Security and privacy in medical data sharing through blockchain

KARIM GUETTACHE[1] , **CHHAGAN LAL**[1] , **MAURO CONTI**[1]
[1]TU Delft

Abstract

The sharing of medical data is becoming ever more important. More and more health-related data is being generated everyday and as will be shown later, its primary as well as its secondary usage brings many benefits to the healthcare system. However, medical data systems are not fail proof and are often the target of cyber-attacks, compromising the privacy of patients and the availability of the system. In fact, there are several security and privacy parameters which a medical data sharing system should ideally adhere to. These include strong authentication and unforgeability, integrity and confidentiality of data, proper consent management and access control mechanisms, availability of data and services, identity anonymity, data anonymity and unlinkability. With the emergence of blockchain technology a new possibility has emerged to realize a secure and trusted medical data sharing system across different institutes, where patients are in control of their data. Through fundamental features of blockchain, like digital signatures, a peer-to-peer network, a distributed and immutable data structure, decentralized consensus, off-chain storage and smart contracts the various security and privacy goals of a medical data sharing system can be met. Primarily these features do a successful job in addressing the security related requirements of a medical data sharing system. Authentication and unforgeability are provided through digital signatures, integrity and confidentiality are met through digital signatures and the immutable ledger, consent management and access control can be implemented with smart contracts and finally availability is achieved through the peer-to-peer nature of the blockchain. Privacy requirements are not handled well through these inherent features and require additional techniques to be met. Identity mixers protocols, relying on zero-knowledge proofs, can be used to achieve identity anonymity and unlinkability.

1 Introduction

The healthcare industry has seen tremendous advances in terms of efficiency and data management through the invention of the modern day computer and the internet [20] [23]. Nowadays almost all of a patient's health related data is stored and managed electronically in the form of Electronic Medical Record (EMR) [7] [14]. The sharing and management of this data is a hot topic as the data concerns sensitive information and is vulnerable to attacks. The sensitive nature of EMRs make them a compelling target for cyber-criminals and attackers, which look to steal this data and potentially sell them to third parties. In 2015, almost 80 million users where affected by a data breach of *Anthem Inc.*, one of the largest healthcare insurance providers in the United States [8]. In the same year *UCLA medical centre Santa Monica* was also attacked, affecting the personal health data of another 4.5 million users [9]. Despite these risks, the sharing of medical data is necessary for primary usage like treatments, operations and diagnoses and has lead to improvements in drugs and treatments in secondary usages [41] [28]. Therefore the sharing of medical data is definitely a worthwhile goal. Thus one of the biggest challenges in the industry is designing and implementing systems that allow for the managing and sharing of EMRs while guaranteeing the security and privacy (S&P) of the system and the patient's data.

With the advent of bitcoin in 2008 [31] and the subsequent emergence of blockchain technology, various research efforts and proposals have been made to use blockchain technology to create secure and privacy preserving medical data management systems. Inherent features of blockchain (BC) like 1. **Decentralization**, 2. **Transparency**, 3. **Immutability**, 4. **Distributed ledger** and 5. **Smart contracts (SCs)** [29] [44] can ensure privacy, safety and reliability of systems built on top of it. Currently there exists plenty literature on proposed blockchain based healthcare systems [17] [39] [34] [4] [10] [40] [35] [11] [13] [12] and how blockchain systems in general address various S&P requirements [29] [36] [18] [16] [43]. What lacks is an in-depth study on how blockchain handles various S&P requirements of a medical data sharing system. Therefore the question this research effort will answer is *what key benefits do blockchain technologies and smart contracts provide in the realization of a secure and trusted decentralized medical data sharing system across different institutes?*

1.1 Contribution

This paper provides an in-depth analysis of how numerous security and privacy goals are met through inherent features of blockchain based systems, all within the context of a medical data sharing platform. The contributions are as follows:

- A compilation of the requirements that make a medical data sharing system across different institutes secure, trusted and privacy preserving.
- An in-depth analysis of how the S&P of a blockchain based medical data sharing system (BCbmds) is achieved through inherent features of blockchain and smart contract, as per the identified S&P requirements.
- A showcase of how two proposed BCbmds tackle the S&P issues.
- An overview of the S&P limitations in a BCbmds.
- A look at an alternative method and how it can be integrated in a BCbmds to address the S&P limitations.

