UNIVERSITÀ POLITECNICA DELLE MARCHE

Faculty of Engineering
Department of Information Engineering
Master of Science in Biomedical Engineering

PRIVACY PRESERVING PROTOCOLS FOR HEALTHCARE BASED ON BLOCKCHAIN TECHNOLOGY

Supervisor
Prof. Marco Baldi

Candidate
Dalila Calabrese

Co-Supervisor
Prof. Franco Chiaraluce

Academic Year 2018-2019

Abstract

Blockchain technology is experiencing an exponential growth in research and industry. It enables a cryptographically secured and irrevocable data sharing leveraging distributed process. The use of blockchain in the healthcare sector turns out to be challenging in order to store data. The treatment of healthcare data is strictly regulated by the General Data Protection Regulation (GDPR), which requires transparency, security, privacy, integrity, non-repudiation and erasure of those data. Blockchain technology, while being an element of innovation that may provide to patients full control on their health information, is intrinsically in contradiction with the GDPR key factors. In literature proliferate prototypes that aim to obtain the compliance of the blockchain with the GDPR but still there is no security analysis on these models that propose a proper way to achieve a compliant mechanism for data exchange. Moreover, blockchain technology is not totally immune to emerging security threats with consequences in terms privacy and treatments effectiveness.

The purpose of this Master's thesis is to abstract models of privacy preserving protocol based on blockchain technology in the healthcare sector. For each model the multiple cryptographic protocols exploited, in order to address the GDPR-compliance, are investigated. Thereafter their safety is analysed, and the obtained results are used to quantify the risk related to their use. Permissioned blockchain infrastructure, based on particle Byzantine fault tolerance (PBFT) consensus mechanism, that preserves on-chain the hash digest of the healthcare data, seems to be the safest among those analysed. Its core technology components in fact, reduce its risk of exposure related to different types of attack.

Contents

List of Figures

Introduction

The citizen who meets health facilities for diagnosis, treatment, medical services, administrative operations must be guaranteed the most absolute confidentiality and the widest respect for his fundamental rights and his dignity. In a completely digitalized world, health management, as well as other sectors, has been subject to a dematerialization of documents, which has changed the conception, creation and conservation of the document itself. In the last thirty years, the need of physical support to produce and preserve data, information and documents has been eliminated. Analog supports have been left, due to the rising of digital technologies which, by reducing production costs and storage spaces and increasing the availability of the document, require a continuous updating of the knowledge of users in order to reduce the risks associated with them. In the health sector, large amounts of heterogeneous data are continuously used, transmitted and processed. Data sharing and data interoperability are the key concepts to ensure an efficient health care system. The treatment of patients data, in health facilities, is strictly regulated by the D.Lgs no. 101/2018, for the compliance of the Italian system to the EU Regulation no. 679 of 2016, the so-called General Data Protection Regulation (GDPR), which has introduced data protection laws that have a deep impact on health data management. Health data are those "*related to the physical or mental health of a natural person, including the provision of health care services, which reveal information relating to his or her state of health*" (art. 4 GDPR), and as such require transparency, security, privacy, integrity and

non-repudiation. In a context of increasingly complex and extended information systems, the blockchain represents, for the health field, an element of innovation that provides an immutable and secure architecture for the exchange and sharing of data within physically distant health information systems. The blockchain, transforming the way in which the main actors of health systems interact with each other, may provide patients with full control on their health information, allowing its management while ensuring data integrity and certification. One of the main principles on which the blockchain technology is based is to keep the data in an immutable public transaction ledger. Once the data are entered in the blockchain it is no more possible to modify or delete them. Indeed, due to the Blockchain immutability by design, they will remain immutable on the ledger. By keeping health data on a public ledger, the concept of confidentiality fails: it turns out to be more challenging to protect data from all individuals who are not allowed to see them. Moreover, restrict the access to information in a public structure is not always possible. Therefore, the blockchain turns out to be intrinsically in contradiction with the GDPR prerequisites, the most innovative concepts introduced by the above-mentioned regulation, such as the same confidentiality and the same right to be forgotten are not guaranteed within the blockchain technology.

The idea of making the use ok blockchain technology compliant with the GDPR prerequisites has prompted the analysis conducted in this work. In literature, several approaches for the creation of a GDPR-compliant blockchain-based model have been proposed in order to overcome this lack.

In this thesis work, these approaches are synthetized and analysed. Through their evaluation, three GDPR-compliant abstract paradigms are proposed, in order to provide a synthetic and clear abstraction of their complex structure and elaborate working principle. These systems by using the blockchain provide GDPR-compliant technical solution for the digitalization, preservation and sharing of dematerialized health data,

ensuring data authenticity, validity and interoperability. Although the technology on which the analysed models are based seems to provide infallible data exchange systems, cryptographically secure, irrevocable and with high reliability and fault tolerance of the system, it is not totally immune to emerging and sophisticated security attacks, which can produce enormous and irreparable damage, not only in terms of privacy, but also in terms of treatments effectiveness, compromising the patient's life. Therefore, for each assessed model, a security analysis of its privacy preserving protocols was carried out.

Aim of this work was to understand whether in the healthcare sector there are GDPR-compliant blockchain-based solutions, which provide safe health data storage systems.

CHAPTER 1
Privacy

1.1 General Features

The term privacy denotes the right of a person to preserve her private life confidentiality. The concept on which it is founded has been characterized by an evolution over time: initially it was referred to the private sphere of a person's life, nowadays it is extended to include the right of any person to control his/her own personal data and to decide about them, verifying how own personal data processing takes place when in case of necessity.
Privacy, recognised in the Universal Declaration of Human Rights, is a fundamental human right and it is interrelated with data protection.
In a global economy governed by technology, individual information turns out to be the most economically desirable electronic asset. Over the years, it has become clear that government agencies, companies and third parties have access to private information, both personal and business.
Defining personal data as “any information that relates to an identified or identifiable living individual”, various technical and legal approaches have been proposed in order to strengthen privacy rights [1]. Indeed, in most countries personal information is protected by laws that limit its collection and use by public and private entities, requiring institutions to give

unambiguous notice to the subject about the type of data collected and the purposes of data collection, which is authorised only upon an explicit consent of the individual. It is therefore clear that privacy and data protection are two interrelated aspects of the internet governance issue: defining privacy as the right to control own personal data, disclosing them or not, the data protection ensures, not only, data confidentiality but, by extending the protection of the individual beyond the sphere of private life, guarantees decision-making self-determination and control over the data circulation. Furthermore, data protection guarantees personal freedom, as physical freedom but also against any illegitimate control or interference by others. A protection to the life of the data processing is added in order to allow an effective protection of personal data.

1.2 General Data Protection Regulation (GDPR)

Data privacy and security are strictly regulated by D.Lgs no. 101/2018, for the compliance of the Italian Code on the protection of personal data to the provisions of EU Regulation 679/2016, so-called General Data Protection regulation (GDPR).

The GDPR is a European Union regulation about the processing of personal data and privacy. It was adopted on 27 April 2016, entering into force on 25 May of the same year, it was subsequently made operational from 25 May 2018 [1], [2]. By unifying and making homogeneous the privacy legislation across the European Union (EU), this regulation aims to strengthen the protection of personal data of EU citizens and residents of the EU, both within and outside the borders of the EU. The GDPR, introducing data privacy Laws, returns to citizens the control of their personal data. The Regulation applies to the processing of personal data, and to the non-automated processing of stored data. Defining personal data “any information relating to an individual,

related to his or her life, whether private, professional or public", the European Commission proposes an extension of the data definitions by offering the following classification of information that allows the unique identification or authentication of a natural person:

· *Personal data* (Article 4 paragraph 1): information relating to a natural person identified or identifiable directly or indirectly.

· *Sensitive data* (Article 9 paragraph 1): personal data relating to religion, politics, health, sexual life, as well as genetic and biometric data.

· *Personal data relating to criminal convictions or crimes* (article 10): personal data relating to criminal allegations, proceedings or convictions

The regulation is addressed to all companies or entities dealing with the processing of personal data of EU citizens, regardless of their legal head office, and the location of processing and archiving systems. The GDPR does not, however, deal with the processing and management of personal data for public policy or national security activities.

Applying both to automated processing and to manual processing, it appears to have a neutral approach from the technological point of view: for example, the type of filing (or archiving) does not provide differentiations within the regulation itself: indeed, GDPR imposes that all data are subjected to the same protection requirements [1], [3].

1.2.1 The GDPR Actors

The main actors involved in the GDPR are respectively the *data subject*, person to whom the personal data refers, the *data controller*, that defines which kind of personal data wants and for what purpose, and the *data processor* that is a natural or legal body which 'processes personal data on behalf of the controller'. The data controller is the key role of the regulation, in fact, by exercising overall control over the purposes and means of the

processing of personal data, he must comply with, and demonstrate compliance with, all the GDPR requirements [2].
At the top of the regulation, on the other hand, there are two players: the European Data Protection Board, that plays a regulatory and supervisory role, thanks also to consultations with the European Union Commission, and the Supervisory Authority, which supervise compliance with the regulation. The Board, in turn, is composed of several authorities which are respectively the European Data Protection Supervisor and the head of one Supervisory Authority of each Member State. The Supervisory Authority is an independent authority that allows local enforcement of legislation.
Furthermore, the GDPR imposes the designation of a Data Protection Officer if the processing of personal data is carried out by a public authority and in organizations that implement treatments that require "regular and systematic monitoring" of data or deal with sensitive or judicial data on a large scale [3].

1.2.2 Legal Issues

The GDPR is founded on a series of principles, rights and obligations, all aimed at protecting personal data. The six basic principles on which the formulation of the rights is based are the following [2]:

- the lawfulness, fairness and transparency with which personal data should be processed
- purpose limitation in the collection of personal data
- personal data minimization to what is necessary for the specified purposes
- accuracy and updating of personal data managed by the controller
- storage limitation of personal data for no longer than necessary for the stated purposes

- integrity and confidentiality of personal data in order to guarantee adequate security.

Furthermore, the GDPR states that, once the data subject gives its consent in a free, informed and unequivocal way to the treatment of the data, the data controller is lawfully authorized to the treatment of the same, until the data subject decides to revoke its consent. Indeed, among the main obligations of data controllers there is that of treating personal data in a lawful manner.
The GDPR rules grant more rights to the data owner, compared to the previous European directives. Among these rights, the most important and those introduced for the first time in the field of privacy are [1], [3], [4]:

- *The right to rectification:* data subjects have the right to promptly update incorrect data. Users have the right to obtain the correction of their personal data in case of inaccuracy or incompleteness. This right also implies that the correction must be communicated to all the third parties involved in the data processing, unless impossibility or excessive difficulty.
- *The right to erasure ('right to be forgotten'):* data subjects are entitled to request to the data controller the removal and the interruption of each dissemination of personal data that are no longer necessary for legitimate processing purposes. This right can be rejected where the personal data are processed for health or archiving purposes in the public interest or are required for legal defense.
- *The right to access*: the data subject is authorized to inquire about the data controller, in order to know if its own data are processed, for what purpose, where and how this is done and with whom the data were shared. Although the right to access is closely related to the right to data portability, there is a clear distinction between them.
- *The right to data portability*: connected with the right to access, it is introduced to give the right to the data subject to transmit the data previously provided to another controller, without being hindered by the previous controller. This right applies only to personal data and

does not apply to truly anonymous data, that cannot be linked back to the user. In most cases the organizations must comply with a request without charging a commission. However, in case of an unfounded or excessive request, it is possible to demand a "reasonable fee" to execute the request itself.

· *The right to be informed*: organisations must provide users with information on the data processing activities that they carry out. Such information should be provided in a clear, concise and intelligible manner, usually through a privacy notice/policy, when personal data are concerned.

· *The right to object*: according to the GDPR, the user can object to processing activities of its data carried out by the data controller. It is necessary that the user justifies its objection; only in case of direct marketing purposes, the subject can exercise its right without any motivation.

· *Rights related to automated decision-making and profiling*: data subjects have the right not to be subject to a decision when it is based on automated processing or profiling and it has a legal or equally significant effect on the user. Organisations may only carry out automated decision-making processes if this is necessary for the performance of a contract and is authorised by the law of the EU State applicable to the controller. After the explicit consent of the user, it is possible to make automated decisions based on special category data.

· *The right to limit processing*: users have the right to limit the processing of their personal data if they contest the accuracy of the data and the processing thereof, in the case of unlawful processing and when data are no longer required, but the user requires them in order to establish a legal claim. The restriction should be disclosed to all addressees of third parties involved in the processing of the data in question, unless this is impossible or disproportionately difficult.

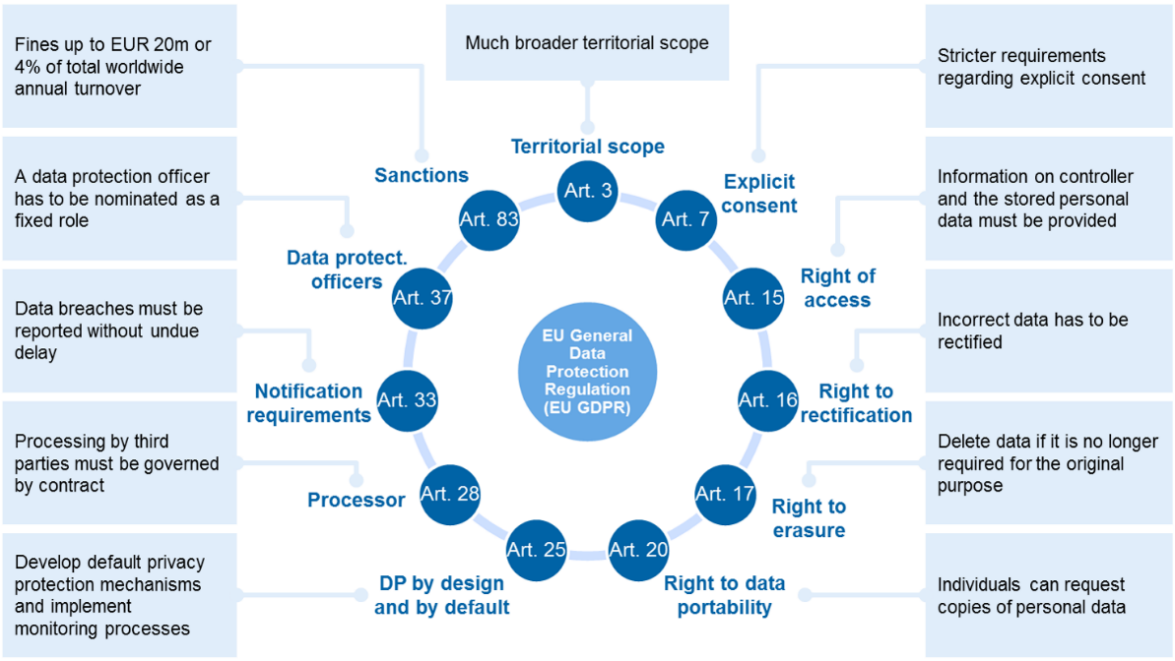

Figure1.2.2: *Data Subject Rights under GDPR*

1.2.3 The GDPR impact on technology

The technology that is based and thrives on certainties, rules and clear requirements often clashes with the GDPR, which appears to be a complex regulation and open to different interpretations. The fact that companies are required to manage all personal data, even though they do not know where these personal data are placed shows a real need for structural reorganization, in order to be GDPR-compliant.

