual Abuse Material: A Comprehensive Survey*; Forensic Science International: Digital Investigation, Vol. 34, 2020, https://doi.org/10.1016/j.fsidi.2020.301022.

[11] NCMEC: *National Center for Missing and Exploited Children (NCMEC)*; Website, 2021, https://www.missingkids.org/HOME.

[12] Mayana Pereira, Rahul Dodhia, Hyrum Anderson, Richard Brown: *Metadata-based detection of child sexual abuse material*; arXiv preprint, 2020, arXiv:2010.02387.

[13] Michael S. Scott, Kelly Dedel: *Clandestine Methamphetamine Labs*; US Department of Justice, Office of Community Oriented Policing Services, 2006, Washington DC, https://popcenter.asu.edu/content/clandestine-methamphetamine-labs-2nd-ed-0

[14] Michael Salter, Elly Hanson: *"I Need You All to Understand How Pervasive This Issue Is": User Efforts to Regulate Child Sexual Offending on Social Media*; Editors Bailey, J., Flynn, A. and Henry, N., The Emerald International Handbook of Technology Facilitated Violence and Abuse(Emerald Studies In Digital Crime, Technology and Social Harms), Emerald Publishing Limited, Bingley, 2021, 729-748 https://doi.org/10.1108/978-1-83982-848-520211053

[15] Swiss Foundation for the Protection of Children: *Sexual violence against children online*; Report, 2018, https://www.ohchr.org/Documents/HRBodies/CRC/GCChildrensDigitalEnvironment/2020/others/swiss-foundation-protection-of-children-2020-11-14.docx

[16] Thorn.org: *Thorn.org*; Website, 2021, https://www.thorn.org/about-our-fight-against-sexual-exploitation-of-children/.

[17] Ray Surette: *The Thinking Eye: Pros and Cons of Second Generation CCTV Surveillance Systems*; Policing: An International Journal of Police Strategies & Management, Vol. 28, No. 1, 2005, 152-73. Emerald Insight, https://doi.org/10.1108/13639510510581039.

[18] Michael Zehnder: *The economics of subjective security and camera surveillance*; WWZ Forschungsbericht, No. 04/09, 2009, Universität Basel, Wirtschaftswissenschaftliches Zentrum (WWZ), Basel, https://www.econstor.eu/bitstream/10419/127522/1/wwz-fb-2009-04.pdf.

[19] Katherine S. Williams, Craig Johnstone: *The Politics of the Selective Gaze: Closed Circuit Television and the Policing of Public Space*; Crime, Law and Social Change, Vol. 34, No. 2, September, 2000, 183-210. Springer Link, https://doi.org/10.1023/A:1008342610872.

[20] Eric L. Piza, Brandon C. Welsh, David P. Farrington, Amanda L. Thomas: *CCTV Surveillance for Crime Prevention*; Criminology & Public Policy, Vol. 18, No. 1, 2019, 135-59. Wiley Online Library, https://doi.org/10.1111/1745-9133.12419.

[21] Ken Pease: *A review of street lighting evaluations: Crime reduction effects*; 2019. In Painter, K. and Tilley, N. (eds.): *Surveillance of public space: CCTV, street lighting and crime prevention*; Crime Prevention Studies: Vol. 10. Monsey, NY: Criminal Justice Press.

[22] Steven Feldstein: *The Global Expansion of AI Surveillance*; Working Paper, September, 2019, 135-59. Carnegie Endowment for International Peace, https://carnegieendowment.org/files/WP-Feldstein-AISurveillance_final1.pdf.

[23] IHS: *Video surveillance: How technology and the cloud is disrupting the market*; Summary Report, 2016, https://cdn.ihs.com/www/pdf/IHS-Markit-Technology-Video-surveillance.pdf.

[24] Wesley G. Skogan: *The Future of CCTV*; Criminology & Public Policy, Vol. 18, No. 1, February, 2019, 161-166, https://doi.org/10.1111/1745-9133.12422.

[25] Kelly A. Gates: *Wanted Dead or Digitized: Facial Recognition Technology and Privacy*; Television & New Media, Vol. 3, No. 2, May, 2002, 235-38, https://doi.org/10.1177/152747640200300217.

[26] Neil C. Thompson, Kristjan Greenewald, Keeheon Lee, Gabriel F. Manso : *The Computational Limits of Deep Learning*, July, 2020, http://arxiv.org/abs/2007.05558.