1.2 Organization

The organisation of this paper can be seen in figure 1 and is structured as follows. Section 2 covers general background information. Starting with an introduction to blockchain in section 2.1, a definition and architectural overview of a general BCbmds in section 2.2 and ending with a list of S&P requirements for a BCbmds in section 2.3. Section 3 contains the main contribution of this paper and starts with an analysis of how each S&P is achieved in a BCbmds through inherent features of BC and SCs. Section 3.1 then showcases two proposed BCbmds and how they tackle some of the S&P parameters. Section 3.2 covers the S&P limitations of a BCbmds and section 3.3 looks at alternative methods to achieve the S&P requirements which blockchain does not handle well, as well as how to integrate them into the BCbmds. Section 4 contains a reflection on the ethical aspects of this research, as well its reproducibility and integrity. Section 5 provides a discussion on the analysis done in section 3. In section 6 this paper comes to an end with a small summary, the answer to the research question in the form of a conclusion and future research directions.

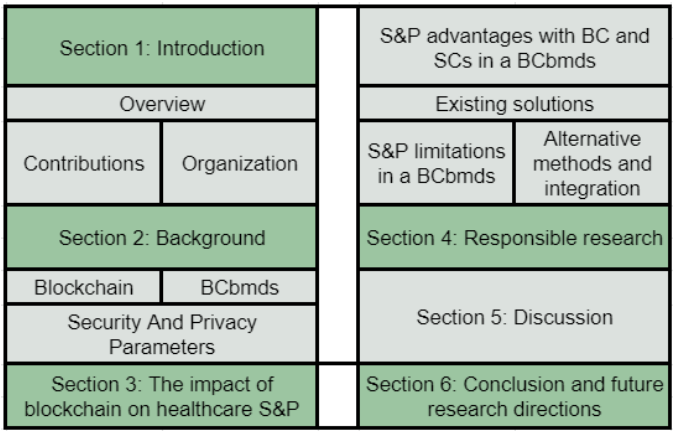

Figure 1: Organisation of this paper

2 Background

2.1 Introduction to blockchain

In 2008 the anonymous Satoshi Nakomoto first published the paper outlining the idea of bitcoin: an open, decentralized, peer-to-peer currency [31]. Bitcoin uses a hash chain data structure, also called a *blockchain (BC)*, to store transactions (txs) and a mechanism called *proof of work (PoW)* to achieve consensus among the peers [38]. The BC data structure, combined with the PoW consensus protocol give bitcoin its praised properties: an immutable ledger within a transparent, trustless network.

Soon after the inception of bitcoin more advanced cryptocurrencies started to emerge that allowed for complex transactions through programmable contracts called *smart contracts (SCs)*. The most popular project that implemented this functionality was the Ethereum protocol [19]. This new paradigm of digital currencies with complex transaction logic through programmable contracts was called BC 2.0. Research into BC technology continued and led to the conclusion that it could be used for many different applications in different domains, not only finance [30]. This new paradigm was called BC 3.0 and envisioned BC being applied to healthcare, supply chain, identity management and many other fields and sectors [30].

Different types of blockchain networks

All BC systems differ in implementation and workings, however these systems generally can be classified into three apparent groups:

- **Permissionless** systems like Bitcoin and Ethereum allow any user to participate in the network and see the ledger state. Users usually have a pseudonymous identifier, which in Bitcoin and Ethereum are their wallet and public key [31] [19]. Furthermore consensus is reached through a challenge-response based protocol to ensure safety of the network. Finally, because of this challenge-response based system such networks are generally less efficient than their counterparts, however the ledger is practically immutable [44].
- **Permissioned** or private BC systems are managed and ran by a single organisation. Each user needs permission to join the system and is authenticated with their real identity. Furthermore actions of users are managed through an access control layer. Because each user is known in private BC system, consensus protocols can be used which are more efficient than their permissionless variants [44]. An example of such a protocol is the practical Byzantine fault tolerance algorithm (PBFT), which uses state machine replication to achieve consensus [43].
- **Consortium** BC systems are a type of permissioned systems where various institutes want to work together but may not trust each other. Users require permission to join the network and are authenticated with their identity. Consortium BCs use more efficient consensus protocols like PBFT [44]. Hyperledger Fabric (HLF) is a well known project that provides an operating system to create consortium BCs in any programming language [15].

Smart contracts

The concept of smart contracts was first coined by Nick Szabo in 1996 [37]. Seventeen years later Ethereum provided one of the first successful SC blockchain implementations [19]. SCs are programs that execute certain code when being 'poked' at. The execution of this code generates certain txs that are commited to the public ledger. This code can contain any logic but usually specifies some sort of agreement between two parties which do not trust each other. The SC enforces their agreed upon terms without the need of a trusted third party. Furthermore SC code is often distributed across the network to ensure immutability and reliability [26].