The development of emerging technologies such as blockchain, artificial intelligence and cloud computing will be strongly influenced by GDPR. These technologies, considered effective means to increase performance and productivity, offer their potential value through huge amounts of data and algorithms. It follows that stricter data management and processing rules are

slowing down the development of these technologies, inevitably increasing the costs that companies need to face [4].

Since many cybersecurity incidents and data breaches have happened in the past, the GDPR requires companies to implement reasonable data protection measures, strengthening their cybersecurity, in order to protect consumers' personal data against data exposure or loss and against threats and violations. Among the key changes made by the GDPR there is the breach notification, in all member state where the violation itself is a risk for the right and individual freedom. In particular, the Supervisory Authority must be informed by the data controller within 72 hours of becoming aware of any data breach. Users, within the same time frame, must be informed of the violation unless the violated data was encrypted or, in general, the violation is unlikely to entail a risk to users' rights and freedoms [3], [4].

The GDPR, by imposing high requirements for data controllers and data processors and for the processing and storage of personal data, requires the companies to make an internal evaluation of their technological platforms and data architecture, in order to be able to guarantee all the users' rights [5].

Therefore, with GDPR radically changes the way data controllers and data processors handle data. In particular, the protection of personal data is no longer seen as a "additional component" of organisations but becomes an integral part of the organisation itself. In fact, it is included in a data processing system, designed specifically for the purposes of processing and the technologies used by organizations.

***Figure 1.2.3**: GDPR Compliance Cycle*

CHAPTER 2
Blockchain Technology

2.1 General Features

The blockchain technology is a disruptive distributed ledger technology, aimed at both protecting data and individual freedom and at simplifying the secure exchange of information and assets [6].
Proposed by Satoshi Nakamoto in 2008, it is a shared, public and immutable data structure organized in blocks [7]. Each block contains a set of records and these represent the recording of a particular event associated with a time instant (timestamp). Being the blockchain an hash-linked timestamps chain of blocks [7], the overall network structure takes the form of a chain since each block is linked to the previous one.
The link between a block and its predecessor is realized through the insertion, in each block, of a reference to the previous block. This backwards dependency finds its end in the initial block of the chain, which usually is generated from scratch.

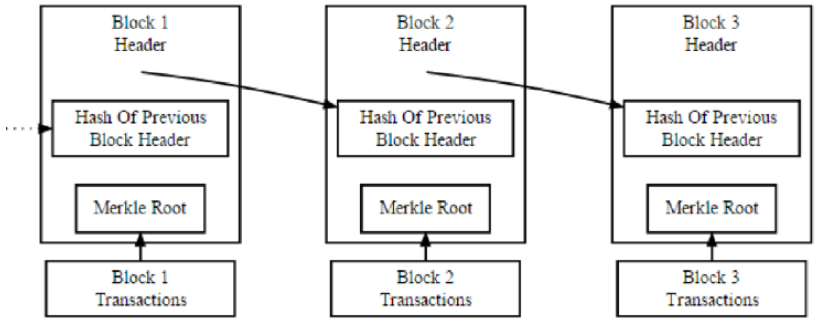

Figure 2.1.1 *Chain of blocks in a blockchain*

In order to obtain a growing list of records, each new block, uniquely identified by a hash, is created by means of a hash computation.

The Secure Hash Algorithm (SHA-256) provides the unique generation of a fixed 256-bit cryptographic hash, necessary to assure the continuity and the immutability of this shared structure; in fact, as said before, each block contains its cryptographic hash and the hash of the previous one, except for the genesis block [8], [9] .The blockchain can be seen as the result of the evolution of the ledger concept, where ledger means the register in which all system transactions are recorded. The blockchain implements a 'distributed ledger', evolution of the 'centralized ledger'. There has been a shift from a rigorously centralised logic, characterised by the recognition of an authority and thus by a "one-to-many" relationship, to a real and complete distributed logic based on a concept of trust among all participants.

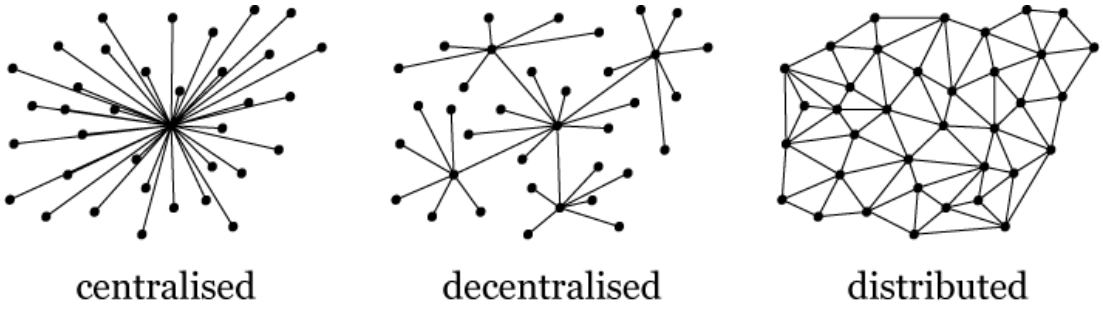

Figure 2.1.2*: From Centralised to Distributed Ledger*

The blockchain evolves through the activity of a distributed timestamp client that takes a block and through the temporal instant (timestamping) marks the block and calculates the hash. Indeed, the blockchain technology is the combination of digital signature and timestamp. The first is used for authentication purposes, ensuring protection from falsification and non-repudiation of origin. While the second is used to mark, in a tamper proof way, the time of each transaction on the network: being each block connected to the previous and to the next in chronological order, it is proof of what occurs in blockchain.

The blockchain is characterized by a radical decentralization of the data that guarantees reliability and trust without the presence of a central authority. It is trusted by consensus, this means that there is a distribution of trust and power among each network's node. The absence of intermediaries allows to have an egalitarian network, in which the concept of trust becomes implicit.

Each node owns a copy of the database file but, at the same time it is responsible for its verification. Each transaction is validated by the network through a trust model based on a group consensus. Therefore, the trust is established independently from a central authority. Once the transaction is validated by the community, through significant computing resources (mining), the addition of the block to the chain of each member is authorized. The data are so logged sequentially and in immutable way on a public ledger [2], [10]. A blockchain has the following properties [10]:

- is distributed: information is replicated and spread over multiple nodes in the network
- is immutable: is a permanent tamper-proof record of transactions. The immutability of this artificial architecture provides the immutability of its contents, which cannot be modified or deleted, once inserted in the structure, without corrupting the entire structure itself
- is transparent: being the blockchain an open file with a full transaction history, this generates traceability in a network in which any user can access

- is consensus driven: this mechanism works without the presence of a central authority. It allows the trust verification of each block through a model of consensus.

2.2 Architecture

The architecture of the blockchain is determined by the absence of intermediaries. Although the non-hierarchical nature of the blockchain structure assures the equality of all the nodes, it is possible to generally distinguish between two different types of nodes that join the blockchain:

- full node: the node must download a copy of the blockchain in order to be able to perform operations and, at the same time, validate the operations carried out by the other users of the blockchain
- lightweight node: the node can still participate in the activities of the blockchain, but without validating the operations of other users. The advantage of a lightweight node is that it does not have to download the entire copy of the blockchain (download only part of it) in order to perform operations.

Basically, the working principle of the network is built on the action of a group of nodes that provides the storage of synchronised copies of the same data [2].

2.3 Types of Blockchain

Proposed as a public permissionless technology, many different types of blockchains have been developed in time and they can be classified from both an amministrative and an implementation point of view into private, public, permissioned and permissionless blockchains.

Specifically, each type is suitable for a particular set of use cases [11]:

- *Permissionless Public:* this is a blockchain in which anyone can control the activity of anyone: everyone can join or leave the network, participating in the consensus. Read and write access is guaranteed to anyone: in this way minimum trust is required between the nodes, thus achieving maximum transparency.
- *Permissioned Public:* this blockchain is characterized by a partial decentralization of the network: reading of the blockchain is granted to anyone, instead writing of new data and the participation to the consensus protocol is allowed according to some permission granted by a consortium of administrators.
- *Permissionless Private:* the sharing of information in this blockchain does not happen publicly. Being permissionless, everyone can join or leave the network at any time. The presence of contracts on these networks establish who has both read and write access to the blockchain.
- *Permissioned Private:* the data storage occurs by means of a permissioned access control system managed by users of the network. The network administrators ensure network membership, providing participants with read and write access to data.

2.4 Transactions

Being the blockchain a shared structure, each single node shares the archive of the entire blockchain and thus each block with all transactions.
In particular, a transaction is a value transfer, recorded in a block. It is an unmodifiable exchange of information between users. Being not encrypted, it is possible to browse and view every transaction ever collected into a block. Specifically, each transaction depends on previous transactions: the resources that will be sent to the new addressee relay on the previous resources' exchanges. For each transaction it is possible to recognize a state, that

represents the status features and a function of state transition. The above-mentioned function takes as input a state and a transaction and transmits as output a new state.
In a blockchain the 'state' is represented from the collection of all the coins, or rather of the Unspend Transaction Output (UXTO). Each UXTO has a specific name and owner defined by a public key (20-byte address). In detail every transaction contains

- one or more input strings, where each input contains a reference to an Existing UTXO and a cryptographic signature produced by a private key associated with the owner's address, which avoids that a user spends coins that does not own.
- one or more output strings, where each output contains a new UTXO that must be added to the state.

Each transaction determines the switching of control from one user to the other, as already said, each operation is visible in the blockchain and can be displayed with a hexadecimal editor.

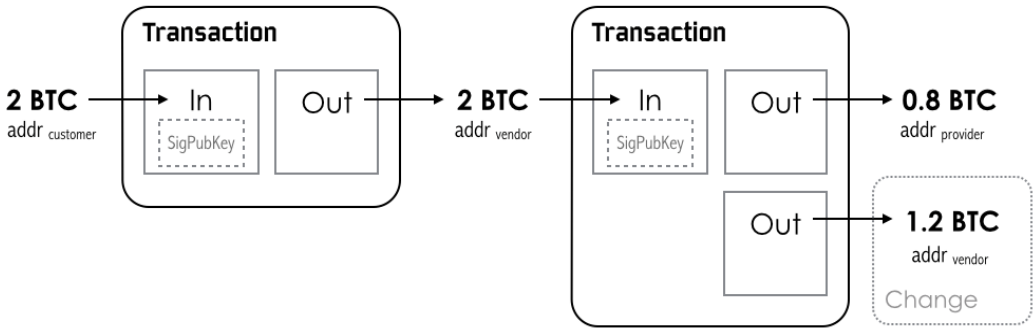

***Figure 2.4.1**: Example of Bitcoin Transaction*

A user who wishes to exchange cryptocurrency on the blockchain must be equipped with a wallet, a pair of public-private keys: the first is an alphanumeric identification code, a public address that allows to receive the money, while the second is a private key that only the owner knows and

thanks to which the money can be sent. The status of transaction is tracked and traced by a transaction hash, that is an identifier generated during the transaction performing.

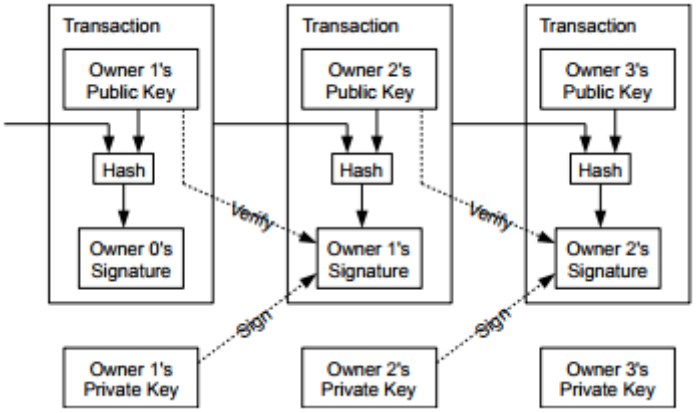

***Figure 2.4.2**: Transaction's Scheme*

The security of each transaction can be evaluated in accordance with a block height, that is the number of the blocks in which this transaction was recorded. Being the block height numerated sequentially, from it the number of block confirmations can be derived, which represents the number of blocks that have been verified by the network after the considered block: greater is the number of block confirmations after the considered block, more secure is considered the transaction included in that block.

The same presence of the timestamp, that marks the moment in which the operation takes place and is recorded, provides a tamper proof tool to know the chronological order of the transactions.

As will be explained in the following section each transaction must be verified/validated by a consensus protocol. Specifically, each transaction carried out, is sent to all the nodes which collect new transactions in a block and participate to a consensus mechanism, in order to verify and to mine the next block. A block, to be accepted must contain inside it only valid

transactions and there should be no double spending (i.e., spending the same money twice) [12], [13].

2.5 Mining and Consensus Mechanisms

The term mining refers to the process of creating a new block realized by means of entities belonging to the network, called miners. For a new block to be generated it is necessary:

- to determine the type of transaction and verify its validity
- to select the last block of the chain, including its hash in the new block
- to resolve a proof-of-work (POW), ensuring the validity of the blocks.

At the structural level each block consists of a timestamp, a random number (nonce), chosen by the miners, a reference to the previous block (hash of the previous block) and a list of transactions that have taken place since the previous block; in this way a persistent and increasing chain is created during time, which is constantly updated as a ledger [14].

