Privacy Preservation in a Blockchain-Based Healthcare System

IVOR ZAGORAC[1] , **CHHAGAN LAL**[1] , **MAURO CONTI**[1]

[1]TU Delft

Abstract

The continuous generation of a large volume of health data from different sources has led to healthcare being a data-intensive domain. To achieve innovative advances in medical treatment procedures and to provide personalized healthcare services to the patients this data needs to be shared among different medical facilities. However, because this data is highly sensitive and personal, several challenges can be faced. Since blockchain technology has features such as transparency, immutability, confidentiality, and auditability, research is being performed to check whether it can be integrated in the healthcare system and thus in the medical data sharing. Nonetheless, privacy is an important aspect of healthcare systems that blockchain technology needs improvement in. This research, first defines the security and privacy requirements for healthcare systems. Then, we look deeper into the privacy requirements that have to be met in blockchain systems and the threats that can arise when systems do not meet them. Next, we present several privacy protection techniques that can be used in a blockchain-based healthcare system and present a design which is a combination of techniques that fulfill the privacy requirements. Lastly, this design is evaluated to see how each component of the design fulfills the requirements necessary.

1 Introduction

The continuous generation of a large volume of health data from different sources has led to healthcare being a data-intensive domain [19]. To achieve innovative advances in medical treatment procedures and to provide personalized healthcare services to the patients this data needs to be shared among different medical facilities [40]. Since blockchain technology has features such as transparency, immutability, confidentiality, and auditability, research is being performed to check whether it can be integrated in the healthcare system [17]. This research will focus on investigating data collection, data sharing, and data processing functionalities in the healthcare domain and their various security and privacy challenges. Furthermore, it will review a possible solution to these challenges, namely blockchain. Recently, accountability for blockchain-based healthcare systems has been researched by Al Omar et al. [1] who proposes 'Medibchain' a privacy preserving platform for healthcare data.

The main goal of this research is to investigate the existing solutions that support confidentiality and auditing in blockchain-based healtcare systems. As a result of this research, a thorough understanding of the confidentiality systems for blockchain and their key benefits and limitations regarding security and privacy parameters is expected. The research question I am aiming to answer is:

How can data confidentiality be achieved in a blockchain-based healthcare system?

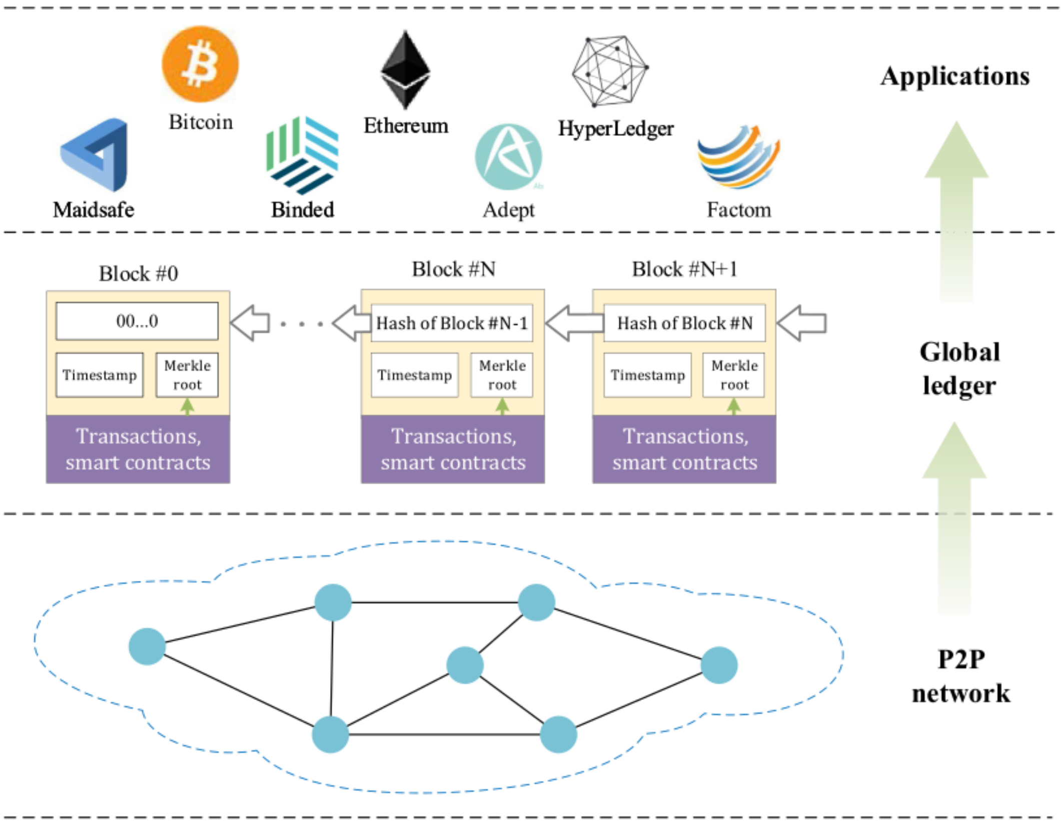
To answer the research question, first the existing state-of-art sytems for confidentiality have to be examined. Secondly, the systems have to be compared using different security and privacy parameters. This leads to the following sub-questions:

1. What are the existing systems that allow for data confidentiality in blockchain-based medical data sharing (BMDS) and how do they work?
2. What are the benefits of the data confidentiality systems in BMDS when comparing them using security and privacy parameters?
3. What are the limitations of the data confidentiality systems in BMDS when comparing them using security and privacy parameters?