2.2 Blockchain based medical data sharing system

To answer the main research question we first need to establish what a general blockchain based medical data sharing system (BCbmds) across various institutes looks like. As discussed in section 2.1, a system concerning various, but specific institutes is best implemented as a permissioned consortium BC. The different institutes in the system are defined as follows:

- Healthcare institutes like hospitals, clinics, medical offices, birth centres, blood banks and surgical centres which produce, store and use patient data
- Research institutes like laboratory or clinical research centres which use patient data for research purposes
- Patients which are everday people interacting with healthcare and research institutes. Patients are the owners of the data which the system handles and are therefore also called data owners. The terms "patient" and "data owner" will be used interchangeably from hereon.

Furthermore the general BCbmds will provide **two** core functionalities: 1. data owners can manage consent regarding their data and other institutes 2. the other institutes can read, write or share patient data based on the consent rules.

Finally, figure 2 depicts an architectural overview of a BCbmds. The systems works as follows: a BC system and SCs are created which handle identities of users, access control rules and consent management. Data owners upload their medical data to secure, offchain storage solutions. Each institute gets a certificate from a *certificate authority (CA)* which allows them to interact with the BC system. Institutes interact with SCs to manage consent or get access to patient data. We assume the BC component is built with hyperledger fabric [15], as it is a well known project sponsored by IBM for building consortium BCs [43] [15]. This presents a general BCbmds and as such this system will be used as a basis whenever we talk about a BCbmds in the rest of this paper.

2.3 Security and privacy parameters

This sub-section lists the S&P parameters that our BCbmds will be evaluated against based on literature regarding the S&P of general healthcare systems [32] [43] [42]. Besides a basic explanation, each parameter also covers its relevance and importance to a BCbmds.

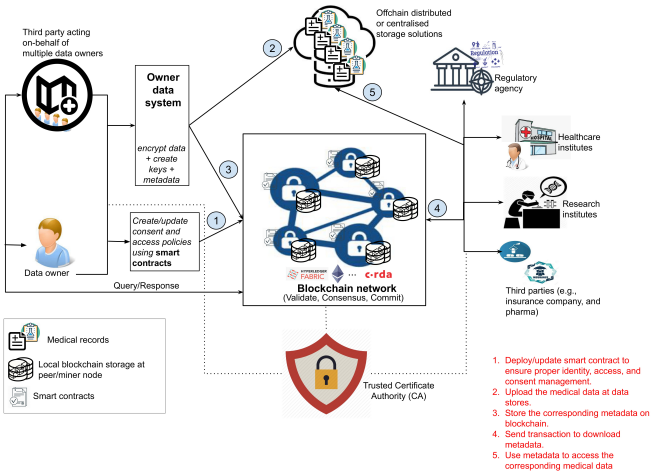

Figure 2: Architectural overview of a BCbmds across various institutes

List of security goals

The security goals are composed of authentication, confidentiality, integrity, non-repudiation, availability and access control.

1. *Authentication & unforgeability:* Any interaction with the BCbmds or exchange of data requires strong authentication of the institutes performing the action. The authentication of institutes should also be unforgeable by anyone other than the institute. Strong authentication ensures the institutes have a high degree of confidence that they are interacting with the correct entity. Furthermore it guarantees integrity and confidentiality of the BCbmds. [27].
2. *Confidentiality:* Sensitive data within the BCbmds should only be accessible by authorized and authenticated users. We can regard the confidentiality through three different goals:
   (a) *Confidentiality of patient data:* Patient data should only be accessible to institutes according to the consent rules.
   (b) *Confidentiality of consent rules:* Consent rules should only be visible to the associated patients and institutes.
   (c) *Confidentiality of data txs:* Information regarding data exchange txs should only be accessible to the institutes issuing the tx and the data owner whose data it concerns.

   Confidentiality ensures privacy of the data and actions of the institutes, as access to this information is confined to authorized participants of the BCbmds.
3. *Integrity:* Patient data, consent rules and data exchange txs within the BCbmds should not be modifiable by unauthorized or unauthenticated users. Patient data can only be modified by institutes with the correct consent permissions. Consent rules can only be updated by the corresponding patient. Date exchange tx information

can not be modified. Integrity of medical data is especially important as it is directly related to a person's health. Treatments, surgery, diagnoses and drug prescriptions all depend on the correctness of patient medical data. Furthermore integrity of consent rules ensures privacy of the patient data. Finally, integrity of exchange txs ensures correctness of the system and makes auditability possible.

4. *Non-repudiation:* Any action performed in the BCbmds; updating consent rules and data exchange txs, done by institutes, or for that matter anyone, should be logged. These logs should not be modifiable or deletable. This guarantees no one can deny an action they performed within the system. This is especially useful to audit the usage of patient data. Furthermore it ensures that institutes behaving maliciously can be held accountable as their actions are logged and they cannot delete or modify those logs in any way.
5. *Availability:* We identify two types of availability:
   (a) *Availability of services:* The BCbmds should always be available to patients wanting to update consent rules and institutes wanting to access data.
   (b) *Availability of data:* Data retrieval should always be fast and occur without delay.