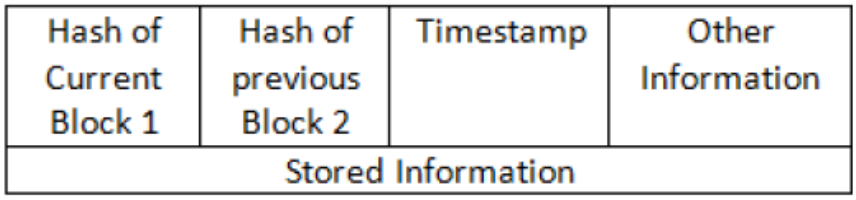

***Figure 2.5.1**: A block structure [15]*

The main mining activity is the Proof-of-work, that is the stage in which the miners continuously try to solve cryptographic puzzles in the form of a hash computation. The POW is based on the use of the hash algorithm SHA-256, whose output is unpredictable, which allows to create a new block through a set of attempts with the repeated increment of the nonce, in order to find a

lower hash value of a target required by the PoW itself and dynamically adjusted. The dynamic adjustment of the target value allows to keep constant the time that the network takes to produce a new block, which is approximately 10 minutes. If the hash digest does not match the fixed target, the nonce is incremented and the contents are rehashed until a valid solution is found. Miners who find a valid solution are then rewarded for this. Once a new valid block is found, it is added to the blockchain, usually in the form of a tree. The cryptographic puzzle is computationally complex to solve, but its solution is easily verifiable within the network. Each decentralized system has its own cryptocurrency, etymologically this term derives from the fusion of the words "cryptography" and "currency". It is an equal and decentralized digital asset used, in a P2P network, to pay users, who employ their computational power in order to solve the cryptographic puzzle to confirm transactions.

Different blockchains utilize varying types of proofs to determine which miner's block will be appended next, in fact not only POW but also Proof-of-Stake and Proof-of-Burn are used.

The first type, suitable for contexts with low computing power, distributes the right to mine according to the amount of currency owned by miners, the second instead allows the miners to win the dispute by proving to have made a 'sacrifice'. In this mechanism the miners' possibility of being selected to extract the next block is proportional to the number of coins that are 'burned', or invested in an irretrievable way, without the possibility of recovering them [8], [16].

It is also possible to find in the blockchain ecosystem implementations of different consensus mechanisms like Round Robin and Practical Byzantine Fault Tolerance (PBFT), whose goal is always to agree on a common truth about adding new blocks to the decentralized and shared ledger.

Round Robin is mostly used in blockchains where mining is restricted only to selected identifiable entities. This prevents the problem of the mining process monopolization: a fair non-monopolized blockchain is generated by

permitted miners, that create a finite number of blocks, in rotation, in a given time window [17].
PBFT, unlike PoW, does not require hashing to add a new block to the chain. In particular, in this type of mechanism the nodes are subdivided into primary and secondary. The execution process is divided into four phases: pre-preparation, preparation, commitment and response: in general, the primary node, a.k.a. leader, announces the transaction that the group must approve. The other nodes sign the transaction validating the correctness of the record and send their response to all the other nodes and to the primary node. The opinion of each node will be published on the network and the status supported by more than two thirds of the nodes is seen as the correct one. This consensus method is less computational effortful than other methods but has an anonymity cost on the system. Indeed, being each decision the result of message-heavy exchange, in order to preserve the anonymity of each of them more resources are required [17], [18].

2.6 Resolution of inconsistencies

Due to the latency of the network, the blockchain, being decentralized, is not always consistent. Different blocks could reach different nodes at different times. As a result, from the structural point of view in the blockchain network it is possible to find states of temporary inconsistency with respect to the normal network structure.
Generally, unanimous consensus amongst the network nodes results in a single blockchain, with verified transactions that the network asserts to be correct. When this unanimous consensus among all the nodes of the network is not reached, a state of inconsistency of the blockchain, called fork, is recorded. The term fork means a doubling of the chain of blocks. It represents, in fact, a condition in which nodes in the network have different views about that blockchain state. Fork generation depends essentially on one

factor: the difficulty of PoW. Greater is the difficulty of the problem to be solved, longer is the time takes to find the solution. The 10-minute mining time is therefore a compromise between block generation speed and fork probability.

The network itself can also form forks, in order to make a variation of the consensus mechanism used. Indeed, through these, it is possible to propose an update of the network, carrying out general improvements.

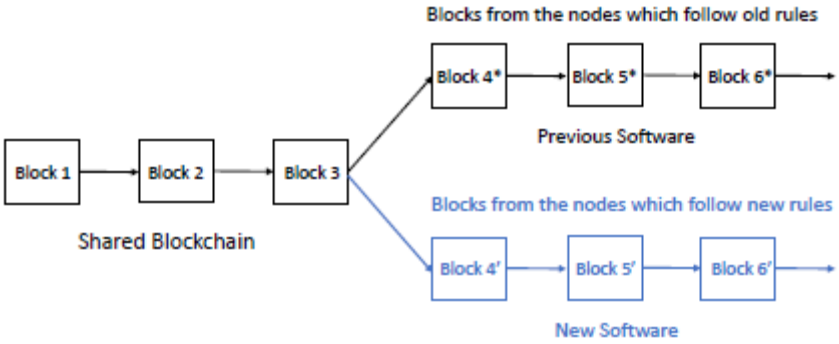

***Figure 2.6.1**: Blockchain forking [11]*

Indicating the forks as the points where a single ideal chain is split in two or more valid chain, they can be divided into two macro-categories:

- *soft forks*: occur when some blocks appear invalid to post-fork nodes. A slight backward compatible splitting is created with the previous blockchain. This means that the changes made do not add anything extra to the existing rules, rather they generate only restrictions
- *hard forks*: occur when new blocks that the network accept appears as invalid to pre-fork nodes. They occur when the blockchain protocol is altered in a non-backwards compatible way. Therefore, necessarily two independent main chains are formed: the nodes that accept the new rules will no longer be able to accept the old ones.

From a more formal perspective, by using the consensus finality concept, firstly proposed by Vukolic [19], it is possible to analyse the fork occurrence in blockchain, according to the utilized consensus mechanism. By defining

the consensus finality as the “impossibility of reaching consensus without fully distributed agreement”, having or not having the consensus finality within the network is equivalent to the impossibility or the possibility of having forks. Consensus mechanisms such as PBFT enjoy the consensus finality: all the involved parties before the consensus, reach an agreement through a heavy exchange of messages. This does not happen for network in which have no communication between the nodes in deciding the validity of the blocks: it means that two or more blocks can be validated at the same height, thus creating the basis of fork occurrence. A fork occurs predominantly in the blockchains that are based on PoW, in which the block validation is the result of a computational challenge, and that does not enjoy consensus finality. Generally, in these networks, "longest chain wins" rule counts to solve the condition of the forks.
A distinction between two other structural inconsistencies, stale blocks and orphan blocks is shown in the figure 2.6.2. These can occur with the consensus process that can leave valid block out of the blockchain.

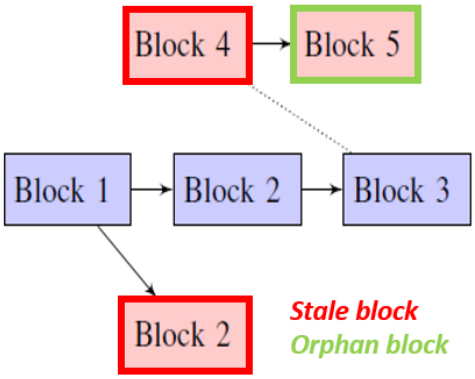

Figure 2.6.2: *Stale vs orphan block.*

The firsts are blocks that were successfully mined but not accepted in the current best blockchain. The presence of the stale blocks is recorded mainly in the public blockchains and it is due to race condition. In creating the next block of the chain, once the miners have found several valid solutions, it is

possible that the network will eventually only accept one of the winning blocks by rejecting the others. Blocks that are not accepted are stale blocks, and as such will not be added to the main blockchain (they are not part of the main blockchain).
While orphan blocks mean those blocks that do not have their parents in the blockchain; that are, all those blocks whose parent block's hash field points to an unauthentic block that is detached from the blockchain. The presence of the orphaned blocks is recorded mainly in system characterized by a small average block computation time [18].

2.7 Ethereum

Ethereum is a public blockchain with a basic programming language built in it, Turing complete language. This essentially means that within Ethereum is possible to write programs that can solve any reasonable computational problem.
Born in 2013, thanks to Vitalik Buterin, a Russian developer, Ethereum, with the presence of this language provides an alternative protocol for the building of decentralized applications where a high degree of application security, rapid development time and interoperability of different applications are essential. These applications are called Smart Contract. The concept of a smart contract, meant as computerized transaction protocol to executes the terms of a contract, was firstly introduced by Nick Szabo, a computer engineer, in 1994 [20].
These contracts, being self-executable under certain conditions, are called "smart". The runtime environment for smart contracts, the calculation centre that allows its execution, is called Ethereum Virtual Machine (EVM). It implements and makes possible the network consensus mechanism. It, also provides a safe, isolated and protected environment from the rest of the execution processes. Ethereum is, therefore, a programmable blockchain that

provides standardized and predefined "operations" but allows users to create different types of applications, not necessarily limited to cryptocurrencies. In fact, it is designed to be adaptable and flexible and to easily create new applications. Each application is written with an objected-oriented, high-level language, called Solidity, and subsequently converted to bytecode, where each byte represents an operation performed inside the EVM. With Solidity, developers can write applications that implement a self-defined business logic integrated into the smart contract, providing a non-repudiable and authoritative transaction.

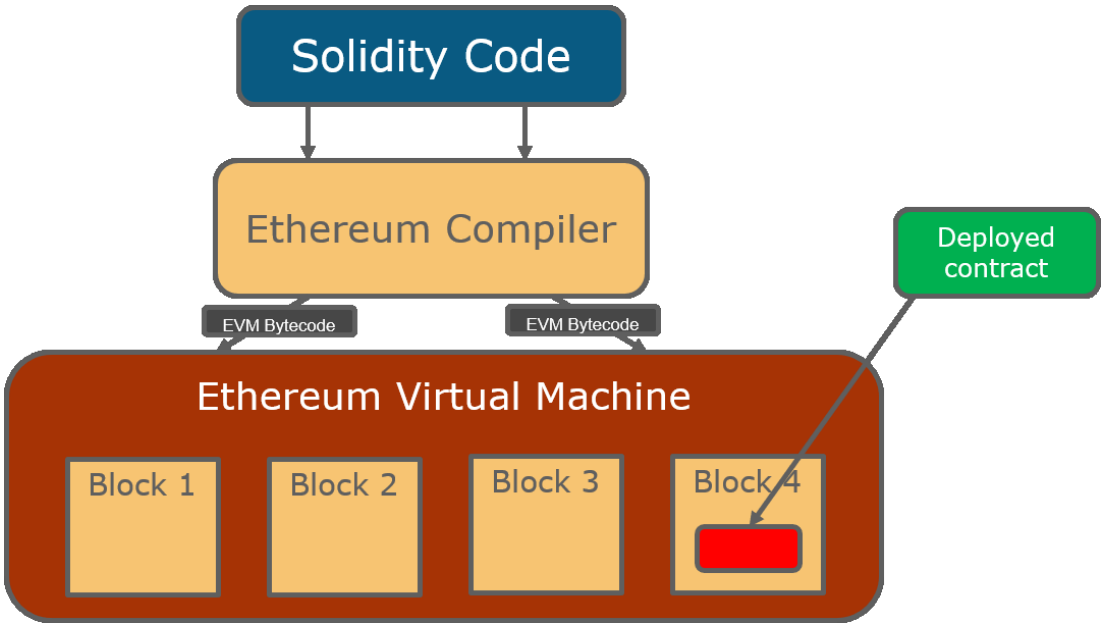

***Figure 2.7.1**: Ethereum Virtual Machine*

Smart Contracts are presented as a shared and replicated ledger aimed at improving efficiency and reducing costs compared with other solutions provided by pre-existing systems based on duplicate business logic and a consent protocol based on reconciliation. Thanks to Smart Contracts, decisions taken consensually in the network can be bonded rather than being subordinated to a central body that authorizes all activities. Although they are called contracts, they must not be completed or filled out. Smart contracts

are simply algorithms, used to execute portions of code that are affected by a transaction; they exercise direct control over the variables involved in a transaction, to ensure traceability and transparency.
In this distributed computing platform, each user can give inputs, obtaining the corresponding output automatically. The basic functioning can be described by the logical succession “if-this-then-that”: a common interest is identified between the parties, a contract that includes the desired conditions and effects is then written between them. The contract is guaranteed by the same Ethereum blockchain and once the network's consent is obtained, the contract meets the desired conditions [20].
The realization of decentralized applications within the Ethereum protocol is rewarded through the cryptocurrency Ether: it is the "virtual currency" with which are remunerated the computational resources of the first node that mines the block successfully, but even before, it represents the processing power necessary to produce the contracts themselves. The Ether is basically a digital asset which is treated as a cryptocurrency exchange with the ticker symbol ETC.
In Ethereum, the status consists of objects called "accounts". Each of them has an address of 20 bytes and the state transitions are direct transfers of value and information between accounts.
Each Ethereum account is structurally composed by the same components, which represent different features, according to the account type. Indeed, in Ethereum, it is possible to distinguish two different types of accounts, that are [21]:

- *Externally owned accounts (EOAs)*, controlled by private keys. These have no code and can send external messages by creating and signing a transaction. Each EOA is structurally composed of:
  -*nonce*: counter that represents the number of transactions sent from the account’s address
  -*balance:* number that represents the financial statement of the account

-codeHash: the hash of the empty string

-storageROOT: the 256-bit hash that encodes the account storage contents. It is initially empty for standard.

- *Contract Accounts,* controlled by their contract code.

Whenever a contract account receives a transaction or message from other contracts, its code is activated, allowing it to read and write to internal storage and send other messages or create contracts in turn. At the structural level, in each Contract Account are present:

-nonce: counter that represents the number of contracts created by the account

-balance: number that represents the financial statement of the account

-codeHash: hash of the empty string is the contract code that can be executed by the network's nodes -containing functions that can be recalled, it can be compared to an object of a language object-oriented

-storageROOT: the 256-bit hash that encodes the account storage contents. It is initially empty for standard.

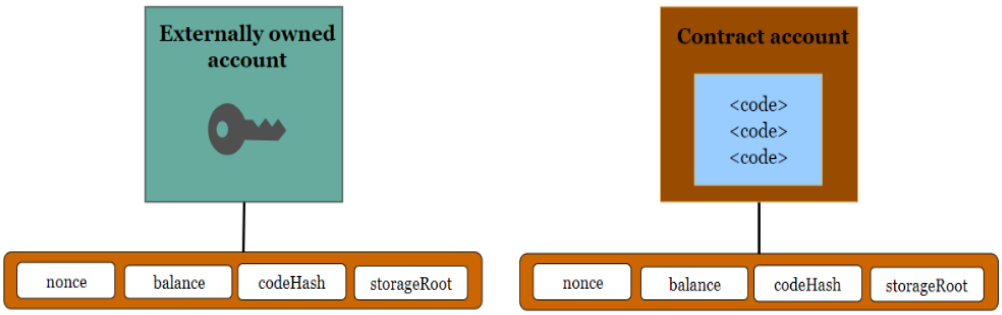

Figure 2.7.2: *Ethereum accounts [21]*

The term transactions in Ethereum means the whole amount of signed data containing a message to be sent from an external account. Within a transaction there are:

- the receiver of the message
- a signature identifying the sender
- the amount of Ether to be transferred to the recipient, which is the subject of the transaction
- an optional data field
- a value representing the maximum number of computational steps that can be executed in the transaction
- a value equal to the commission that the sender pays for the computational step.