The contribution of this paper can be summed up as follows:

- First, we present the privacy requirements that have to be met in a blockchain-based system and discuss the threats that arise when the system lacks in meeting them. Furthermore, we explain how the HIPAA and GDPR regulations can assist in ensuring privacy preservation in these systems.
- Second, we present the benefits and limitations of several techniques that can be used to protect privacy in a blockchain-based healthcare system. Additionally, we present a design that combines these techniques to fulfill the requirements stated earlier.

Figure 1: Overview of the Blockchain architecture [20]



- Lastly, we evaluate the design based on requirements for privacy in healthcare systems found in other research.

The rest of this paper is organized as follows: Section 2 gives background information on important components of the blockchain-based healthcare system. Section 3 gives an overview of multiple researches that present blockchain-based healthcare systems and presents the benefits and limitations of each of the systems. In Section 4, the privacy requirements of a blockchain-based system are discussed. Furthermore, the threats that can arise when a system does not meet these requirements are presented. Next, in Section 5, techniques that can assist in privacy preservation in a blockchain-based system are explained. Section 5.3 proposes the design of a privacy-preservation system whcih combines multiple techniques used for privacy protection. The evaluation of the design is done in Section 6. Finally, the paper is concluded in Section 8.

2 Background

This section will give background information on the important components of a blockchain-based healthcare system. Firstly, a simple explanation of blockchain and smart contracts will be presented. Second, the important security and privacy requirements of healthcare systems will be presented. Lastly, the blockchain framework Hyperledger Fabric is explained because this is a commonly used framework for blockchain-based healthcare systems.

2.1 Blockchain and Smart Contracts

Blockchain is a distributed ledger. The blocks in the chain are chronologically ordered and can not be updated once they are committed to the chain. This is ensured by linking the blocks through hash functions. Each block contains the hash of the previous block which means that if someone was to tamper with one block, they would have to change all the following blocks. Since this is a very time-consuming job, the blockchain becomes immutable and integrity of the blocks is accomplished. Each block contains multiple transactions where each transaction can for example be a transfer of money or in the case of a blockchain-based healthcare system, a change in the permission rules of a patient's data. Smart contracts (SC) were first introduced by Nick Szabo in 1994 who defined SC as "a computerized transaction protocol that executes the terms of a contract" [35]. In blockchain context, SC are scripts stored on the blockchain. A smart contract is triggered when a transaction is invoking the address of the smart contract on the blockchain. The SC executes independently and automatically on every node in the work. It uses the code specified in the SC and the data given in the transaction to get to the new state of blockchain. The SC allow the users of the blockchain to run code on the data without the need of trusted parties because the SC is safely stored on the blockchain as well.

2.2 Healthcare System Requirements

[28] is a research on the security and privacy issues in modern healthcare systems. In this research, the authors found several goals that have to be set for healthcare systems regarding security and privacy. The security goals are:

- *Authentication*: This is a fundamental component for securing healthcare systems [28]. Instead of having weak password authentication schemes like most of the current networked medical devices, the system has to cover aspects such as 'environment setup', 'single or multi-factor authentication' and 'emergency scenarios'.
- *Confidentiality*: Only authorized personnel or entities should have access to components of the system such as device information, system configuration and healthcare data.
- *Integrity*: Unauthorized users' devices or applications should not be able to modify the healthcare system or the data in the system. Without a mechanism that checks for integrity, malicious attackers could possible alter the data in a healthcare system leading to firmware failures [25].
- *Non-repudiation*: Healthcare systems should have some kind kind of logging functionality that makes sure that attackers can not cover their traces. Furthermore, it should not be possible to tamper with these logs or delete them.
- *Availability*: The system should always be available for authorized users, both in normal and emergency situations.

And the privacy goals are:

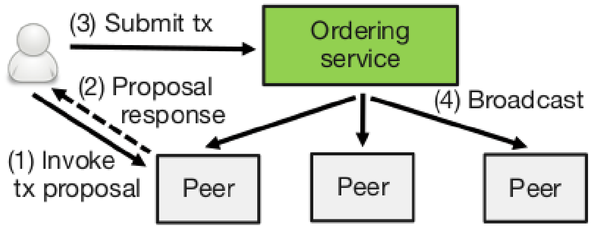
- *Device Anonymity*: This property means that the identity of the medical device is unknown to the system. This is important because it means that attackers or unauthorized entities should not be able to determine properties of the device such as IP or MAC address.
- *Data Anonymity*: Like it suggests, data anonymity means that unauthorized users can not identify another user from their data. Basically, patients and doctors in a healthcare system should use pseudonyms instead of their real identities. Furthermore, this requirement makes sure that sensitive data is not exposed to the outside world.
- *Communication Anonymity*: Communication between user and the system should happen anonymously. This also means that attackers should not be able to detect when a user is communicating with the system.
- *Unlinkability*: Attackers who can scan the data transactions between the sender and the receiver should not be able to find any relationship between data and sender.