   Both points are crucial for patient care as surgeries, treatments and even drug prescriptions require patient data to be readily available. Especially in medical emergencies and life threatening situations it is crucial that doctors can access patient data at any time and as fast as possible.
6. *Access control:* Institutes can only perform the actions which they are authorized for. Only data owners can manage their own consent rules and data exchange txs should be done exclusively according to the consent rules. This ensures privacy as only patients themselves can manage the consent rules of their data. It further ensures privacy because the consent rules are enforced correctly. Furthermore it guarantees integrity of the system.

List of privacy goals

The privacy goals are composed of consent management, identity anonymity, data anonymity, unlinkability and transparency & auditability.

1. *Consent management:* Patients should be able to manage the consent rules regarding their data. This provides patients with privacy of their data as they can specify which institutions can access their data at which time and when they can't access that data anymore. Furthermore they can specify what type of access these institutes have, being *read*, *write* and, or *sharing* rights.
2. *Identity anonymity:* The identities of the institutes in the BCbmds should be anonymous when they perform actions like updating consent or exchanging data. Practically this means that observed actions in the system should not **directly** reveal the identity of the action taker. We thus aim for *pseudonymity*. This ensures privacy of the institutes as their actions in the system are decoupled from their identity. The only exception to this is that data owners *can* see the identities of the institutes taking actions regarding their data. We think this tradeoff is justifiable as a patients ability to audit the usage of their data outweighs the requirement of institutes to remain anonymous to them. Additionally, total identity anonymity is an unrealistic goal as managing consent and exchanging data relies on knowing the identities of the counter parties. Therefore we limit this requirement to only actions performed in the system.
3. *Data anonymity:* Patient data should not be able to identify the patient when this is specified in the consent rules. Furthermore GDPR rectal 26 requires all stored personal data to be either anonymized or pseudonymized [1]. However, based on the consent rules, certain institutes, like a doctor performing medical treatment, should be able to de-anonymize the data when this is permitted. Data anonymity ensures privacy of the patients when research institutes or other parties use or process their data.
4. *Unlinkability:* Aggregating actions performed in the BCbmds should not provide additional information on the action taker than could be obtained from a single action. Analyzing multiple consent rule updates or data exchange txs should not provide additional information on the data owner, the institute or their relationship. This property accordingly guarantees the privacy of the institutes.
5. *Transparency & Auditability:* Consent rules and data exchange txs should be fully transparent to the data owner and the corresponding institutes. Furthermore consent rules and data txs should be linked to identifiable institutes such that their actions can be audited. Transparency ensures a user can audit the usage of their data and hold institutes that misbehave accountable. This disincentives institutes from misbehaving and thus increases the S&P of the patient's data.

3 The impact of blockchain on healthcare security and privacy

Through inherent features of BC and SC technology a trusted and secure BCbmds can be realized. Each distinct feature of BC and SC contributes to the overall S&P of the system. This section presents an analysis of how each inherent BC feature and SCs provide S&P to the BCbmds by addressing the identified S&P parameters of section 2.3. Table 1 contains an overview of each inherent feature of BC and which S&P parameters they address.

Membership service & digital signatures

In BC systems authentication is achieved through asymmetric cryptography, where users have a public and private key [25]. Ownership of the private key authenticates users as the owner of their public key through digital signatures [36]. In consortium BCs users do not merely have a public-private key pair but a certificate that contains additional information on them [44]. A common standard for such certificates

is the X.509 standard [6]. Each participant in a consortium BC gets their certificate from a trusted third party called the Certificate Authority (CA) [15], which legitimizes the certificate by supplying their own digital signature. Furthermore in hyperledger fabric, each transaction or communication requires a digital signature of the sender's certificate. These mechanisms ensure strong *authentication* of the institutes in a BCbmds. Each institute is pre-authenticated by receiving a certificate from the CA, which contains identifiable information on them [6] [15]. Furthermore the institutes in the BCbmds BC only accept messages or transactions from other institutes which contain digital signatures of certificates issued by the CA. This way the institutes in the BCbmds can ensure communications are coming from other legitimate institutes. By using strong cryptographic algorithms [2] to create digital signatures, a high degree of *unforgeability* is ensured as it is extremely difficult to compute or derive a private key from a signature or public key [25].