The message that is sent by contracts to other contacts is partly like a transaction; structurally it is composed of the same fields excluding the value equal to the commissions paid by the sender. It is never serialized, and it is a virtual object that exists only in the running environment, without ever being published on the blockchain: can be considered as an internal message that is produced by a contract and not by an external actor, which happens instead for the transaction. A message is, therefore, the result of the execution of a call, during the implementation of the code by a contract [14], [22], [20].

CHAPTER 3

GDPR-Compliant Biomedical Blockchain Architectures

3.1 Blockchain Application in Biomedical Domain

Security, transparency privacy, integrity and non-repudiation are very delicate issues for the biomedical sector, characterized by an enormous attention to the management of data of extreme importance and sensitivity.

The blockchain represents, for the healthcare field, an element of innovation that provides a secure architecture for the exchange and sharing of data within the healthcare system. At the same time, it provides patients with full control of their information, by allowing a safe management of healthcare data.

In the healthcare sector, a system founded on blockchain is based on the centrality of the patient and transparency, aimed at improving the safety and efficiency of the system itself. Being blockchain-based medical systems a hot topic, in recent years several models have been introduced that give the blockchain a central role in the validation and storage of health data.

Yue et al. [23] proposed an APP for sharing healthcare data, where patients control, send and own their data easily. Qi et al. [24] proposed MeDShare, a

system that can address the problem of medical data sharing among medical big data servers in a trust-less environment. Ekblaw et al. [25] presented MedRec, a decentralized record management system to implement electronic health records by using blockchain. As defined by the International Organization for Standardization (ISO), electronic health records include “any computer processable repository of information regarding the health status of an individual”. It is possible to identify two main types of electronic health records: medical records, produced mainly by hospital departments and generally focused on medical care; and personal health records, controlled by the patient and generally containing information at least partly entered by the patient. Blockchain technologies show a potential to address several security and integration issues regarding health records [26]. To understand how the blockchain can improve the efficiency and security of storage and sharing of healthcare data it is necessary to consider its key features. As previously described, the blockchain is based on three principles:

- data are stored in an immutable public transaction ledger
- it is implemented through a decentralized network of computing nodes that makes them robust against faults and DoS attacks
- each transaction is described by data that are usable to all users.

This decentralized network is based on the pseudo-anonymity and on a public key infrastructure that allows the blockchain to encrypt content in a way that makes decryption expensive and prohibitive.

In order to implement a system of health records based on blockchain, documents related to the patient's health status should be published on a healthcare blockchain. Each blockchain transaction could then be identified by a digital signature of the healthcare provider, which identifies the origin and will have as its recipient the patient's ID. Patients could retrieve contents from the blockchain by means of mobile or web-based applications. Among the advantages of this system of record-keeping, there is the centrality of patients, who own and directly control his/her own health data.

It is possible to exploit blockchain in order to store two different kinds of health data. The first ones are health data collected by healthcare professionals, while the second ones are data collected by means of Internet of Things (IoT)-based devices, during home remote monitoring of the patient. Both are entered in blockchain after an acquisition process, digital conversion and validation. These data, shared on a blockchain, are subject to an exposure risk: in order to preserve their privacy and security, it is necessary that anonymization/pseudonymization procedures is carried out before moving them onto the blockchain [8], [26], [27], [28].

3.2 Blockchain's compliance with the GDPR

Though blockchain and GDPR started with very different goals—creating a currency independent from a central authority versus introducing data protection laws—the two initiatives are aligned on the principles of secured and self-sovereign data (individuals in charge of their own data).

The sharing of information in the blockchain network is made easier thanks to the creation of an open timestamp ledger. The working principle of the network is built on the action of a group of nodes that provides the storage of synchronised copies of the same data.

The blockchain, thanks to the characteristics of immutability, transparency, verifiability, data encryption and operational resilience offers a radically different approach to data security, compared to centralised system. Data security develops around three basic elements: confidentiality, integrity and availability, known by the acronym CIA. The concept of confidentiality corresponds to the ability to protect data from all individuals who are not allowed to see them, limiting the access to information. It is a prerequisite for ensuring privacy and it is usually guaranteed through cryptographic techniques. Integrity is the ability to prevent data modification in an unwanted and unauthorised manner. integrity allows to maintain reliability,

consistency and accuracy of data. Availability refers to the possibility of accessing to the data when requested.

In blockchain both integrity and availability are fulfilled, the architecture of the network is a key feature for their achievement; on the other hand, it clashes with the concept of confidentiality and privacy, whose achievement turns out to be more challenging [15].

Indeed, using a public blockchain to store and exchange transaction data creates major privacy hurdles: by default, all data entered in the ledger is in clear. Since each node has a complete copy of the ledger, confidentiality of data cannot be preserved. Therefore, if an attacker can access the network, he may also gain access to the data. Although the public nature of the blockchain allows that specific access and authentication controls are not necessary, there are some blockchain implementations that begin to address the issues of data privacy and access control. It is desirable that security controls, such as access controls, are implemented directly at the application level. Examples of private blockchains, in fact, solve this issue by means of an authentication protocol, in order to guarantee secure access to data [29].

Due to the Blockchain immutability by design, this shared, decentralized and public ledger technology is not considered in full accordance with the GDPR, but there are some particular use cases and applications that can be seen as GDPR-compliant [30].

Being hashing a one-way transformation and being the blockchain based on cryptography and hashing, once personal data and information are stored on blockchain they are immutable and cannot be deleted or modified; it is clear that this feature of the blockchain clashes with one of the key requirements of the GDPR, known as Data Erasure or Right to be Forgotten. It is important to understand how the blockchain technology can adapt its immutability of data to the GDPR. Data erasure can be implemented in different ways in this technology that guarantees that nothing will be deleted.

One solution is to encrypt personal information written to the system, such that when the time comes, the destruction of encryption keys ensures that sensitive information is no longer accessible [9].
Another possible solution is to provide proof of integrity by writing only the hash digest of data into the blockchain, while these are stored outside the blockchain. This maintains the integrity of the transactions, while allowing the possibility of deleting data, leaving only residual traces of information in the blockchain.
Furthermore, the decentralized model of the blockchain technology may be an issue itself with respect to GDPR compliance. In particular, since the roles concerning data management in the blockchain architecture are not specifically defined, no data controllers and data processors can be held legally responsible for the processing of the data themselves [30].
However, the National Commission on Informatics and Liberty (CNIL) observes that participants, who have the right to write on the chain and who decide to send data for validation by the miners, can be considered as data controllers. The CNIL is an independent French administrative regulatory body whose mission is to ensure that data privacy law is applied to the collection, storage, and use of personal data. Regarding the GDPR, the CNIL, providing solutions for a responsible use of the blockchain in the context of personal data, states the following. Blockchain participants, by defining the purposes and means of data processing, can be defined as data controllers when the processing of the data is not limited to a strictly personal activity. It is therefore clear that in a decentralized structure not all participants can be considered data controllers. The same miners who have the function of validating the transactions, not defining the purposes and the means of the transactions, cannot be considered as such [30].
In the same way the CNIL believes that data processors exist in a blockchain, according to the GDPR: such as smart contract developers who process personal data on behalf of the data controller or in some cases, the miners. In fact, within the meaning of the GDPR, could be considered data processor

all those users that follow the data controllers' instructions when checking whether the transaction meets technical criteria [31].
While it does not appear to be fully GDPR-compliant, the blockchain seems to be helpful to increase transparency in data processing. In fact, each transaction added to a blockchain is digitally signed with an indication of date and time. This means that any transaction is identifiable on time.
This feature ensures that the authenticity of the signature on a file is not duplicated, since each transaction is cryptographically associated to a user. Therefore, the reliability of the system is increased through the detection of tampering attempts or fraudulent transactions.
Moreover, any new transaction added to a blockchain will result in changing the global status of the ledger. Consequently, with each new iteration of the system, the previous state will be stored, resulting in a fully traceable chronological register. From the point of view of computer security, this provides companies with an additional level of security that the data is authentic and not tampered [30].

3.3 GDPR-compliant blockchain paradigms

Although during years several blockchain applications in the biomedical domain have been presented, it is possible to collect the protocols found in the literature in a few basic paradigms in order to provide a synthetic and clear abstraction of the complex structure and complex working principle of the GDPR-compliant blockchains used in the biomedical sector. In principle, medical data integrity can be ensured by storing hashes in the blockchain, by using a cryptographic version of the data or by putting the medical data in a local storage, outside the blockchain [26].
As a matter of principle, the modus operandi of the paradigms that will be described next is aimed at preserving data privacy, in a sector without

standardization concerning archive keeping. Through the main characteristics of the blockchain - decentralized consensus, trustless and immutability - safe and efficient platforms are provided for the health data management. In order to address the main difficulties that follows from the GDPR regarding the management of sensitive data, the following paradigms propose new blockchain models, characterized by different approaches to sensitive data storage, ensuring the GDPR compliance itself. In particular, analysing them, it is possible to notice how the right to be forgotten is commonly addressed by using multiple cryptographic protocols on an immutable distributed ledger.

To maintain the tracking of the data life cycle, the first solution widely found in the literature is the one that inserts in the blockchain not the private healthcare data but the hash value of the data item. This paradigm could simply be described as follows: on-chain hash/off-chain healthcare data. It is mainly based on two separate storage systems: private data is always saved in a local storage instead of the blockchain and, after consent is given, only the hash of this data is inserted into the blockchain. This process generates a blockchain transaction depending on a key and on a value. The key, in this case, specifies the hash digest written into the blockchain, while the value represents the reference to the off-chain data repository. What characterizes the blockchain storage models that are based on this paradigm is the removal of the link between the blockchain and the private health data: in order to preserve data confidentiality, health data are entered in the ledger not in clear, but only after a unidirectional hash computation. User, controller and processor are the actors of the developed prototype, that is aimed to provide a secure, transparent and GDPR friendly healthcare data management system [32], [33], [34].

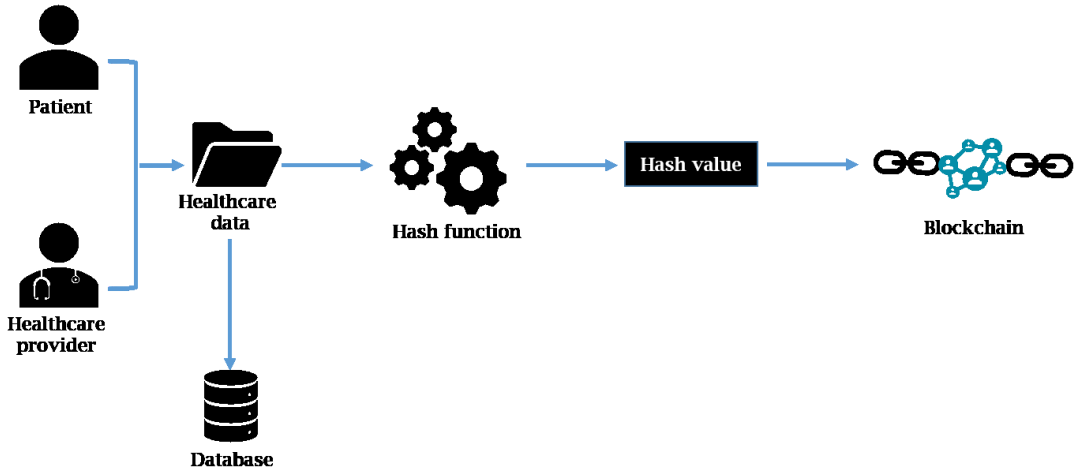

Figure 3.3.1: *Health data storage according to the first paradigm*

Cryptography in blockchains is primarily used for two purposes: securing the identity of the sender of transactions and ensuring that past records cannot be tampered. In the literature there are blockchain models which employ advanced cryptographic techniques for further security, that is, for storing encrypted data in order to resolve the data preserving issues. It is possible to schematically define the operating principle of this paradigm. As it happens for the previous paradigm, in order to guarantee the confidentiality of the health data, data are entered in blockchain not in clear. In this specific case the cryptographic version of health data is stored on the blockchain. Encrypted versions of the data are obtained by means of a public or private key encryption.

Thanks to a tailor-made architecture, the paradigm executes the project's core principles: compliance to the GDPR, control of the data access, and the guarantee that the data won't be stolen. By making, also, an analysis of the paradigm it is possible to notice how this guarantees Pseudonymization, Privacy, Integrity Accountability and Security [9], [35], [36], [37], [38].

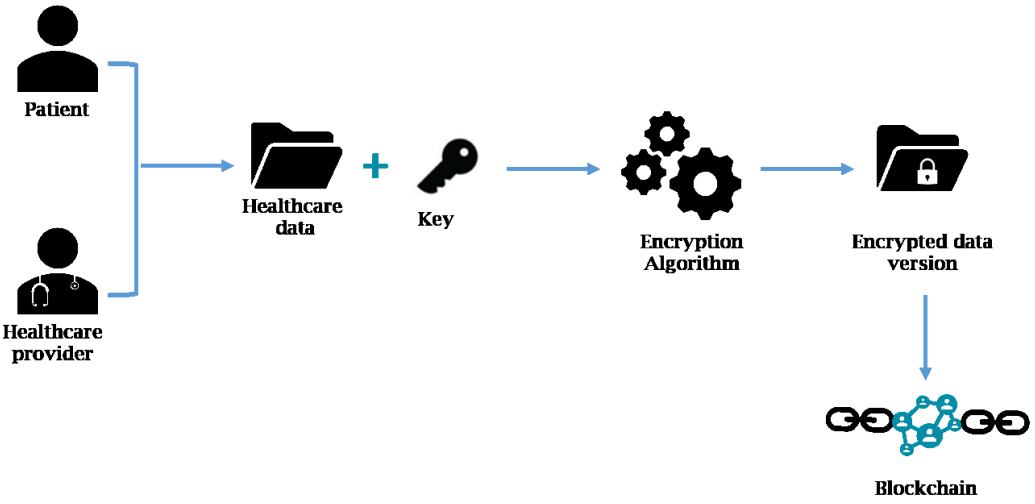

Figure 3.3.2: *Health data storage according to the second paradigm*

Unlike what has been seen until now, in the third and last paradigm derived from the analysis of the blockchain GDPR-compliant models, the storage of healthcare data is performed off-chain. The blockchain, although it plays a central role in the filing of healthcare data, is used here only for the purposes of data management and data integrity, besides the management of consent and authorization. The blockchain stores the pointers to the data, memorizing the position of the data stored pursuant on a local database, thus regulating their use. Each pointer consists of a query string that, if executed, returns a subset of patient data; this query string is placed with the hash of the data to guarantee the non-alteration of the data at the source. Additional information indicates where the data is in the local database, in order to allow their retrieving by healthcare providers. The data are, therefore, inserted into a compliant cloud in an appropriate geographical domain so that it can be integrated, monitored and verified by the patient. The entered data are thus accessible to doctors and specialists in order to provide global medical solutions to patients. This paradigm, using blockchain technology, guarantees the non-modification of the data and produces a safe, fast and reliable way to

exchange data. Solutions like these could be useful implementations to mitigate the transparency/confidentiality issues [25], [39], [40], [27].