These goals, or also referred to as requirements, will be used to evaluate the related work in Section 3. Furthermore, the privacy requirements will be part of the evaluation in Section 6.

2.3 Hyperledger Fabric

Hyperledger Fabric [2] is a permissioned blockchain platform and is commonly used as a component of a blockchain-based healthcare system. The nodes that have access to the ledger are called peers and each peer is part of an organization. When a user wants to add a transaction, this starts a two-phase process. First, the user has to approach one or more peers with a transaction proposal and ask them to execute and endorse it. The peers execute a SC, this is called the chaincode in Hyperledger Fabric, to check whether they will endorse the transaction. They also run the chaincode to see how the transaction would change the state of the ledger. Second, once sufficiently many endorsements are obtained, the user sends the transaction with the endorsements to an ordering service. This ordering service keeps track of the order of transactions coming in and adds the transactions to the ledger. Figure 2 gives an overview of the transaction flow in Hyperledger Fabric.

Figure 2: Overview of the transaction flow in Hyperledger Fabric [10]



3 Related Work

In recent years, various papers have proposed designs or frameworks for a blockchain-based healthcare system. This section will provide a brief description of each of these proposals and discuss their limitations. Table 1 also provides an overview of the characteristics of each of the frameworks based on the requirements set in 2.2.

[4] presents a blockchain implementation that addresses the major issues of the current healthcare systems, namely fragmented, slow access to medical data; system interoperability; patient agency; improved data quality and quantity for medical research. The system allows the participants to be fully informed on their medical history and any modifications to it by providing a log which is not only comprehensive, but also accessible and credible. However, the system does have some limitations. Firstly, it relies on the local system admin of individual databases for its security. Furthermore, it does not prevent an attacker from applying data forensics or frequency analysis on the transactions in the blockchain. In other words, even though a person's name is private, someone could analyze the amount of interactions between a person and a certain provider.

[15] proposes Ancile, a blockchain system which uses Ethereum tools to be both cost and storage effective for blockchain technology. The system is designed in such a way

Table 1: Overview of the related work

Papers	Authentication	Confidentiality	Integrity	Availability	Device Anonymity	Data Anonymity	Communication Anonymity	Unlinkability
[4]	●	●	●	◐	●	◐	◐	○
[15]	●	●	●	●	●	◐	◐	◐
[24]	●	●	●	●	●	●	◐	◐
[39]	●	●	●	●	●	●	●	○
[16]	●	●	●	●	●	◐	◐	○
[1]	●	●	●	●	●	◐	◐	○
[36]	●	●	●	●	◐	○	○	○

that it gives ownership and final control of EHRs to the patient. Furthermore, it controls and tracks who can access and use health documents, grants secure transfer of these documents and is HIPAA compliant. When looking specifically at privacy preservation, Ancile makes identifying a specific patient difficult by only using their Ethereum addresses. Moreover, since the system uses multiple smart contracts to separate information, there is a heightened level of data obfuscation. Nevertheless, the system does lack effort when it comes to transaction encryption. The authors mention the possibility of using differential privacy [18] to overcome the problem of malicious blockchain analysis.

In [24], the authors display the design of a blockchain-based privacy-preserving data sharing system for EMRs called BPDS. They solve the potential security risks of data centralized storage by storing the EMRs in the cloud and the indexes of the records in a tamper-proof consortium blockchain. There are four factors that provide strong privacy preservation in the system, namely anonymity, cloud storage, content extraction signature and improved DPoS. Each participant has multiple public keys and uses them for different transactions which makes the transactions anonymous. The original EMRs are encrypted and stored in the cloud which solves the risk of original medical data leakage. Content extration signature lets patients remove any sensitive portions in the original data such that there is no data privacy leakage. Lastly, the system uses an improved Delegated Proof of Stake consensus algorithm which guarantees reliability of data sharing.

[39] proposes a blockchain-based privacy preserving scheme for health data called Healthchain. In Healthchain, the collected health data of IoT devices can be uploaded and publisched as transactions. Furthermore, users of the system can read doctors' diagnoses of the uploaded data. The system uses two blockchains called Userchain and Docchain for users' health data and doctors' diagnoses respectively. To store this data, the system uses interplanetary file system (IPFS). IPFS is a distributed file system that stores data with high integrity and resiliency. To create a privacy-preserving scheme, the authors set the design goal that each user's health data can only be obtained by them and by their authorized professional healthcare staff. The paper mentions the use of encryption on both Userchain and Docchain to securely and privately store the data but the system lacks a way of making sure that privacy of patients is preserved in the transactions on the chain. Like mentioned before, this means that an attacker could use blockchain analysis to get sensitive information of the users.

MedChain, which is proposed in [16], is a blockchain based framework for EMRs that aims at providing interoperable, secure, and efficient access to health data whilst maintaining the patients' privacy. The system uses timed-based smart contracts to govern the transactions on the blockchain. It is designed to be compatible with existing EMRs databases and to improve current management systems. To provide privacy preservation, MedChain uses encryption to store the data privately.