[27] Marian J. Borg: *The structure of social monitoring in the process of social control*; Deviant Behavior, Vol. 18, No. 3, 273-293, 1997,https://doi.org/10.1080/01639625.1997.9968059.

[28] Yi Zeng, Enmeng Lu, Yinqian Sun, Ruochen Tian: *Responsible Facial Recognition and Beyond*; September, 2019, http://arxiv.org/abs/1909.12935.

[29] Melissa Roux: *Why Facial Recognition Is Important For Banking Services*; Sightcorp (blog), March, 2019, https://sightcorp.com/blog/why-facial-recognition-is-important-for-banking-services/.

[30] Joy Buolamwini, Vicente Ordenez, Jamie Morgenstern, Erik Learned-Miller: *Facial Recognition Technologies: A Primer*, 2020, https://people.cs.umass.edu/~elm/papers/FRTprimer.pdf.

[31] Douglas Yeung, Rebecca Balebako, Carlos Ignacio Gutierrez Gaviria, Michael Chaykowsky: *Face Recognition Technologies: Designing Systems that Protect Privacy and Prevent Bias*; Homeland Security Operational Analysis Center operated by the RAND Corporation, 2020, https://www.rand.org/pubs/research_reports/RR4226.html.

[32] Frank Hersey: *In-Car Biometrics to Detect Drunk-Driving Could Add to Data Build up*; Biometric Update, August, 2021, https://www.biometricupdate.com/202108/in-car-biometrics-to-detect-drunk-driving-could-add-to-data-build-up.

[33] David Gershgorn: *Retail Stores Are Packed with Unchecked Facial Recognition*; The Verge, 14 July, 2021, https://www.theverge.com/2021/7/14/22576236/retail-stores-facial-recognition-civil-rights-organizations-ban.

[34] Karamizadeh Sasan, Abdullah Shahidan, Zamani Mazdak: *An Overview of Holistic Face Recognition*; International Journal of Research in Computer and Communication Technology. No. 2, 2013,738-741, https://www.researchgate.net/publication/262725509_An_Overview_of_Holistic_Face_Recognition.

[35] Cambridge Dictionary: *CCTV*; https://dictionary.cambridge.org/fr/dictionnaire/anglais/cctv. Accessed 16 Dec. 2021.

[36] Andrew E Taslitz: *The Fourth Amendment in the Twenty-First Century: Technology, Privacy, and Human Emotions*; Law and Contemporary Problems, Vol. 65, No. 2, 2002, https://doi.org/10.2307/1192242.

[37] Christopher Slobogin: *Public Privacy: Camera Surveillance of Public Places And The Right to Anonymity*; SSRN Scholarly Paper, Social Science Research Network, February, 2003, https://papers.ssrn.com/abstract=364600.

[38] Richard A. Wasserstrom,: *Privacy: Some Arguments and Assumptions*: Philosophical Dimensions of Privacy: An Anthology, edited by Ferdinand David Schoeman, Cambridge University Press, 1984, pp. 317-32. Cambridge University Press, https://doi.org/10.1017/CBO9780511625138.015.

[39] Ribeiro, R. H., Rodrigues, B. B., Killer, C., Baumann, L., Franco, M. F., Scheid, E. J., & Stiller, B. (2021, June). ASIMOV: a Fully Passive WiFi Device Tracking. In 2021 IFIP Networking Conference (IFIP Networking) (pp. 1-3). IEEE.

[40] B. Rodrigues, C. Halter, M. Franco, E. J. Scheid, C. Killer and B. Stiller, "BluePIL: a Bluetooth-based PassIve Localization Method," 2021 IFIP/IEEE International Symposium on Integrated Network Management (IM), 2021, pp. 28-36.

[41] B. Rodrigues, L. Müller, E. J. Scheid, M. F. Franco, C. Killer and B. Stiller, "LaFlector: a Privacy-preserving LiDAR-based Approach for Accurate Indoor Tracking," 2021 IEEE 46th Conference on Local Computer Networks (LCN), 2021, pp. 367-370, doi: 10.1109/LCN52139.2021.9524945.