Certificates and digital signatures further certify that *access control* and *consent management* can be implemented since these mechanisms rely on the identities, or attributes derived from identities, of the institutes. Signatures provide proof of ownership of data for patients, which consent management mechanisms can build upon. Additionally these properties ensure *confidentiality of data* and *integrity*. Transactions cannot be done without digital signatures of institute certificates and thus there is strong proof of the correct entities having performed txs. Next to that, the public key of institutions can be used to encrypt data which they are allowed to access before it is sent to them. This further strengthens confidentiality as only the recipient institution can decrypt this data. Finally, *non-repudation* [25] and *auditability* are also achieved with digital signatures from certificates as all actions performed by institutes are tied to a signature, which is tied to their identity. Thus they can not repudiate any action they have done, and furthermore their actions can be audited.

With all these benefits it must be said that the reliability of the aforementioned security services depend for a big part on the security of the private key [33]. Techniques like biometric authentication can be used to strengthen the safety of the public key infrastructure and as such the security of the system [33].

Peer-to-peer network

BC systems consist of nodes or peers that communicate in a peer-to-peer (p2p) fashion. Each peer contains a copy of the BC and the SCs [44]. This property greatly benefits the *availability* of the system, as there is no single point of failure like in a client server model. There is a great likelihood that some nodes will still be available after an attack on the BCbmds nodes is performed. *Ransomware* or *denial-of-service (dos)* attacks on some nodes do not affect the overall availability of the system [43]. Therefore, institutes can be confident that the system services are always available to them.

Distributed and immutable data structure

One of the core features of BC systems is a distribute, practically immutable data structure that often takes the form of a hash-chain. From now on we will refer to this data structure as the ledger. The immutable ledger provides many benefits to a BCbmds. First of all it ensures there is a single source of truth for *access control* mechanisms. The ledger can be used to keep record of the consent rules, through which an access control mechanism can be implemented. *Integrity* of the consent rules is also provided this way as they are directly stored on the ledger, which automatically gives them the immutability property. Patient data, however, is often stored off-chain to preserve privacy, however a hashed version of the data can be stored on the ledger. This provides a form of integrity to the patient data as the most up-to-date hash represent an immutable truth of what the data should be. The fact that we can only store the hash of the patient data on the ledger and keep the rest off-chain also strengthens the *data anonymity*. The hash ensures that nodes observing the ledger cannot derive any information on the patient's data or identity.

The ledger is distributed and stored locally on all nodes, which provides the system with great *transparency*. Furthermore *non-repudiation* and *auditability* are also achieved by having an immutable ledger as this ledger can function as a bookkeeping mechanism of all actions performed in the system. Institutes can not perform actions without them being committed to the ledger thus they cannot repudiate actions they have performed in the past. This further gives data owners the ability to audit the usage of their data, as they also store a local copy of the ledger.

Consensus

Consortium BC often use PBFT algorithms to achieve consensus among the nodes [44] [15]. Hyperledger fabric can process up to 10.000 transactions per second [24], which is sufficient to address the data exchange tx that happen in a BCbmds. The *availability* requirement of fast data access is satisfied by this metric. Moreover, *access control* in a BC system fundamentally relies on consensus of the nodes. A strong consensus algorithm like PBFT ensures that access control rules are enforced correctly even in the presence of a number of faulty nodes. This also leads to greater *confidentiality of data* and *integrity* as nodes will always agree on what the correct txs in the system are. At last, consensus strengthens *non-repudiation* as it serves a proof that an institute has performed an action related to the committed tx.

Smart contracts

A consortium BC built in hyperledger fabric uses chaincodes to provide programmable contract functionality [15]. *Consent management* logic can be implemented in SCs. The SCs are stored on all the peers and the code is stored on the immutable ledger. This way the SC contract code guarantees execution of the consent management logic as was initially specified. Similarly, *access control* techniques can also be implemented with SCs. SCs can further be used to provide *data anonymity* by implementing an "anonymize" option in the consent logic. Institutes requesting data can then get an anonymized version of that data based on this option, through a SC that handles the data exchanges in the BCbmds. Finally, *transparency & auditability* can be achieved by programming a SC that returns all on-chain data regarding a patients medical data. This contract can return a history and current state of consent rules as well as a history of data exchange txs that

happened. This way the patient can see and audit all the actions done regarding their data.

Off-chain storage

Patient data can be stored off-chain in a cloud or on institute systems. Distributed data storage techniques can be used to ensure this data is encrypted and thus remains *confidential*. Only by interacting with the BC can the actual data be retrieved and modified. As such, *integrity* is also maintained. Finally, replicated storage can enhance the *availability* of the data. If one storage solution fails, the replicated nodes will ensure the data is still accessible. Finally off-chain storage, as opposed to on-chain storage, reduces the resources required to maintain the ledger. If all the patient data were stored directly on the ledger, the nodes would require significantly more resources to store this data locally [5]. Off-chain allows us to leverage the cloud, or devices that we designed specifically for the purpose of storing large amounts of data.