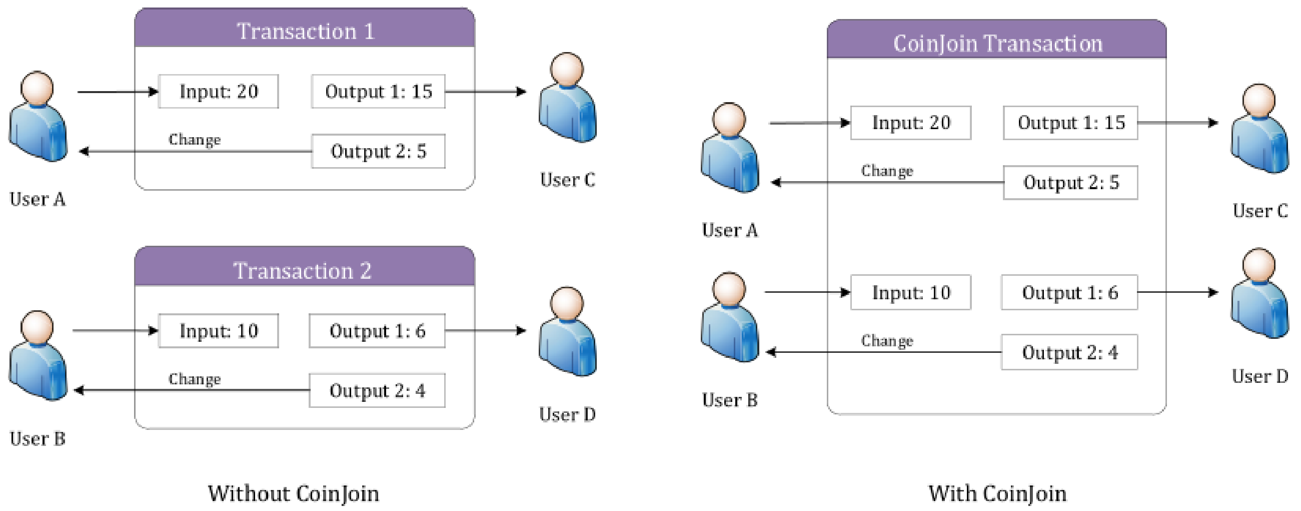
[1] explains the design of MediBchain, a platform that returns the control of the patients' private data to themselves. The authors state that present healthcare systems lack in pseudonimity but that MediBchain gives the pseudonimity of patients. The main goal of the platform is to retain accountability, integrity, pseudonimity, security and privacy as opposed to current EHR systems where these features are being lost. In order to achieve privacy, the framework has a 'Registration Unit' which handles logging into the system in a private manner. Furthermore, the encryption of the health data provides privacy too. The technical detail, however, lacks for this framework. The paper is a very quick overview of what a blockchain-based healthcare system could look like but the components are not described in detail.

[36] proposes PACEX, a Patient-Centric EMR Exchange model for healthcare systems using blockchain. PACEX allows the users to have complete control of their EMRs by only enabling transfer of the records according to the patient's consent. The system introduces an easy-to-use application that provides the ability to manage multiple accounts spread across different hospitals. The access control is performed completely using Blockchain technology. The qualitative analysis of the system shows that it has the required features: authentication, integrity, access control and traceability. Even though the system fulfills these requirements, it does lack in scalability. The proposed framework can not process requests parallelly so implementing multi-threading algorithms could make sure that the system does scale well.

4 Privacy Requirements and Threats for Blockchain-Based Systems

This section describes which privacy requirements have to be met in a blockchain-based system. Furthermore, there is a

Figure 3: Basic overview of how the decentralized mixing service CoinJoin works [20]



section that explains how the HIPAA rules [23] and GDPR regulations [37] can ensure greater privacy of healthcare information. Lastly, it discusses the threats that arise when the privacy requirements are not met.

4.1 Privacy Requirements

There are two main requirements which a blockchain-based system need to satisfy in order to protect privacy. Firstly, the links between transactions should not be visible or discoverable. Secondly, the content of the transactions is only known to their partakers [20]. These requirements are also called 'identity privacy' and 'transaction privacy', respectively.

Identity Privacy means that the real identities of the users of the blockchain can not be discovered by looking at the transactions in the blockchain. This goes further than just using encryption to create pseudonyms because this only provides a limited identity privacy. If an attacker were to monitor or analyze the transactions, they could get information about the users by using analysis strategies such as 'anti-money laundering' [33] or 'know your customer policy' [21]. They technical details of these strategies are out of the scope of this paper.

Transaction Privacy means that the contents of the transactions can only be accessed or seen by specified users, which is most cases are the parties involved in the transaction. This gives an increased level of privacy which is important when operating with sensitive information, in this case EMRs.

4.2 HIPAA and GDPR Regulations

The Healthcare Insurance Portability and Accountability Act (HIPAA) [23] is a privacy rule drawn up by the United States. The HIPAA has multiple purposes, namely defining standards for healthcare environments, improving data sharing within these environments and protecting personal healthcare data which is often sensitive information. The rules and standards set by the act cover numerous grounds. These are healthcare plans, healthcare provision, Healthcare Clearinghouses and Business Associates. Blockchain technology can benefit from the privacy rules and standards in HIPAA and this way improve security and reliability of patient personal information in the healthcare environment [17].