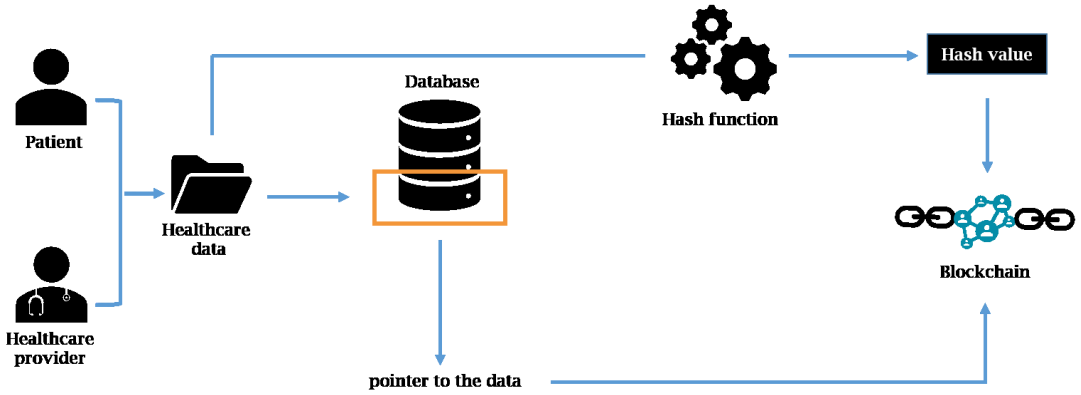

***Figure 3.3.3**: Health data storage according to the third paradigm*

3.4 Core technology components of analysed models

The analysed paradigms are based on core technology that can be schematically divided into four levels: communication, archiving, security mechanism and consent mechanism. In all the paradigms considered in this analysis, to achieve a reliable mechanism for secure and efficient medical record exchange, communications take place according to P2P technology through multicasting routing on Ethereum: each node posts the activities involved, patients and service provider's records, to the blockchain and forwards information to all other nodes, without the presence of a central authority. The security of the blockchain is instead guaranteed differently in the three paradigms, through a variety of cryptographic principles for data encryption and privacy protection. Therefore, depending on the considered paradigm, data storage takes place, as previously shown on-chain, off-chain or with a combination of both. Furthermore, several consensus mechanisms

to manage medical blockchain-based ledgers model have been proposed: proof-of-work, practical Byzantine fault-tolerance and round robin.
The analysis that allowed to extrapolate the most commonly used blockchain paradigms in the health field was conducted on different blockchain models; although in most cases they are still in the design phase, some seem to have a promising future. The extrapolation of each single paradigm described above has occurred by grouping blockchain models that preserve data privacy according to the same type of storage and security mechanism. In detail, the paradigm characterized by storage of healthcare data hashes in blockchain is the result of system architectures evaluations such as BlocHIE [41], MyHealthmyData (MHMD) [34], Blockchain based Personally Identifiable Information Management System (BcPIIMS) [42] and Blockchain-based eHealth Model (BEIM) [43]. Furthermore, the analysis has considered MediBchain [9], MIStore [44] and Advanced-Block Chain (ABC) [45] models where data are encrypted and stored on-chain (embedded into transactions). Then, the MediChain [39] model has also been considered, concerning to the off-chain database storage.
BlocHIE is a blockchain-based platform for healthcare information exchange, it is based on two loosely coupled blockchains, that exploit the difference of requirements of healthcare data during the storing and sharing processes. The two blockchains are respectively EMR-Chain, which contains EMRs, and PHD-Chain which includes personal healthcare data, generated by each patient. EMR-Chain contains data with a high criticality, which require high privacy and authenticity and moderate throughput, latency and fairness. In EMR-Chain techniques for off-chain storage in a distributed database and on-chain hash storage and verification are integrated in order to achieve privacy and authentication. Each EMR copy submitted to the blockchain network contains a timestamp, the hash value of the medical record, the hospital signature, the patient signature, a set of keywords, and an extra description. The hash value of the medical record is generated using a hash function such as MD5. This solution employs PoW, in order to have a model that provides

privacy, accountability and searchability. In order to store data from individuals, that require a high throughput, in PHD-Chain two fairness-based transaction packing algorithms are introduced [41].

Blockchain based Personally Identifiable Information Management System (BcPIIMS) is a blockchain-based model that bases its architecture on the differentiation of data in Personally Identifiable Information (PII) and Non-Personally Identifiable Information (NPII). Although it is not a model designed for the treatment of healthcare data, it proposes an off-chain blockchain architecture which uses both a local database and a distributed ledger to preserve a trustable PII life cycle, which could be re-adapted to healthcare, guaranteeing a data-storing architecture compliant with the GDPR rules. It is a Korean model that proposes a separate procedure of storing PII and the rest of the data. In this solution, the PII is saved in a local database and non-PII with the hash of the PII is stored in the blockchain, through a global approval procedure that makes tampering difficult. Trust among peers over the state of the ledger is guaranteed by the use of a Round Robin consensus algorithm: the nodes must create blocks in rotation and in order for a new block to be approved is required that, at least 75% of the total nodes must agree with the choice of its addition to the chain. Once a new block is added, the personal data can be shared and eventually modified or deleted. The model was implemented in 2018 using the help of Multichain 2.0 getting as a result a private blockchain, where the network consensus requires minimum difficulties to be achieved [33].

Blockchain-based eHealth Model (BEIM) is a blockchain-based model that ensures secure data exchange, data sharing and information integrity in eHealth system, an information system that manages medical data, such EHRs, permissions given by a care subject, audit logs, billing information. In BEIM, the blockchain in mainly used in order to create a rigorous and tamper-resistant data register. The storage of medical data occurs off-chain and in a permissioned blockchain the hash digests of the data are kept secure and in turn, as additional level of security, are encrypted through an asymmetric

encryption algorithm. This model allows information removal, which is a legal requirement in EU countries, using the Practical Byzantine Fault Tolerant algorithm as consensus mechanism. In fact, this prototype provides the solution to the problem of transactional transparency based on a blockchain with the option to remove documents or logs. The model allows the removal of any record. The only trace of a deleted record is two transactions with no information about the content or the origin of that document [43].

MyHealthmyData (MHMD), born as a project funded by the European Commission, it aims to achieve by 2020 the first open network of biomedical information based on the connection between patients and health organizations. The key elements on which this system will be based are the following:

- use of a distributed digital master tree on which the information concerning the distributed storage of the health data is trimmed in hash-based language code. Notably, by using a hash function each health data is mapped in a fixed-size digest
- minimization of the risk of fraudulent activities by distributing control over the entire network
- differentiation of the use of data through dynamic consent, according to the preferences of the patients. The presence of smart contracts allows to grant, deny or withdraw the consent to the access to the data for different uses.

This GDPR compliant blockchain model provides the traceability of the data life cycle, an efficient implementation of the right to be forgotten and a localization of the usage of the data. It can be described as an abstraction model, in which the uses are part of a public blockchain network that orchestrates the secure data sharing process. The healthcare data in clear are always stored in the data controller's facility and never in the blockchain; what is instead insert in blockchain is the hash digest of the same data [32], [34].

MediBchain is a blockchain-based platform proposed in 2017: it is a patient centric healthcare data management system, which uses the blockchain for storage to achieve privacy. Indeed, starting from the blockchain, data preserving vulnerabilities are solved by using cryptographic function. The cryptographic versions of the health data are stored on-chain. A high-level view of the system protocol provides the differentiation of some entities such as a data sender (patient), a data receiver (medical care provider), a registration unit, a private accessible unit (PAU) and lastly a permissioned blockchain, as a repository of health data, which requires authentication for access. The health data are encrypted with a symmetric encryption technique, according to the following encryption function (1):

$$\text{Enc}(\text{key}, \text{HealthData}) = \text{Encrypted User HealthData} \quad (1)$$

While, the private data are encrypted by means of a public key encryption technique like Elliptic Curve Cryptography, that provides encryption, signature and key exchange approaches. Privacy, pseudonymization, integrity, accountability and security of healthcare data are maintained by the system: only registered parties can be able to interact with the system, the security of the health data is provided by data symmetric encryption, that keeps them safe from eavesdropper. In a state of disguised identity, no entity can identify any system party and on the same time only authentic party's interaction is ensured. In MediBchain the data accountability is system centric: in order to interact with a particular block is required to each involved party to hold his own ID-block [9], [46].

MIStore, differently from the models summarized up to now, it is a blockchain-based medical-insurance system proposed by Zhou et alt. in 2018. This system, using the Ethereum blockchain, provides an architecture that allows the medical insurance related health record data encryption and consequently the immutable storage on the blockchain for future use by medical insurance companies. The basic instance of the MIStore model

contains 4 different parties, that are patient, hospital, insurance company and n servers. In the implementation of MIStore, ECDSA, ECIES and SHA-256 are respectively used as signature scheme, encryption scheme and hash function. In each transaction health data are encrypted with ECIES: a hybrid cryptosystem based on the computational Diffie-Hellman problem, that combines the convenience of public-key cryptosystem with the efficiency of a symmetric-key cryptosystem. By using the Practical Byzantine Fault Tolerant algorithm as consensus mechanism, it is assumed that 2/3 of MIStore's nodes are honest in the system. MIStore includes two different kind of nodes, record-nodes and light-nodes: the firsts are responsible of the maintenance of the blockchain and in turn store the entire copy of ledger list. While for the light-nodes this does not happen: in order to verify a newly published block, they only keep the block-headers, without transaction information [44].

Advanced Block-Chain (ABC) Architecture is an approach proposed as a solution to supply a reliable mechanism for secure and efficient medical record exchanges. In the ABC architecture each occurrence of healthcare record generation is published on the e-Health Block Chain, in a block format. Each block is backwards dependent with the previous block and at the same time linked with the next one. Healthcare data are encrypted and stored on-chain, specifically, encrypted data are embedded into transaction; each block of the chain contains the block ID and the hash of the previous block, the Timestamp, the encrypted patient activities, the digital signature of the healthcare service provider and patient [45].

MediChain is a Chinese project that combines the blockchain technology with an off-chain database storage. It allows the storage of medical data off-chain, in an appropriate geographic domain. It is a Medical Big-Data Platform that gives access to doctors and specialists anywhere. Patients, doctors, and hospitals put healthcare data into an off-chain database storage, that becomes part of the MediChain ecosystem and pointers and rules on usage and anonymity are stored in the blockchain. MediChain utilizes the

characteristics of blockchain, such as decentralized consensus, 'trustless' and immutable storage in order to provide a a function capable of mapping the healthcare data localization [39].

The core technology components of each medical blockchain-based model are respectively synthetized and grouped, according to the security mechanism, in Table I.

Table I: *Core technology components of analysed models*

Healthcare blockchain-based model	***Comunication***	***Type of Blockchain***	***Storage of Medical Data***	***Security Mechanism (Privacy Protection)***	***Consensus Mechanism***
BlochHIE	P2P on Ethereum	Permissionless	On-chain/Off-chain	Hash version	PoW
BcPIIMS	P2P on Multichain 2.0	Permissioned Private	On-chain/Off-chain	Hash version	Round Robin
BEIM	P2P on Ethereum	Permissioned Public	On-chain/Off-chain	Hash version	PBFT
MHMD	P2P on Ethereum	Permissionless	On-chain/Off-chain	Hash version	PoW
MediBchain	P2P on Ethereum	Permissioned Public	On-chain	Symmetric Encrypted version	PoW
MIStore	P2P on Ethereum	Permissioned Public	On-chain	Asymmetric Encrypted version	PBFT
ABC	P2P on Ethereum	Permissioned	On-chain	Encrypted version	PoW
MediChain	P2P on Ethereum	Permissionless/ Permissioned	Off-chain	Local Database	-

In Table II, it is possible to observe, for each model considered in the analysis, which are the fulfilled GDPR requisites. Each model, through the combination of blockchain technology and particular security adopted mechanisms, manages to guarantee the integrity and confidentiality of the health data. Equally for each model, using different storage mechanisms the right of erasure, imposed by the GDPR, is guaranteed. If on the one hand the solutions that use on-chain storage fail to guarantee the requirement of geographical location of data, this GDPR requisite is instead guaranteed for model that provide off-chain storage solutions.

Table II: *Guaranteed GDPR requirements in the analysed models*

Healthcare blockchain-based model	*Guaranteed GDPR requirements*			
	Integrity	*Confidentiality*	*Data erasure*	*Territorial scope*
BlochHIE	√	√	√	√
BcPIIMS	√	√	√	√
BEIM	√	√	√	√
MHMD	√	√	√	√
MediBchain	√	√	√	✗
MIStore	√	√	√	✗
ABC	√	√	√	✗
MediChain	√	√	√	√

CHAPTER 4

Security Assessment of Proposed Paradigms

4.1 Vulnerability to cyberattacks

E-health, by using the blockchain technology, tries to provide facilities to improve patient health with an efficient and safe healthcare data sharing system. Although the blockchain is a computationally expensive tool, and although it requires a large bandwidth and huge computational power, it seems to be a promising tool to improve security, privacy and functionality of the healthcare data. The healthcare field represents an extremely dynamic, delicate and vulnerable sector, as the widespread use of new technologies inevitably intertwines with the delicacy and relevance of the data processed. This peculiar structure of health system consequently involves the presence of a multiplicity of "security-induced safety cases" by virtue of which cases of security breaches can entail significant consequences in terms of the effectiveness of the treatments or even lead to situations of risk for the life of the patient. The healthcare, according to the Vormetic report on data security 2016 is one of the sectors that is most vulnerable to cyberattacks

[47]. In order to prevent security threats, many architectures promise to correctly and accurately apply security and privacy measures. In an area where the accuracy of data plays a key role in the diagnostic process but also in the treatment of patients, cyber threats may affect the performance and efficiency of the systems. Although the blockchain technology with immutability, consolidated trust, distributed identity and consensus, and efficient and everlasting verification may seem an infallible technology for the security of healthcare data, it is worth noting that attacks are more and more sophisticated and effective and can cause enormous and irreparable damages.