3.1 How existing systems tackle S&P requirements

Two proposed BCbmds are analyzed on how they tackle some of the S&P requirements. The first system has no name [21] and uses a consortium BC built on HLF. This is similar to our proposed system. The second system is MedRec which uses a different approach; MedRec is a permissionless BCbmds that uses ethereum smart contracts for authentication and other operations [17].

Alevtina Dubovitska et. al. [21]

1. ***Authentication:*** A membership service plugged into hyperledger is used to authenticate users in a blockchain based data management and sharing system for radiation oncology. The membership service registers users with different roles. The different roles are used to specify which actions can be performed with the chaincodes used in the system. To verify the legitimacy of the users the membership service consults a national practitioner data bank through which the users can identify themselves as legitimate healthcare institutes or research institutes. Furthermore the membership service hosts a CA that generates key pairs for signing and encryption for each user.
2. ***Consent management & access control:*** "Permission blocks" stored in the world state and chaincodes are used to manage consent and access control. Every permission block stores a timestamped consent form containing the ID of the clinician which has obtained consent, the timeframe of the consent, the category of data which the clinician can access, the type of consent they have (read, write or share), study IDs to allow the clinician to share the data for research purposes and an anonymity tag for the latter to specify if the data must be anonymized before sharing it with researchers. By allowing the patient to specify the timeframe and the category of data the patient is enabled with fine-grained access control of their data. Furthermore the timestamp makes each block unique and allows for updating permissions as the permission block with the latest timestamp is considered the current state.
3. ***Integrity:*** The hash of off-chain patient data is stored on the ledger as well as a timestamp, clinician ID and more to ensure unforgeability of data. Furthermore data can only be added or modified by a clinician based on the permissions on the ledger.
4. **Confidentiality of data:** Each user generates a symmetric encryption key with which they create a pseudonymous identity and encrypt their data. The users can then encrypt this key with authorized participants public key to allow them to view the identity and the data. Confidentiality is maintained as only the authorized participants can decrypt the shared key with their private key. If the symmetric key is lost or abused the user can simply generate a new one and use a proxy re-encryption algorithm to re-encrypt their data.
5. **Availability:** All off-chain data is stored on a cloud. This cloud can be accessed by role-based APIs as long as there is a node registered and connected to the network. Furthermore this system uses symmetric key encryption and proxy re-encryption algorithms to make sure that on-and off-chain data can still be recovered if private keys are lost.

MedRec [17]

1. ***Authentication:*** A special smart contract called the registrat contract binds a readable, generally known identifier like a hospital's name, to an Ethereum address to authenticate each participant. Furthermore SC policies are used to handle new registrations and to change existing ones. These policies coded in the SC ensure registration is restricted to only trusted and verified institutes.
2. ***Consent management & access control:*** MedRec uses a Patient Provider Relationship (PPR) smart contract on Ethereum to provide consent management and fine-grained access control of patient data. For each relationship where a data provider stores the data of a data owner (for example a hospital for a patient) a PPR SC is issued which contains a mapping of addresses and query strings that can be used to access the data when executed on the provider's machine. The data owner can allow other participants to access their data by adding a new entry in the PPR SC consisting of the Ethereum address of the other party and a query string that defines which data is retrieved and which isn't. Through specifying the query string the data owners are thus enabled with fine-grained access control. When a third party tries to access data with their query string the data providers machine can query the PPR SC to check if the third party has access before executing the query string. When access is revoked the data provider's machine can simply reject an invalid data request from the third party which might have remembered the query string.
3. ***Integrity:*** Each query string in a PPR SC contains a hash of the data which can verify its integrity.
4. ***Transparency & auditability:*** MedRec uses a summary SC which holds a list of references to peers. This list represents all the previous and current relationships in the

Inherent features of BC						
S&P parameters	Membership service & digital signatures	P2P network	Distributed and immutable ledger	consensus	Smart contracts	Off-chain storage
Authentication & unforgeability	x					
Confidentiality	x		x	x		x
Integrity	x		x	x		x
Non-repudiation	x		x	x		
Availability		x	x	x		x
Access control	x		x	x	x	
Consent management					x	
Identity anonymity						
Data anonymity			x		x	
Unlinkability						
Transparency & auditability	x		x		x	

Table 1: A mapping of inherent BC features to S&P parameters

system. A data owner can call this SC and get a summary of all the relationships regarding them and their data, in the present and in the past. This provides them with full transparency of the state regarding their data and they can audit who has been using their data and who hasn't. Furthermore participants can leave and rejoin the network at any time, and can always still retrieve their history. Auditability is achieved because all the public keys of the participants are linked to a real life identifier.

3.2 Security and Privacy limitations

In spite of the many advantages BC provides to meet S&P goals, it also has it limits. This section addresses the S&P parameters which BC does not handle well. Most of these are related to privacy.