The General Data Protection Regulation (GDPR) [37] aims to avoid the collection of personal data when it is not essential for the intended purpose. This ensures both privacy-by-desing and by-default [6]. The rights in the GDPR need to be satisfied by the blockchain technology in healthcare as well. These rights include the right to be informed, right to withdraw consent, direct access to data, and the right to be informed on data breaches. The challenge is to combine these rights with blockchain technology because it for example states that citizens should have to ability to erase their data which may come in conflict with blockchain technology. As stated earlier, the Blockchain should be immutable, persistent and unmodifiable therefore blockchain-based healthcare systems should comply with GDPR whilst protecting the users' privacy.

4.3 Privacy Threats

The authors of [20] state that: "Due to the public nature of the blockchain network, it is possible to trace the flow of transactions to extract the userss physical identities or other additional information by data mining". The most common way of attacking is by 'de-anonymization'. De-anonymization is the use of static analysis on the blockchain to unmask users. There are multiple ways of attacking that can de-anonymize users' real identities. Since the details of these ways of attacking is out of the scope of this research, we will just briefly mention them. The attacks are:

1. Network Analysis: Because the blockchain is an open and public ledger, attackers can perform a static or network analysis to unmask users of the blockchain.
2. Address Clustering: Attackers can partition the network into different clusters of addresses because of the inherent properties of a transaction in the blockchain.

3. Transaction Fingerprinting: [3] talks about six attributes in a transaction on the Bitcoin blockchain that may define the involved parties of the transaction. These attributes are random-time interval (RTI), hour of day (HOD), time of hour (TOH), time of day (TOD), coin flow (CF) and input/output balance (IOB). By looking closer into these attributes, an attacker could possibly de-anonymize the user performing the transactions.

5 Privacy Protection Techniques

This section will describe how different techniques can be used to ensure that the privacy of the patients is protected and the requirements specified in 4.1. Firstly, mixing services are explained and it is shown how this technique can help fulfill the requirement of both identity privacy and transaction privacy. As mentioned before, it is important that only involved parties can have access to the details of a transaction and this can be ensured by mixing services. Secondly, techniques for private smart contracts will be discussed. Because a blockchain-based healthcare system uses smart contracts for access control and consent management, it is important that peers who execute the contract do not get any information on the data used in the contract. Two ways of making smart contracts private will be shown, namely Trusted Execution Environments (TEE) and Secure Multi-Party Computing (SMPC).

5.1 Mixing Services

Mixing services, which were introduced by [12], allows users of the blockchain to hide who is involved in a transaction and the content of a transaction. In this design, the mixing service will serve as a way to reduce the risk of de-anonymization and thus provide identity privacy. Furthermore, this privacy protection technique also allows the users to hide the content of the transaction and make sure it is only available for involved parties, thus ensuring transaction privacy. There are two main types of mixing services: centralized mixing and decentralized mixing.

Centralized mixing

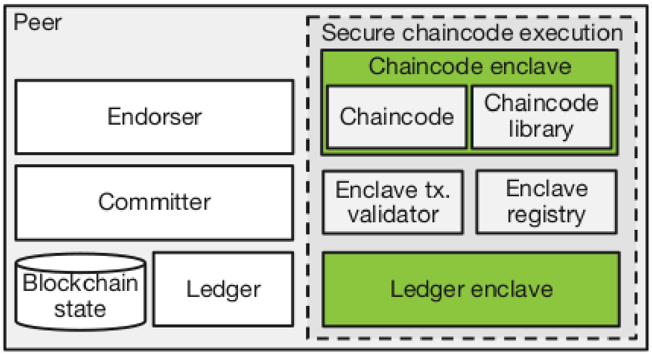
A lot of centralized mixing websites exist such as [8], [9] and [22]. They swap the transactions among different users which hides the relationship between their incoming and outgoing transactions. This is done anonymously and for most of the websites you need a TOR network to use the service. This TOR network allows the communication with the mixing service to be free and worldwide while keeping the users anonymous. There are, however, three main limitations to centralized mixing services [20]. First, the delay for transactions that is caused by the mixing service because multiple participants have to be involved in order for transactions to be mixed, is quite high. Second, the mixing server may be vulnerable to DOS attacks because it has a single point of failure. Third, users need to pay a fairly high mixing fee for most of the mixing services that are in practice right now.

Decentralized mixing services

The DOS threat caused by centralized services can be solved by using decentralized mixing services which enable a group of mutually untrusted organizations to make transactions anonymously without a third-party involved [20]. Furthermore, there are no mixing fees that have to be paid by the users since there is no extra party responsible for mixing the transactions. Different decentralized mixing services have been proposed by [26], [29], [7] and [41]. Figure 3 gives a simple overview of a transaction using [26].

To conclude, decentralized mixing services are a relatively easy way to fulfill the privacy requirement of both identity privacy and transaction privacy. And finally, most of the mixing services are compatible with existing blockchains which is also a major benefit [6].

Figure 4: Architecture of a system that uses TEE in Hyperledger Fabric [10]



5.2 Private Smart Contracts

Because a blockchain-based healthcare system relies heavily on the functionality of smart contracts (SCs) for access control and consent management, making sure that the computations in the SCs are run securely and the data in the computations is kept private is an essential element of the privacy-protection in the system. There are many different techniques that can help to create private smart contracts like SMPC, homomorphic encryption, indistinguishability obfuscation and TEE. In the next sections, we will focus on the benefits and limitations of TEE and SMPC.