4.2 Taxonomy of attacks and applicability evaluation

In order to assess the security of the previously described paradigms, the realization of different types of attacks has been considered. The feasibility of each attack in blockchain applications varies depending on the network topology, adversarial requirements, peer behaviour, and incentives [48]. In general, since each paradigm consists of medical blockchain-based models characterized by a P2P communication, a first class of attacks are those that aim at structural vulnerabilities. The feasibility of attacks on the blockchain structure itself is considered; Attacks that damage the type of communication within a P2P the network among equal nodes, are also considered. In the hypothesis of a structural attack, the forks and stale blocks are the ones that, by representing an inconsistent state of the blockchain structure, can be exploited by adversaries to cause confusion, fraudulent transactions, and distrust within the network.

As previously described, some hard forks have the simple purpose of updating the register, while others move away and may led to a split in cryptocurrency. During the propagation of the hard fork effects on the

network, some risks emerge that endanger the reliability and immutability of the transactions that have taken place. In fact, it may happen that the malicious subjects replicate the transactions just made on the original chain, on the newly created chain. This could create what is known as a **Replay Attack**. This kind of attack represent a well-known threat to network security. During the replay attack occurrence, the original messages are intercepted and are retransmitted literally; in this way any attack's attempt is considered, by the network security protocols, as a normal data transaction. This is because the original data comes from an authorized user and therefore, they are valid themselves. Therefore, a replay attack would imply that the two cryptocurrencies, that of the legacy chain and that of the forked chain could be unlocked, at the time of the sale, with the same private key, thus exposing the transactions at risk of be duplicated. The fact that these attacks do not exploit any data decryption makes them an effective attack approach for opponents who do not have to deal with complex cryptographic protocols. Protocol changes, known as hard fork, are therefore a particularly relevant risk for this type of attack. When these hard forks occur, it is theoretically possible for hackers to use replay attacks against blockchain ledger. A transaction processed on a ledger by a user, whose wallet is valid before the hard fork, will also be valid on the new version. As a result, a user who has received a certain number of cryptocurrency units from someone else through a register could switch to the other register, replicate the transaction and unlawfully transfer a number of units identical to their account a second time. The same considerations made previously about the occurrence of the fork clearly explain how a replay attack is favoured in blockchains that are based on PoW and PoS consensus mechanisms, instead of in the blockchain-based models that use PBFT [18], [49], [19].

Forks can also be caused by malicious aims such as implanting "Sybil nodes" that follow conflicting validation rules. In a P2P system, a **Sibyl Attack** relates to a dissociation of the personality: the attacker disrupts the network by creating a number of misbehaving nodes. In particular, it consists in

impersonating multiple identities in a network characterized by the absence of a central entity that verifies the identity of the users. An attacker could impersonate multiple distinct identities in order to subvert the integrity of the network by increasing its own voting power. In the Sibyl attacks not only, the attackers might put in the minority the honest nodes creating sufficient false identities, but might create the basis of a 51% attack, by changing the order of the transactions and preventing the confirmation of the same. In medical blockchain-based models that use PoW as a consensus mechanism, the creation of a new identity would entail the addition of computing power to solve the cryptographic problem, thus making Sybil attacks expensive for opponents. It is therefore clear that the use of PoW offers an intrinsic resistance to Sybil attacks. The threat of a Sybil attack remains valid for all the approaches that instead rely on other consensus mechanisms. In particular, the PBFT mechanism shows a susceptibility to Sybil attacks in situations where the network has few nodes [50].

The structural inconsistency of the stale block, instead, as will be seen later, can be exploited by possible attackers with the aim of depriving, in the network, an honest miner of his reward [18].

The blockchain technology, although relying on a decentralized infrastructure, like all online services, is prone to **Denial of Service (DoS) Attack**. A DoS attack is an attack aimed to disrupt the regular operation of the network by flooding the nodes. Typically, this is accomplished by overloading the target with a massive amount of traffic. This could produce a network congestion, denying certain nodes access to some data. It is possible that many malicious machines are directed to target a single network node: this attack mode is called **Distributed Denial of Service (DDoS) Attack**. By attacking the target in a distributed manner, it is much more likely to saturate the target of the attack, than with a single source. This operating mode is often chosen by attackers, also because expanding the origin of the attack makes more complex to go back to the source of the attack. In order to be feasible in a public blockchain, this type of attack requires an opponent

that must have control over the majority of the network hash rate. Therefore, an attack like this is very expensive. However, in PBFT-based blockchains a DDoS attack can be launched if the opponents control at least 33% of the replies, considering and assuming that with this consensus mechanism 2/3 of the participants are honest. An attacker may compromise the security of a service by delaying non-faulty nodes or the communication between them until they are tagged as faulty and excluded from the from the group of nodes that are in charge of the response task. In this type of blockchain it is assumed that the size of the network is known by the participants: in fact, an attacker can calculate the number of Sibyl nodes needed to carry out the attack and can obstruct the verification process within the network. Considering a network of size n the attacker should introduce f sibyl nodes such that n <3f+1: in this way the activity of the whole system will be suspended because for the transaction processing at least 3f + 1 approvals are required. It is therefore evident that in systems based on the PBFT mechanism this type of attack could be more feasible than in systems using other consensus mechanisms since for the attacker it is necessary to have control over only 33% of the nodes to be able to apply the attack [18].

By assessing the mining process, two types of attacks can be considered. Majority attacks and Selfish Mining attacks: both these attacks exploit the mining features in order to manipulate the network.

The Majority Attack, better known as the "51% attack", would occur if a miner or mining pool alone could manage 51% of the network's power and thus overturn the network itself. After reaching 51%, and therefore the majority, of the hash rate of the network a malicious user can make invalid some transaction, preventing its verification, reverse it in order to allow double-spending, split the network, or forking the main blockchain.

Through this attack the concept of decentralized network is lost, in fact it involves a manipulation of the transactions by a network majority. Moreover, depending on how overwhelming is the majority which the attacker has, although it can be burdensome, with this type of attack is possible to go back

in time to create alternative versions of the old blocks, obtaining modification of already existing blocks.

The probability that a miner has to launch a successful attack is a function of his normalized hash rate (x) and the number of confirmations (k) obtained from the network. Considering (y) the remaining hashing power, such that x + y = 1, the probability of success P (s) is:

$$P(s) = \begin{cases} 1, & if\ x > y \\ \left(\frac{x}{y}\right)^k, & if\ x < y \end{cases} \quad (2)$$

It is necessary that a miner publishes a long chain with valid PoW, so that the entire network switches to his forked version. In models based on a consensus and validation mechanism different from the PoW, which are not based on the heavy computing power, any attack attempt becomes more profitable; in fact, these mechanisms, in relation to 51% attack, are considered less robust than PoW [18].

Selfish Mining Attack is an attack strategy in blockchains based on a PoW consensus mechanism. This attack is led by a miner with a large hash rate, over 25% of the total hash rate, who manages to validate a block before the others, but without announcing it to the network. The miner exploits the acquired advantage trying immediately to hook up another valid block and so on, building its own hidden chain. A competition is thus achieved between the two chains, the private one of the selfish miner, who has enough power to keep his chain secret and the public one of honest miners.

The selfish miner, opportunistically, reveals his private chain once the public chain approaches the length of his chain. Having demonstrated more proof-of-work compared to other miners in the mining pool, selfish miner achieves thus greater rewards.

Although the entire blockchain technology turns out to be weak against this kind of attack, it's possible to evaluate the attack feasibility on medical blockchain-based models considered in this work.

The presence of mining pool can be used as a discriminating factor because selfish mining is a specific attack vector to PoW consensus. It is known that the purpose of mining pools is to centralize power in order to overcome the calculation challenge while this attack is born with the purpose to "waste" the computational resources of honest miners.
On the other hand, the attack became ineffective in all the blockchain based model that are built on a PBFT or round robin consensus mechanism because there are no expensive resources involved in the generation of a new block (i.e. no challenge like PoW). Finally, networks that don't use mechanism such as Mining pools and hardware or the ASIC family (dedicated hardware to undermine) don't have any incentive having a selfish miner inside and therefore it goes to self-preservation [48], [51].
As a further type of attacks, **cryptographic primitive attacks** are considered, in order to assess security of the protocols used in each considered approach. In order to obtain a realistic analysis, it is necessary to remember that there are two types of primitives in the blockchain-based models analysed: one or more primitives at the basis of the consensus mechanism and one or more primitives at the basis of the privacy preserving mechanism. Within the blockchain infrastructure, different primitives are exploited depending on the consensus mechanism used. For example, at the base of the PoW consensus mechanism there is a hash computation according to SHA-256. To date, attacks on this kind of primitive turn out to be ineffective. The situation is different for blockchains based on the PBFT mechanism, which do not require hashing power in the validation phase of the transaction, but a heavy exchange of messages between the different nodes is required, characterized by a cryptographic proof. To add new valid blocks to the underlying blockchain Message Authentication Codes (MACs) and RSA (Rivest–Shamir–Adleman) digital signature algorithm are mainly used. The cryptographic primitives that are at the basis of privacy-preserving protocols may be less robust than those used in the blockchain infrastructure: for example, in the blockchain-based protocols that aim at preserving the health data integrity

through a hash digest, the presence of a pseudo-random salt is often not guaranteed, which instead would make the hashing process more secure. Brute-force or dictionary attacks can therefore be applied with the aim those that can lead to guessing of the hash. **Brute Force attacks** attempt to solve the hash guessing problem by trying every possible combination of the input. Instead, **Dictionary attacks** only try a reduced set of inputs that are likely to work. Although they require considerable computations, it is possible to derive the computational effort required and therefore the vulnerability of the privacy preserving system depending on the digest algorithm used. As said before, the calculation of the hash of a file consists in the application of a hashing algorithm that from the file extracts a string that identifies it uniquely. This is not a reversible process; therefore, it cannot be traced back to the origin of the hash unless by infinite attempts. The hash is a one-way and collision-resistant mathematical process: this means that given an n-bit hash, that is a map from arbitrary length messages to n-bit hash digests, in order to find any message that hashes to that value it is required a work equivalent to about 2^n hash computation, where n is precisely the number of digest bits. Considering the birthday paradox in probability theory, according to which the probability of collision is largely greater than the collective imaginary, it should be stressed that this probabilistic model was used to reduce the complexity of finding a collision for a hash function, in a cryptographic attack called the birthday attack. Having said that, each hash algorithm can provide no more than n/2 bits of security against collision attack: therefore, finding any message which hashes to the same value should require a work equivalent to 2^(n/2) hash computations.

Among the best-known hash algorithms there are MD5, SHA-1, SHA-2, whose main features are summarized in Table III.

***Table III**: Comparison of the features of the best-known hash algorithms*

Hash Algorithm	*Bit-hash*	*Hash characters*	*Computational effort*	*Security against Collision Attack*
MD5	128	32	2^{128}	2^{64}
SHA-1	160	40	2^{160}	2^{80}
SHA-2	256	64	2^{256}	2^{128}

Among the models analysed, only one has specifications relating to the algorithm used: BlocHIE exploits MD5. The other solutions, although based on hash computations, do not report specifics about the algorithm used to compute the digest stored in blockchain. Brute force attacks and dictionary attacks against hash functions apply for all those protocols that store a hash digest on a public blockchain, like BlocHIE and MHMD. In order to preserve the data protection of healthcare data, in most of the cases, the considered solutions exploit permissioned distributed ledger techniques. Although they are permissioned, they are based on weak authentication systems.

As previously underlined, in these models, in order to gain access to the desired resource, the user enters the pair (ID, Password), where the ID is an identity statement, and the password is the proof that supports this statement. Authentication schemes of this type are defined as 1 Factor (1F) authentication. Their security may be compromised through attacks to the primitive, in case it is weak, or attacks to the key, like brute force attacks and dictionary attacks.

Specifically, 1F authentication systems may be subject to attacks such as:

- Re-use of a fixed password: an attacker can get hold of a user's password directly from the source. The password can be acquired directly from the server during the brief interval of verification, in which it is clear.

- Exhaustive password search: an attacker systematically or randomly tries to apply the unidirectional function to a string until it finds an encrypted password in the system
- Password guessing: consists in guessing the password by searching for it in decreasing order of probability.

In healthcare blockchain models built on a permissioned network type, an authentication procedure is required, in order to join the network.

Once the permissioned wall is broken by means of a possible fake profile, pretending to be a legitimate user or by means the kinds of attack previously described, it is possible to proceed with the guessing of the hash, as mentioned above in the session of cryptographic primitive attacks.

Scalability it is the ability of a system to adapt to an increased demand. The blockchain is an intrinsically scalable system: as the number of nodes increases, increases the stability of the system which becomes more secure. Connected to the scalability concept there is the efficiency issue. The blockchain requires that all data must be replicated on the mining nodes, this creates a huge redundancy, in a constantly growing network and, therefore a low efficiency in managing large amounts of data. The same blockchain maximum pre-established block size, greatly affects this issue. Therefore, transmitting and storing a large amount of data on blockchain rise efficiency concerns, connected with an increase in transactions' time and costs. Related to the concept of efficiency, the applicability of an **Overload attack** has been evaluated. It is possible to exploit, indeed, the high redundancy of the system and its correlated low efficiency in order to hypothesize an attack on the system.

In the blockchain-based models that preserve healthcare data with an on-chain storage, by assuming an increase in the amount of data, the aforementioned efficiency issue would be recorded: by increasing the amount of data, an increase in the flow inside the system would follow. The preconditions of an overcrowding of the system itself are so created. These, in the long term, will damage the regular functioning of the network.