Identity anonymity

A BCbmds running a consortium BC makes it difficult to achieve identity anonymity. Consortium BCs rely on knowing the identities of the participants as not everyone is allowed to join the network. Each institute in a BCbmds has a x.509 [6] certificate that contains identifiable information on the organisation they belong to and other identifiable information. Furthermore the nature of BC also requires each tx to be signed, which in a consortium BC inherently couples the identity of the tx issuer to the tx. Therefore identities are known and not anonymous on each action performed in the system. Finally the identities of institutes are necessary to provide elementary functions of the system like consent management and access control as these rely on parties knowing each other.

Unlinkability

Unlinkability is also not inherently supported in a consortium BC. All txs get verified by all peers in the network. Furthermore all txs are coupled to the certificates of the institutes and so is the ledger state in general. Through this, txs and ledger updates can be analyzed to find patterns and gain additional information [43]. Hyperledger fabric endorsement policies could provide a fix to this since not all peers execute the txs in the system. However the validator nodes still see all the read write sets as well as the identities of the endorsers that endorsed the txs and they have access to the public ledger [15]. With this information certain linkages can still be made between patients and healthcare organizations. Tx fingerprints can reveal information about the user. Six attributes of txs that can reveal information and de-anonymize the issuer 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) [16]. The authors of [16] even conducted a study where 40% of users profiles were recovered after analyzing their txs in the bitcoin network. Even if they used another wallet for each tx [16]. Similar techniques can be applied in a BCbmds by certain institutes to learn more about the other institutes.

Confidentiality of consent rules

Confidentiality of data is achieved in a BCbmds as analyzed above. However confidentiality of consent rules is more difficult as 1. all txs need to be verified by all nodes, which allows them to see consent rule updates and 2. the ledger which contains the consent rules is locally available to all nodes. This

means each institute running a node in the BCbmds can see the consent updates and current consent state, as well as the history of all the patients.

Confidentiality of data transactions
Similarly to the point made above, confidentiality of data exchange txs is difficult because of the open nature of BC and the fact that all txs need to be verified by all peers. Therefore all peers will see the data exchange txs that happen in the network, which destroys the confidentiality of them.

3.3 Alternative solutions

This section describes an alternative solution to achieve identity anonymity and how this solution can be integrated in a BCbmds.

Identity mixers are cryptographic protocols that can be used to provide *identity anonymity*, as well as *unlinkability* [18] [3]. These protocols work by replacing standard x.509 certificates with a new kind of certificate. From these certificates special *presentation* tokens can be derived, which act as a proof of digital signatures on certain attributes without disclosing the signature or the attribute value itself. This is achieved through zero-knowledge proofs [22]. Verifiers verify these tokens based on an access control policy called the *presentation policy*, which specifies which attributes, or predicate about attributes the user should include in the presentation token [3].

Identity anonymity and unlinkability are provided through this as on each tx an institute can generate a fresh new presentation token which is completely unlinkable to previously used presentation tokens. "Neither the CA, nor the verifiers can tell if two presentation tokens were derived from the same or two different certificates" [3]. Furthermore the presentation tokens do not reveal any information on the user or attributes of the user, but merely a proof that a signature on some attribute is valid and that the user is in possession of the corresponding private key.

The properties of authentication and unforgeability are still maintained with identity mixers as the certificate issuance process is similar to the standard x.509 certificate procedure: the certificates, which are a set of attributes, are digitally signed by the CA. Furthermore each institute possesses a private key associated to the certificate. Next to that, auditability is also not compromised as identity mixers allow certain specially assigned institutes to break the unlinkability of certain txs under certain circumstances [3].

Finally, there exists an identity mixer MSP that can be integrated in hyperledger fabric. It works as follows [3]:

1. The CA generates a signing key pair and the public key is made public to all participants
2. An institute generates its own private key
3. The institute requests a certificate from the CA
4. The CA issues the certificate in the form of an identity mixer
5. The certificate contains the attributes that the institute has and is stored along with the private key on the institutes device
6. The institute can generate a new unlinkable presentation token each time it needs to sign a tx
7. The token proves possession of a certificate, as well as possession of the attributes required for the presentation policy
8. The verifier can verify the presentation token with the CA's public key

4 Responsible Research

This research effort mainly focused on the security and privacy of medical data sharing through BC. We evaluated a BCbmds against a list of S&P parameters and looked at how BC handles each one of them. This section will first describe the ethical concerns associated with the topic of S&P and BCbmds. Then we will reflect on the reproducibility of the research done and finally on the integrity.