Trusted Execution Environments

According to [31], a Trusted Execution Environment (TEE) is “a tamper-resistant processing environment that runs on a separation kernel.” This means that it makes sure that code ran on a TEE is authentic and kept confidential. The separation kernel, which was already mentioned in the definition, is a very important part of the TEE and was introduced in [30]. Because of the time limit of this research, it will not dig deeper into the technology of a TEE but rather look at how it can help to preserve the privacy of the users of a healthcare system. [31] states the following main benefits of TEE: it can prove its trustworthiness to third-parties, its content can be securely updated, attacks on the main memory, both software and hardware, are not effective against a TEE and exploiting backdoor security flaws are not possible. The TEE that is most notably used in blockchain-based systems is Intel's SGX [13]. However, [32] shows multiple papers have exposed the vulnerabilities of SGX, namely [11], [34] and [27]. Furthermore, they state that "all the current proposed and existing blockchains whose security rest on SGX aren't providing detailed explanations and proofs on how they are defending against these attacks".

TEE and Hyperledger Fabric
In [10], the authors introduce an architecture and a prototype for smart-contract execution with Intel SGX for Hyperledger Fabric. They state that to prevent data leakage, every peer in the network has a CPU with in integrated SGX that executes transactions inside an 'enclave'. Figure 4 shows an overview of the architecture system presented in [10].

Secure Multi-Party Computing
Secure multi-party computation [14] splits the data used in smart contracts between N parties using secret sharing. This can also be done with the states of blockchain's smart contract, also referred to as program states. Each peer contributing to the SMPC process only gets a part of the data which makes sure that the peer running the smart contract can not see what the original data is. SMPC requires the majority of the participants in the system to be honest which makes it less suitable for permission-less blockchains. This is not a problem for healthcare systems, however, since they always use permissioned blockchains. The process does, on the other hand, require exchanging of data between the peers which leads to network latency. [32] introduces a design that uses SMPC for private smart contracts. The authors state that "secure multi-party computation is more mature than the fully homomorphic methods, and has a less trusting threat model than trusted execution approaches". Another example of the use of SMPC for private SC can be found in [38]. This blockchain uses SMPC such that no third-party has to be assist in managing the accounts and keys used in the system as well as execute smart contracts securely.

SMPC and Hyperledger Fabric
[5] investigates the use of SMPC for supporting private data on Hyperledger Fabric. The authors state that the parties involved in the system store their private data on the ledger and encrypt the data with their own secret key. Furthermore, when this private data is needed for chaincode computations, the party that owns the data decrypts it and uses it as local input. This allows the chaincode to both have public data that is stored openly on the ledger and the private data as input.

5.3 Design for Privacy Protection

We can establish a final design by combining the techniques that allow for privacy protection. The first component is a decentralized mixing service that allows the users of the blockchain-based healthcare system to stay anonymous. Furthermore, it makes sure that only involved parties can see the transaction content and details. Besides ensuring identity and transaction privacy, we want the design to allow for private smart contracts such that the sensitive data in the system can be protected from malicious users. Because the TEE relies too much on the trust on hardware, which can be exploited as shown in Section 5.2, we opt to use SMPC to execute the chaincode in Hyperledger Fabric. This allows access control and consent management, which are critical functionalities of a healthcare system, to be performed in a privacy-preserving manner. As Section 5.2 discussed, there is already research being done on how SMPC can be specifically implemented in Hyperledger Fabric and after looking at benefits and limitations we can conclude that this is the best option for private smart contracts.

6 Evaluation of the Design

In this section, the design proposed in Section 5.3 will be evaluated using the privacy requirements set in Section 2.2. Firstly, it will be shown how mixing services provide device anonymity. Secondly, it will be explained how private smart contracts ensure data anonymity. Thirdly, we will discuss how mixing services also provide communication anonymity. Lastly, we will look into unlinkability and how the combination of mixing services and the private smart contracts ensure this requirement.

6.1 Device anonymity

As stated before, device anonymity means that the system makes sure that no properties of the medical device can be traced back when data is being shared. In our blockchain-based healthcare system, decentralized mixing services provide this functionality. Because the technique changes how transactions are stored in the blockchain, it makes sure that the devices stay anonymous and only involved parties can see what the original transactions were.

6.2 Data anonymity

The data anonymity requirement is covered by using private smart contracts to make sure that data can not be traced back to a specific user. As explained in Section 5, SMPC makes sure that peers running the chaincode only get a part of the data instead of the whole original data. This provides data anonymity because no one in the network can see whose data they are using for the smart contract.

6.3 Communication anonymity

Communication anonymity is a hard requirement to ensure for blockchain-based systems since all the transactions are visible in the blockchain and thus all communication with the system is visible in the blockchain. However, the mixing services can ensure that it is hard to define which user exactly is communicating at what point. This also means that the connection between user and system is hard to establish for an attacker.

6.4 Unlinkability

Keeping the user of the system and the content of the transaction anonymous or secret is important for unlinkability. This functionality is provided by mixing services in the first place. Next to that, the SMPC execution of the SC makes sure that it is not possible to establish a relationship between sender and data because the peers do not get the original data. By mixing the transactions, attackers that listen in on the network are not able to detect which users are making transactions and what the messages of these transactions are.