Moreover, by evaluating the models based on PBFT mechanism it seems that they work well in their classical form with small consensus group sizes due to the cumbersome amount of communication that is required to reach the consensus. Therefore, a consequent increase and overload of the network would entail remarkable consequences in terms of efficiency and scalability.
The hypothesis of the **Padding Oracle Attack** was recorded for all the models that use data storage in the blockchain in an encrypted way. This type of attack uses cryptographic message fill or padding validation to decrypt ciphertext. Usually in cryptography, plaintext messages with variable length often need to be expanded in order to be compatible with the underlying cryptographic primitive. This type of attack is associated with symmetric cryptographic systems that operate in Cipher Block Chaining (CBC) mode: an oracle, generally a server, leaks data about whether the filling/padding of an encrypted message is correct or not. The messages can therefore be decrypted by the attackers through the use of these data. In detail, in CBC encryption a message must have multiple length of the block length. Therefore, a padding scheme is used to obtain the requirement. Let 'L' be the length of the block. Let 'b' be the number of bytes to be added to the message. It turns out $1 \leq b \leq L$. Strategy used to favour padding is the following: the string containing b (a byte) is added to the message exactly b times. When decrypting, the padding is checked and removed, before returning the plaintext. If the padding is not correct, an error is generated and returned. A Padding Oracle attack allows to manipulate a cipher, to send the manipulated version and to observe the obtained response from the server and finally, to repeat the procedure if it fails to calculate the word. This decryption system allows to decipher any encrypted code and gives an error if the code is incorrect. Using this information, therefore, an attacker can completely recover the plaintext. In particular, the process that allows the decryption of the data takes place without knowing the encryption key, but through the work of the oracle. Although even the padding modes for asymmetric algorithms, like OAEP, can also be vulnerable to Oracle Padding attacks, the probability that this happens

is limited by the difficulty of the algorithm itself. Systems that insert the data on the blockchain, after a symmetric encryption, are more exposed [52]. Finally, considering the conservation mechanism of the health data on an off-chain database storage, only some of the previous attacks can be applied. Since in this type of paradigm the storage of health data is off-chain, the classic attacks on a database have to be taken into consideration. Furthermore, being the blockchain used to map the location of the data itself, verifying their integrity, DDoS attacks on the network would lead to a loss of the document index. This loss would lead to a dissociation between the health data and their owner; the reference to data inserted in the blockchain may be lost, thus creating an outage. Attacks of this type could obscure some of the blockchain nodes and make them unreachable for a certain period of time: going to undermine the mechanism for reconstructing the data. Indeed, keeping in the blockchain the data pointer, with this attack any reference to the location of the data is momentarily lost. The health data itself does not undergo alterations, remains saved in a database, without however having the instructions that guarantee its reconstruction, in the time frame of attack execution.

For the analysed models that exploit an off-chain database storage it is also necessary to point out that the use of obsolete database or of a database with an old software can be a high-risk factor by itself. The possible malware infiltration associated with use of an obsolete infrastructure, could potentially produce alteration of the healthcare data itself by threatening their privacy, in case of unauthorized reading or release, their integrity, in the event of unauthorized changes, and their availability. A malware infiltration can be disastrous; consequences include data theft, extortion or the crippling of network systems. This last example of attack also affects the models that have been taken in account for the abstraction of the protocol which inserts the health data into the blockchain by means of a hash computation, with a combination of on- and off-chain storage.

4.2 Technical Analysis of Security Vulnerabilities

The previously presented taxonomy of attacks highlights the main features of every single attack and the vulnerabilities to which each one aims. An analysis of the security of the medical blockchain-based models taken into account has been carried out, establishing for each model the degree of applicability of each single attack.

This structural analysis starts from the technical components of each model, previously schematized in the table II, and the characteristics of each attack. As output, it provides a detailed picture of the risk related to the use of each individual GDPR-compliant model.

Three degrees of applicability related to each attack have been defined: not applicable, partially applicable and applicable, referring to the technological components of the solutions. Depending on the degree of applicability of the attack, a low (green), medium (orange) and high (red) vulnerability has finally been obtained for each single solution. Table IV shows, in relation to each type of attack, the level of vulnerability related to the use of each healthcare blockchain-based model considered in the analysis.

Table IV: *Obtained results of technical analysis of security vulnerabilities*

Healthcare blockchain-based model	*Network-based Attack*			*Mining-based Attack*		*Security Mechanism-based Attack*			
	Replay Attack	*Sybil Attack*	*DDos Attack*	*Selfish Mining Attack*	*Majority Attack*	*Padding Oracle Attack*	*Dictionay Attack*	*Brute Force Attack*	*Overload Attack*
BlocHIE	🔴	🟢	🟢	🔴	🟠	🟢	🔴	🔴	🟢
BcPIIMS	🟠	🔴	🔴	🟢	🔴	🟢	🟠	🟠	🟢
BEIM	🟢	🔴	🔴	🟢	🔴	🟢	🟢	🟢	🟢
MHMD	🔴	🟢	🟢	🔴	🟠	🟢	🔴	🔴	🟢
MediBchain	🔴	🟢	🟢	🔴	🟠	🔴	🟠	🟠	🔴
MIStore	🟢	🔴	🔴	🟢	🔴	🟠	🟠	🟠	🔴
ABC	🔴	🟢	🟢	🔴	🟠	🟠	🟠	🟠	🔴
MediChain	🔵	🔵	🔴	🔵	🔵	🔵	🔵	🔵	🔵

In the figure 4.2.1 are considered network-based attacks: specifically Replay attack, Sybil attack and DDos attack.
For the *Replay attack* the risk of exposure is correlated to the occurrence of hard fork within the network. It is estimated that in BloCHIE, BcPIIMS, MHM, MedBchain and ABCD models there is a high risk of exposure to this attack due to the use of PoW as a consensus mechanism. In BcPIIMS, since it is used round robin as a mechanism for the validation of a new block, the risk is medium. A low risk is assessed in BEIM and MIStore as these are based on a PBFT mechanism.In case of MediChain, the attack does not apply. For the *Sybil attack* it is possible to notice that BloCHIE, MHMD, MedBchain and ABC that use PoW as a consensus mechanism, have an intrinsic resistance to this attack. In fact, these models have a low exposure risk. In contrast, BcPIIMS, BEIM and MIStore, using consensus mechanisms different from PoW, have a high exposure risk. In case of MediChain, the attack does not apply.
For the *DDos attack* a low risk of exposure is assigned for BloCHIE, MHMD, MediBchain, ABC in which, this type of attack is really expensive, considering the huge computational power required by PoW mechanism. BcPIIMS, BEIM, MIStore, have a high risk of exposure, since in these models the size of the network is known, and therefore it is possible to trace the number of Sibyl nodes needed to carry out the attack and obstruct the verification process. In case of MediChain, the risk of exposure to this attack is high, since it would damage the data reconstruction process.

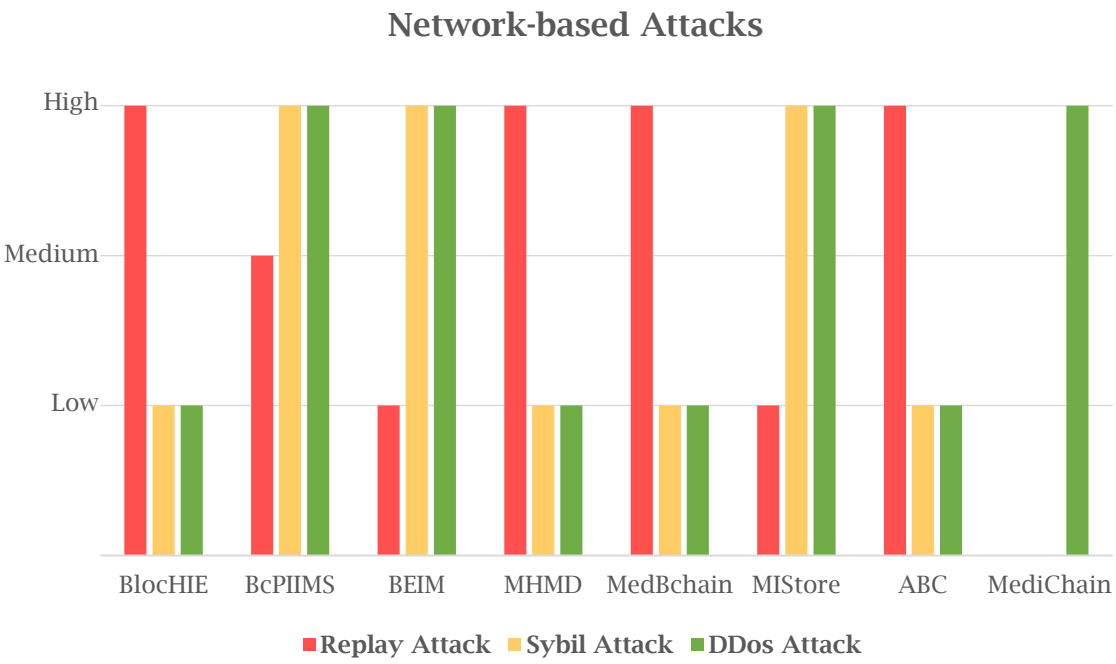

***Figure 4.2.1**: Risk of exposure to network-based attacks*

In the figure 4.2.2 the mining-based attacks are considered, in the specific Selfish Mining attack and Majority attack. For the *Selfish Mining attack,* a low risk of exposure has been assigned to the BcPIIMS, BEIM and MIStore models, since in them there are no expensive resources involved in the new block generation (i.e. no challenge like PoW). A high exposure risk is assigned to BlocHIE, MHMD, MedBchain, ABCD which use PoW as a consensus mechanism and therefore have incentive having a selfish miner inside. In case of MediChain the attack does not apply. The reasoning for the *Majority attack* is, instead, different. To BcPIIMS, BEIM and MIStore models is assigned a high risk of exposure because, using mechanisms different from PoW, they are less robust with respect to this type of attack. Any attempt of this attack becomes less profitable, but still possible, in model as BlocHIE, MHMD, MedBchain, ABCD that are based on PoW. Indeed, these models are assigned a medium risk of exposure. In case of MediChain, the attack does not apply.

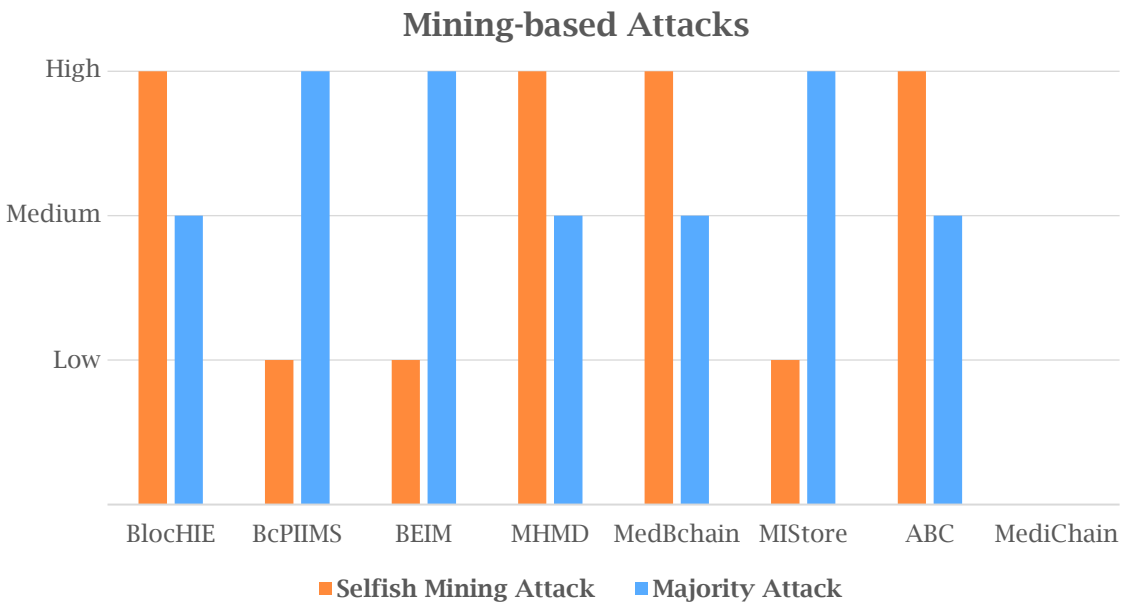

Figure 4.2.2: *Risk of exposure to mining-based attacks*

In the figure 4.2.3 are considered the Security mechanism-based attack, specifically Padding Oracle attack, Dictionary attack, Brute Force attack and Overload Attack. BloCHIE, BcPIIMS, BEIM and MHMD, appear to have a low exposure risk to the *Padding Oracle attack* and to the *Overload Attack*. This is intrinsically due to the mechanism of data storage. In fact, these models, providing a combination of off-chain and on-chain storage through a hash computation, are safe with respect to these types of attacks. In case of MediChain the attack does not apply. As far as the *Padding Oracle attack* is concerned, to MediBchain has been assigned a high risk, as it preserves the health data on-chain after a symmetric encryption. To MIStore, which uses a public encryption key and to ABC, that does not specify which kind of encryption function it uses, have been assigned a medium risk. In case of MediChain the attack does not apply. Instead, for the *Overload attack*, since MediBchain MIStore and ABC are characterized by an on-chain storage, the risk of getting an overcrowd of the system by large amounts of data is high. In case of MediChain the attack does not apply. For the *Dictionary attack* and the *Brute Force attack*, the same considerations are applied. BloCHIE and

MHMD which propose a type of permissionless blockchain, have a high risk of exposure to the aforementioned attacks; in a public environment, the guessing of hash of the health data by means of dictionary attack and brute force attack is greatly favoured by the absence of a pseudo-random salt. The other models choose a blockchain type different from the permissionless one. In order to join the network a 1F authentication system is required, and since it is considered weak, for these models have been assigned a medium risk of exposure. BEIM, with respect to the other solutions provide an additional level of security: not only the data are kept secure in a permissioned environment with the on-chain data hash storage, but the digest hash is then, asymmetrically encrypted. Being the computational effort required too high, the risk of exposure is low. In case of MediChain the attack does not apply.