[42] Peter Fussey, Daragh Murray: *Independent report on the London Metropolitan Police Service's trial of live facial recognition technology*, 2019, http://repository.essex.ac.uk/24946/1/London-Met-Police-Trial-of-Facial-Recognition-Tech-Report-2.pdf

[43] Sarah Ravani: *Oakland Bans Use of Facial Recognition Technology, Citing Bias Concerns*; San Francisco Chronicle, 17 July, 2019, https://www.sfchronicle.com/bayarea/article/Oakland-bans-use-of-facial-recognition-14101253.php.

[44] Yannick Chavanne: *Ces visages qui n'existent pas vont créer des outils de reconnaissance faciale*; ICT Journal, September, 2021,https://www.ictjournal.ch/articles/2021-10-12/ces-visages-qui-nexistent-pas-vont-creer-des-outils-de-reconnaissance-faciale. Accessed 16 Dec. 2021.

[45] Yassin Kortli, Maher Jridi, Ayman Al Falou, Mohamed Atri: *Face Recognition Systems: A Survey*; Sensors, Vol. 20, No. 2, January, 2020, https://doi.org/10.3390/s20020342.

[46] David Leslie: *Understanding Bias in Facial Recognition Technologies: an explainer*; The Alan turing Institute, September, 2020, https://doi.org/10.5281/zenodo.4050457.

[47] Joy Buolamwini, Gebru Timnit: *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*; Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR, 2018, pp. 77–91. proceedings.mlr.press, https://proceedings.mlr.press/v81/buolamwini18a.html.

[48] National Center for Statistics and Analysis: *Overview of motor vehicle crashes in 2019*; Traffic Safety Facts Research Note, Report No. DOT HS 813 060, National Highway Traffic Safety Administration, December, 2020, https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813060

[49] Chris Schwarz, John Gaspara, Thomas Millerb, Reza Yousefianb: *The Detection of Drowsiness Using a Driver Monitoring System*; Traffic Injury Prevention, Vol. 20, No. 1, June, 2019, 57-61, https://doi.org/10.1080/15389588.2019.1622005.

[50] Aptiv: *What Is a Driver-Monitoring System?*; Aptiv, 26 April, 2021, https://www.aptiv.com/en/newsroom/article/what-is-a-driver-monitoring-system.Accessed 16 Dec. 2021.

[51] Dave Davies: *Facial Recognition And Beyond: Journalist Ventures Inside Chinas Surveillance State*; NPR, January, 2021, https://www.npr.org/2021/01/05/953515627/facial-recognition-and-beyond-journalist-ventures-inside-chinas-surveillance-sta?t=1639682559812.

[52] Michael, Standaert:*Smile for the Camera: The Dark Side of China's Emotion-Recognition Tech.* The Guardian, 3 March, 2021. The Guardian, https://www.theguardian.com/global-development/2021/mar/03/china-positive-energy-emotion-surveillance-recognition-tech.

[53] Alfred Ng: *China Tightens Control with Facial Recognition, Public Shaming*; CNET, https://www.cnet.com/news/in-china-facial-recognition-public-shaming-and-control-go-hand-in-hand/. Accessed 05 December. 2021.

[54] Simon Denyer: *Beijing bets on facial recognition in a big drive for total surveillance*; The Washington Post, https://www.washingtonpost.com/news/world/wp/2018/01/07/feature/in-china-facial-recognition-is-sharp-end-of-a-drive-for-total-surveillance/. Accessed 04 December. 2021.

[55] Chris Burt. *Human Rights Watch Calls on China to Shut down Police Cloud*; Biometric Update, 21 November, 2017, https://www.biometricupdate.com/201711/human-rights-watch-calls-on-china-to-shut-down-police-cloud

[56] Yuan Yang: *Why Hong Kong Protesters Fear the City's "Smart Lamp Posts"*; Financial Times, 8 January, 2020, https://www.ft.com/content/f0300b66-30dd-11ea-9703-eea0cae3f0de.

[57] Seungha Lee: *Coming into Focus: China's Facial Recognition Regulations*; 4 May, 2020, https://www.csis.org/blogs/trustee-china-hand/coming-focus-chinas-facial-recognition-regulations. Accessed 16 Dec. 2021.

[58] Benjamin J Goold: *CCTV and Human Rights* in *Citizens, Cities and Video Surveillance: Towards a Democratic and Responsible Use of CCTV*; Paris: European Forum for Urban Security, 2010, https://commons.allard.ubc.ca/fac_pubs/152/.