7 Responsible Research

This section reflects on the integrity and reproducibility of our research. Because this research does not involve any experiments or other reproducible methods, reproducibility is not relevant for this research and thus shall not be further discussed. As for integrity, during this research all cited research has been extensively checked to be reliable and trustworthy.

8 Discussion, Conclusions and Future Work

This research has been done by first getting insights in the world of blockchain. Background information on blockchain itself but also on smart contracts and Hyperledger Fabric was necessary to start answering the research question. Besides the technical knowledge, it was important to dive deeper into the security and privacy requirements of healthcare systems in general. Once these were established, it was time to look into the privacy requirements and threats of blockchain systems. By looking at the technology behind multiple techniques that allow for privacy protection and thus fulfilling the privacy requirements, we could get a better grip on what combination of techniques could tick all the requirements boxes. In this investigation, it was important to look at the benefits and limitations of each of these techniques. In the end we came to the conclusion that a combination of mixing services and SMPC is the best starting point for a blockchain-based healthcare system that wants to preserve and protect the privacy of its users. Because of time limitations, it was not possible to go into deep technical detail of these techniques and it was not possible to provide information on how they can be implemented in the system.

To come back to the research question: "How can data confidentiality be achieved in a blockchain-based healthcare system?". This research shows that multiple techniques used for privacy protection in blockchain systems have to be combined to meet the requirements set for healthcare and blockchain systems. Firstly, mixing services can be used to fulfill the blockchain privacy requirement of identity privacy and transaction privacy. Furthermore, this technique assists in providing device anonymity, communication anonymity and unlinkability which are privacy requirements that healthcare systems should meet according to research. Besides mixing services, SMPC is a powerful privacy-preservation technique that allows for private smart contracts. Private smart contracts are essential in blockchain-based healthcare systems because they allow chaincode of Hyperledger Fabric to be run privately and securely. Moreover, SMPC assists in providing data anonymity and unlinkability.

Future research in this field should include investigating how these techniques can be implemented in more technical detail. This holds for both the mixing services as SMPC. Although some research has been done in using SMPC and TEE in Hyperledger Fabric, more research is necessary to make sure all privacy requirements can be met when these techniques are implemented in healthcare systems. Additionally, the limitations of TEE should be investigated more thoroughly to see whether they can be overcome such that TEE could become a component of a blockchain-based healthcare system.

References

[1] Abdullah Al Omar, Mohammad Shahriar Rahman, Anirban Basu, and Shinsaku Kiyomoto. Medibchain: A blockchain based privacy preserving platform for healthcare data. In *International conference on security, privacy and anonymity in computation, communication and storage*, pages 534–543. Springer, 2017.

[2] Elli Androulaki, Artem Barger, Vita Bortnikov, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Christopher Ferris, Gennady Laventman, Yacov Manevich, et al. Hyperledger fabric: a distributed operating system for permissioned blockchains. In *Proceedings of the thirteenth EuroSys conference*, pages 1–15, 2018.

[3] Elli Androulaki, Ghassan O Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun. Evaluating user privacy in bitcoin. In *International Conference on Financial Cryptography and Data Security*, pages 34–51. Springer, 2013.

[4] Asaph Azaria, Ariel Ekblaw, Thiago Vieira, and Andrew Lippman. Medrec: Using blockchain for medical data access and permission management. In *2016 2nd International Conference on Open and Big Data (OBD)*, pages 25–30. IEEE, 2016.

[5] Fabrice Benhamouda, Shai Halevi, and Tzipora Halevi. Supporting private data on hyperledger fabric with secure multiparty computation. *IBM Journal of Research and Development*, 63(2/3):3–1, 2019.

[6] Jorge Bernal Bernabe, Jose Luis Canovas, Jose L Hernandez-Ramos, Rafael Torres Moreno, and Antonio Skarmeta. Privacy-preserving solutions for blockchain: Review and challenges. *IEEE Access*, 7:164908–164940, 2019.

[7] George Bissias, A Pinar Ozisik, Brian N Levine, and Marc Liberatore. Sybil-resistant mixing for bitcoin. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, pages 149–158, 2014.

[8] BitBlender. Bitblender. https://bitblender.cc. Accessed: 2021-06-22.

[9] Bitcoin Fog. Bitcoin fog. https://www.bitcoinfog.site. Accessed: 2021-06-22.

[10] Marcus Brandenburger, Christian Cachin, Rüdiger Kapitza, and Alessandro Sorniotti. Blockchain and trusted computing: Problems, pitfalls, and a solution for hyperledger fabric. *arXiv preprint arXiv:1805.08541*, 2018.

[11] Ferdinand Brasser, Urs Müller, Alexandra Dmitrienko, Kari Kostiainen, Srdjan Capkun, and Ahmad-Reza

Sadeghi. Software grand exposure:{SGX} cache attacks are practical. In *11th {USENIX} Workshop on Offensive Technologies ({WOOT} 17)*, 2017.

[12] David L Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–90, 1981.

[13] Victor Costan and Srinivas Devadas. Intel sgx explained. *IACR Cryptol. ePrint Arch.*, 2016(86):1–118, 2016.

[14] Ronald Cramer, Ivan Bjerre Damgård, et al. *Secure multiparty computation*. Cambridge University Press, 2015.