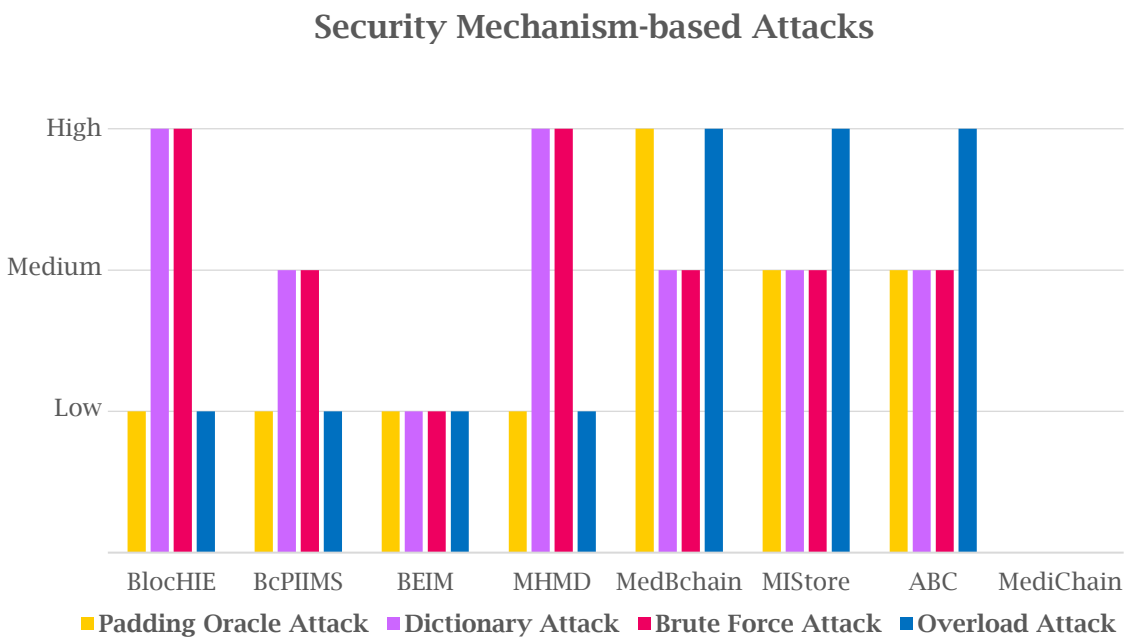

***Figure 4.2.3**: Risk of exposure to security mechanism-based attacks*

4.3 Discussion

The Table IV shows that none of the considered GDPR-compliant medical blockchain-based models is totally immune to the whole set of considered attacks. It is clear that the consensus mechanism is a discriminating factor in the assessment of the applicability of the considered attacks. The quantitative analysis of the results obtained shows that the most promising solutions for the treatment of health data are those that, in order to preserve the privacy of users in compliance with the GDPR, combine a permissioned blockchain with a consensus mechanism different from PoW. Compared to the public counterpart, a permissioned blockchain infrastructure offers better performance, in that it uses consensus mechanisms more suited to a permissioned environment and less heavy computationally, such as the PoW. However, the fact that in these solutions other consensus mechanism than the PoW are used exposes them to the risk of Sibyl attacks, which as previously pointed out found a solution in the complex computational power required by PoW. The susceptibility of the PBFT mechanism to Sybil attacks could be solved by increasing the number of nodes participating in the network, as a low number of nodes within the network facilitates this type of attack. The increase in nodes, reducing the risk of Sybil attacks, would however lead to problems of scalability, since models based on the PBFT mechanism work well in their classic form, that is, with a small consensus group.

It would therefore be advisable to find a trade-off: adding nodes to the network ensuring a reduction in the applicability of Sybil attacks, but at the same time limiting this addition to strictly necessary numbers in order to reduce scalability problems associated with an increase in network size. It is also worth observing that among the solutions that use PBFT in a permissioned blockchain, those that preserve the data by combining an on-chain/off-chain storage by means of a hash

computation are unencumbered from the risk of Overload and Padding Oracle attacks. It is possible to conclude that among the solutions analysed those permissioned and based on consensus mechanisms different from PoW, that also belong to the paradigm that conserves health data by combining on-chain and off-chain storage through hash computation, turn out to be the most secure ones.

This is because health data is stored in a network that requires user authentication and that uses a consensus mechanism suitable for a permissioned environment. In addition, the feasibility of attacks that aim at hash guessing appears to depend heavily on the permissioned blockchain framework. In fact, attacks of this type are only possible after breaking users' authentication systems, which although weak, being 1F systems, exist and cannot be neglected.

Conclusions

This work stems from the idea of providing abstraction models and a subsequently security assessment of privacy preserving protocol that using the blockchain technology provide a GDPR-compliant, innovative, secure and patient-centric tools for the healthcare data sharing.

After analysing the requirements of the GDPR, imposed on the treatment and processing of healthcare data, an overview of the blockchain technology was provided, analysing its general characteristics, architecture and operating mechanisms. Due to its immutability by default and its public and distributed nature this technology turns out to be intrinsically in contradiction with the GDPR. In this study, several projects that try to overcome this lack have been analysed and grouped in three GDPR-compliant paradigms. These three groups are characterized by different storage mechanisms and supply many healthcare challenges, such as, data integrity, data interoperability, data protection and data sharing. Starting from the core technology components of each model and considering the current healthcare scenario and the apparent unstoppable growth, qualitative and quantitative, of the attacks, an analysis of the vulnerabilities was then carried out, in order to evaluate the security of the mechanisms with which the patient data protection is guaranteed.

Although recently blockchain-based medical systems are a hot topic, according to the latest studies there is no security analysis related to their use. This study firstly offers a taxonomy of the attack and secondly hypothesises different types of attack that aim at the structure of the network, the type of mining, the process of filling the chain and the security

mechanism. Even though all the analysed models have the objective of preserving the privacy of the patient, the permissioned blockchains, based on a consensus mechanism different from PoW, that requires a process of authentication of the involved parts, even if weak, are seemed the most secure.

To conclude, this study could represent a starting point for further research. In order to demonstrate an evaluation on the security level of privacy preserving protocol based on blockchain applications in the healthcare field.

Bibliography

[1] P. Voigt and A. Von dem Bussche, “The EU General Data Protection Regulation (GDPR)”, *A Practical Guide, 1st Ed., Cham: Springer International Publishing,* 2017.

[2] T. Lyons, L. Courcelas and K. Timsit, “Blockchain and the GDPR-thematic report”, European Union Blockchain Observatory and Forum, 16 october 2018.

[3] Iubenda, “GDPR Guide”, [Online]. Available: https://www.iubenda.com/en/help/5428-gdpr-guide.

[4] Information Commissioner's Office, Guide to the General Data Protection Regulation (GDPR), 22 May 2019 .

[5] H. Li, L. Yu and W. He, “The Impact of GDPR on Global Technology Development”, *JOURNAL OF GLOBAL INFORMATION TECHNOLOGY MANAGEMENT,* vol. 22, no. 1, pp. 1-6, 24 Jan 2019.

[6] M. Pilkington, “11 Blockchain technology: principles and applications”, in *Research handbook on digital transformations 225*, 2016.

[7] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System” *Bitcoin.org: 9,* DOI 10.1007/s10838-008-9062-0, 2008.

[8] A. Dhar Dwivedi, G. Srivastava, S. Dhar and R. Singh, “A Decentralized Privacy-Preserving Healthcare Blockchain for IoT,” *Sensors,* vol. 9, no. 326, pp. 1-17, 15 January 2019.

[9] A. Al Omar, M. Shahriar Rahman, A. Basu and S. Kiyomoto, “MediBchain: A Blockchain Based Privacy Preserving Platform for Healthcare Data”, *Springer International Publishing ,* no. DOI: 10.1007/978-3-319-72395-2_49, pp. 534-543, 2017.

[10] K. Sultan, U. Ruhi and R. Lakhani, “CONCEPTUALIZING BLOCKCHAINS: CHARACTERISTICS & APPLICATIONS”, in *11th IADIS International Conference Information Systems*, 2018.

[11] M. Raikwar, D. Gligoroski and K. Kralevska, “SoK of Used Cryptography in Blockchain”, 2019.

[12] Z. Zheng, S. Xie, H. Dai, X. Chen and H. Wang, “An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends”, in *IEEE 6th International Congress on Big Data*, June 2017.

[13] D. Mohanty, “Deploying Smart Contracts”, *Ethereum for Architects and Developers,* pp. 105-138, Springer, 2018.

[14] G. Wood, “ETHEREUM: A SECURE DECENTRALISED GENERALISED TRANSACTION LEDGER”, 2014.

[15] A. Ali and M. Mazhar Afza, “Confidentiality in Blockchain”, *International Journal of Engineering Science Invention (IJESI),* vol. 7, pp. 50-52, 2018.

[16] G. G. Daghera, J. Mohler, M. Milojkovicc and P. Babu Marellaa, “Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology”, *Sustainable Cities and Society,* vol. 39, p. 283–297, 2018.

[17] Fede, “Consensus in the Blockchain Ecosystem-What is Consensus, and how to achieve it in a Blockchain ecosystem”, The Blockchain Lion, 7 April 2019. [Online]. Available: https://blockchainlion.com/consensus-blockchain/.

[18] M. Saad, J. Spaulding, L. Njilla, C. Kamhoua, S. Shetty, D. Nyang and A. Mohaisen, “Exploring the Attack Surface of Blockchain: A Systematic Overview”, pp. 1-30, ArXiv, 6 Apr 2019.

[19] M. Vukolic, “The quest for scalable blockchain fabric: Proof-of-work vs. bft replication”, *Spinger, International Workshop on Open Problems in Network Security,* pp. 112-125, 2015.

[20] N. Szabo, “Smart Contracts”, 1994. [Online]. Available: http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html.

[21] H. Kenneth, “Ethereum account”, Medium: Coinmonks, 7 May 2018. [Online]. Available: https://medium.com/coinmonks/ethereum-account-212feb9c4154.

[22] Ethereum, “Ethash”, 2016. [Online]. Available: https://github.com/ethereum/.

[23] X. Yue, H. Wang, D. Jin, M. Li and W. Jiang, “Healthcare data gateways: Found healthcare intelligence on blockchain with novel privacy risk control”, *Journal of Medical Systems - Springer,* vol. 40(10) :218, 2016.

[24] Q. Xia, E. B. Sifah, K. O. Asamoah and e. al., “MeDShare: Trustless medical data sharing among cloud service providers via blockchain”, *IEEE Access,* vol. 5, pp. 14757-14767, 2017.

[25] A. Ekblaw, A. Azaria, J. D. Halamka and e. al., “A Case Study for Blockchain in Healthcare: “MedRec” prototype for electronic health records and medical research data”, *Proceedings of IEEE Open and Big Data Conference,* vol. 13, pp. 1-13, 2016.

[26] G. Drosatos and E. Kaldoudi, “Blockchain Applications in the Biomedical Domain: A Scoping Review”, *Computational and Structural Biotechnology Journal,* no. 17, p. 229–240, 2019.

[27] T.-T. Kuo, H.-E. K. Kim and L. Ohno-M, “Blockchain distributed ledger technologies for biomedical and health care applications”, *Journal of the American Medical Informatics Association,* vol. 24, no. 6, p. 1211–1220, 2017.

[28] I. Drew, “Moving Toward a Blockchain-based Method for the Secure Storage of Patient Records”, Agoust 2016.

[29] E. Piscini, D. Dalton and L. Kehoe, “Blockchain & Cyber Security Point of View”, Deloitte.

[30] IBM Security, “Blockchain and GDPR-How blockchain could address five areas associated with GDPR compliance”, United States of America, March 2018.

[31] CNIL-Commission Nationale Informatique&Libertès, “BLOCKCHAIN-Solutions for a responsible use of the blockchain in the context of personal data”.

[32] A. Bayle, M. Koscina, D. Manset and O. Perez-Kempner, “When Blockchain Meets the Right to be Forgotten: Technology Versus Law in the Healthcare Industry”, in *IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 2018.

[33] N. Al-Zaben, M. M. H. Onik, J. Yang, N.-Y. Lee and C.-S. Kim, “General Data Protection Regulation Complied Blockchain Architecture for Personally Identifiable Information Management”, pp. 77-82, 2018.

[34] E. Morley-Fletcher, “MHMD: My Health, My Data”, Rome, 2017.

[35] BCD-Blockchain Certified Data, “BCDiploma-WhitePaper v2.2”, April 20th, 2018.

[36] “Protecting Data through Encryption in a Public Blockchain”, [Online]. Available: https://www.platform6.io/2019/03/20/protecting-data-through-encryption-in-a-public-blockchain/.

[37] Platform, “Protecting Data through Encryption in a Public Blockchain”, [Online]. Available: https://www.platform6.io/2019/03/20/protecting-data-through-encryption-in-a-public-blockchain/.

[38] Hackernoon, “Understanding Blockchain Privacy — Anonymity, Encryption and Decentralization”, 23rd June 2018. [Online]. Available: https://hackernoon.com/understanding-blockchain-privacy-anonymity-encryption-and-decentralization-30ed4fddb808.

[39] Medichain, “Whitepaper”, 2017. [Online]. Available: https : / / icorating . com / upload /whitepaper/vU8xMs6F0MYeJPg6KcIxJpkItW0Oln2jyiGdUXbx.pdf.

[40] A. Azaria, A. Ekblaw, T. Vieira and A. Lippman, “MedRec: Using Blockchain for Medical Data Access and Permission Management”, in *2nd International Conference on Open and Big Data*, 2016.

[41] S. Jiang, J. Cao, H. Wu, Y. Yang, M. Ma and J. He, “BlocHIE: a BLOCkchain-based platform for Healthcare Information Exchange”, in *IEEE International Conference on Smart Computing*, 2018.

[42] N. Al-Zaben, M. M. Hassan Onik, J. Yang, N.-Y. Lee and C.-S. Kim, “General Data Protection Regulation Complied Blockchain Architecture for Personally Identifiable Information Management”, *IEEE,* pp. 77-82, 2018.

[43] T. Hyla and J. Peja´, “eHealth Integrity Model Based on Permissioned Blockchain”, *Future internet,* vol. 11, no. 76, pp. 1-14, 2019.

[44] L. Z. W. Sun, “MIStore: a Blockchain-Based Medical Insurance Storage System”, *Journal of Medical Systems,* vol. 42, no. 149, pp. 1-17, 2018.

[45] W. Liu, S. Zhu, T. Mundie and U. Krieger, “Advanced Block-Chain Architecture for e-Health Systems”, in *19th International Conference on E-health Networking, Application & Services (HealthCom): 2nd IEEE ETPHA* , 2017.

[46] A.-Z. Mishall, Z. Zhang and J. Zhang, “Efficient and Secure ECDSA Algorithm and its Applications: A Survey”, *arXiv,* pp. 1-31, 27 Feb 2019.

[47] B. Garret, “Vormetric data threat report”, RESEARCH, Thales Group, Tech. Rep., 2016.

[48] M. Saad, L. Njilla, C. Kamhoua and A. Mohaisen, “Countering Selfish Mining in Blockchains”, *arXiv,* pp. 1-5, 2018.

[49] R. Garavaglia, “Le fork sulla blockchain: un equilibrio fra opportunità e rischi di una governance decentralizzata”, *Ict Security Magazine,* 2019.

[50] AGID CERT-PA, “CRIPTOVALUTE- cosa sono e come usarle al meglio”.

[51] S. Lee and S. Kim, “Pooled Mining Makes Selfish Mining Tricky”, 2018.

[52] J. Manger, “A Chosen Ciphertext Attack on RSA Optimal Asymmetric Encryption Padding (OAEP) as Standardized in PKCS #1 v2.0”.

[53] Klaytn Docs, “Consensus Mechanism”, [Online]. Available: https://docs.klaytn.com/klaytn/design/consensus-mechanism.