[15] Gaby G Dagher, Jordan Mohler, Matea Milojkovic, and Praneeth Babu Marella. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustainable cities and society*, 39:283–297, 2018.

[16] Eman-Yasser Daraghmi, Yousef-Awwad Daraghmi, and Shyan-Ming Yuan. Medchain: a design of blockchain-based system for medical records access and permissions management. *IEEE Access*, 7:164595–164613, 2019.

[17] Erikson Júlio De Aguiar, Bruno S Faiçal, Bhaskar Krishnamachari, and Jó Ueyama. A survey of blockchain-based strategies for healthcare. *ACM Computing Surveys (CSUR)*, 53(2):1–27, 2020.

[18] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

[19] Ruogu Fang, Samira Pouyanfar, Yimin Yang, Shu-Ching Chen, and SS Iyengar. Computational health informatics in the big data age: a survey. *ACM Computing Surveys (CSUR)*, 49(1):1–36, 2016.

[20] Qi Feng, Debiao He, Sherali Zeadally, Muhammad Khurram Khan, and Neeraj Kumar. A survey on privacy protection in blockchain system. *Journal of Network and Computer Applications*, 126:45–58, 2019.

[21] Martin Gill and Geoff Taylor. Preventing money laundering or obstructing business? financial companies' perspectives on 'know your customer'procedures. *British Journal of Criminology*, 44(4):582–594, 2004.

[22] Grams. Helix light by grams. https://www.grams-helix-light.com. Accessed: 2021-06-22.

[23] Rebecca Herold and Kevin Beaver. *The Practical Guide to HIPAA Privacy and Security Compliance*. Auerbach Publications, USA, 2nd edition, 2014.

[24] Jingwei Liu, Xiaolu Li, Lin Ye, Hongli Zhang, Xiaojiang Du, and Mohsen Guizani. Bpds: A blockchain based privacy-preserving data sharing for electronic medical records. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2018.

[25] Kelvin Ly and Yier Jin. Security studies on wearable fitness trackers. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE*, 2016.

[26] Gregory Maxwell. Coinjoin: Bitcoin privacy for the real world. In *Post on Bitcoin forum*, 2013.

[27] Ahmad Moghimi, Gorka Irazoqui, and Thomas Eisenbarth. Cachezoom: How sgx amplifies the power of cache attacks. In *International Conference on Cryptographic Hardware and Embedded Systems*, pages 69–90. Springer, 2017.

[28] AKM Newaz, Amit Kumar Sikder, Mohammad Ashiqur Rahman, and A Selcuk Uluagac. A survey on security and privacy issues in modern healthcare systems: Attacks and defenses. *arXiv preprint arXiv:2005.07359*, 2020.

[29] Tim Ruffing, Pedro Moreno-Sanchez, and Aniket Kate. Coinshuffle: Practical decentralized coin mixing for bitcoin. In *European Symposium on Research in Computer Security*, pages 345–364. Springer, 2014.

[30] John M Rushby. Design and verification of secure systems. *ACM SIGOPS Operating Systems Review*, 15(5):12–21, 1981.

[31] Mohamed Sabt, Mohammed Achemlal, and Abdelmadjid Bouabdallah. Trusted execution environment: what it is, and what it is not. In *2015 IEEE Trustcom/BigDataSE/ISPA*, volume 1, pages 57–64. IEEE, 2015.

[32] David Cerezo Sánchez. Raziel: Private and verifiable smart contracts on blockchains. *arXiv preprint arXiv:1807.09484*, 2018.

[33] Paul Allan Schott. *Reference guide to anti-money laundering and combating the financing of terrorism*. World Bank Publications, 2006.

[34] Michael Schwarz, Samuel Weiser, Daniel Gruss, Clémentine Maurice, and Stefan Mangard. Malware guard extension: Using sgx to conceal cache attacks. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 3–24. Springer, 2017.

[35] Nick Szabo. Formalizing and securing relationships on public networks. *First monday*, 1997.

[36] Bhavesh Toshniwal, Prashanth Podili, Ravula Jaysimha Reddy, and Kotaro Kataoka. Pacex: Patient-centric emr exchange in healthcare systems using blockchain. In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 0954–0960. IEEE, 2019.

[37] Paul Voigt and Axel von dem Bussche. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer Publishing Company, Incorporated, 1st edition, 2017.

[38] Wanchain Foundation. Building super financial markets for the new digital economy. https://www.wanchain.org/files/Wanchain-Whitepaper-EN-version.pdf. Accessed: 2021-06-27.

[39] Jie Xu, Kaiping Xue, Shaohua Li, Hangyu Tian, Jianan Hong, Peilin Hong, and Nenghai Yu. Healthchain: A

blockchain-based privacy preserving scheme for large-scale health data. *IEEE Internet of Things Journal*, 6(5):8770–8781, 2019.

[40] Ying Yu, Min Li, Liangliang Liu, Yaohang Li, and Jianxin Wang. Clinical big data and deep learning: Applications, challenges, and future outlooks. *Big Data Mining and Analytics*, 2(4):288–305, 2019.

[41] Jan Henrik Ziegeldorf, Fred Grossmann, Martin Henze, Nicolas Inden, and Klaus Wehrle. Coinparty: Secure multi-party mixing of bitcoins. In *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*, pages 75–86, 2015.