Ethical concerns
From an ethical perspective the research around the S&P of medical data is beneficial to the patients, as it can potentially guard them from future harm and mistreatment. However, what is easily neglected is that the average patient may not have the required knowledge to adequately analyze the S&P benefits of a BCbmds. This knowledge asymmetry may lead the implementer of such a system to take advantage of the patients, because the patients do not know how to evaluate the system they are interacting with. Besides that the patients may have a hard time to get convinced of the S&P benefits of a BCbmds and they may therefore choose to not use the system despite those benefits.

Another point to be made is that the permissioned nature of the described BCbmds can lead to discrimination against certain users. At the end of the day, the CA decides who can and who can not use the system, as they distribute the certificates needed to join the system. Through this they can unjustly deny certain patients access to the healthcare system, based on political or ethnicity reasons. Furthermore they could unjustly deny certain institutions from joining the system based on political or competitive reasons.

Reproducibility
The research done is adequately reproducible as it consists of an analysis done mostly on publicly available information. Next to that, no experiments where conducted so there was not any reliance on randomness or other unpredictable factors. However, some of the information used in this research came from academic papers which are only accessible through subscriptions to academic sites. An average person might not have the resources to take out those subscriptions to obtain access to all those papers. Moreover, the reproduction of this research does require some level of knowledge on computer science (CS) related topics, that an average person is not expected to have. As mentioned in the first point of ethical concerns, an average patient wanting to assess the S&P of the BCbmds will have a hard time doing so if they are not familiar with CS or BC. All in all though this research if fairly reproducible.

Integrity
Integrity is maintained as all the sections are consistent with the presented problems and research question in the introduction. Next to that a lot of statements made in this paper are backed up by scientific references, which further adds to the integrity of the presented information. However, the research was done in a relatively short period by someone who is not an industry expert. This could potentially have affected the integrity of the research. Finally, this research effort was done within a research group, but this paper was written entirely by one person. So, some other perspectives were taken into account when writing this paper but the most contribution came from one person only. Thus, compared to a collaborative research effort, this paper naturally lacks some variety in perspective.

5 Discussion

A secure and trusted decentralized medical data sharing system across different institutes can only be achieved if the system has strong S&P guarantees. Many such systems have been proposed [20], but none of them have addressed the underlying assumption that BC and SC technology can successfully provide the necessary S&P services for such applications. This paper aimed to fill that gap by mapping medical data sharing S&P requirements to inherent features of BC and SCs. A few remarks can be made from this analysis. First of all, BC and SC technologies are still in their infancy stage and it is to be expected that many new breakthroughs and refinements are still to come in the future. Therefore the inherent features of BC and SCs might differ from the ones that have been identified here, although we do not expect this difference to be so big that a paradigm shift is caused. Second of all, while we considered some features "inherent" to BC and SCs and others not, this distinction is not set in stone and falls within a gray area, given the diversity of the different BC implementations. Especially regarding various cryptography techniques or off-chain storage solutions it is hard to say whether they are inherent to the BC model or just additional features. As a third point it has to be noted that, while this paper analyzed each S&P requirement mostly in isolation, in reality, different trade-offs have to be made between the S&P requirements when realizing a BCbmds. For the most part these trade-offs regard the *transparency* and *privacy* aspects of such a system. Fourth, the analysis done in this paper relied on certain properties of BC models, without going into detail about how well these properties are actually enforced. For example, the immutability guarantees in a consortium BC using a PBFT consensus algorithm differs quite a lot from those in a PoW permissionless model [44]. When actually designing a BCbmds, such details should be taken into consideration. We merely assumed the properties and derived our analysis from those assumptions, without analyzing to what extent those assumptions hold in various circumstances.

6 Conclusions and Future Work

This paper aimed to present an analysis of various S&P requirements of a medical data sharing system and how BC and SC technology can be used to address those requirements. To achieve this first of all a set of S&P requirements was compiled from analyzing the existing literature. This set included parameters like *authentication*, *integrity*, *confidentiality*, *consent management*, *identity privacy* and *unlinkability*. Then, supporting on existing literature of BC features and security and privacy services, an analysis was done to see how BC and SCs inherently meet the S&P requirements of a medical data sharing system. Some inherent features of BC include *digital signatures*, *distributed immutable ledger*, *peer-to-peer network* and *decentralized consensus*. The inherent features of BC and SCs do a great job of addressing the security parameters that were identified. However, they lacked in addressing the various privacy requirements that were identified. Fortunately there are various techniques to mitigate these limitations, which we described in section 3.3. One of those techniques are *identity mixers*, which improve identity anonymity and unlinkability in the system, without strongly compromising other features like authentication and auditability. To conclude, the answer to our main research question "*what key benefits do blockchain technologies and smart contracts provide in the realization of a secure and trusted decentralized medical data sharing system across different institutes?*" is that inherent features of BC and SCs provide great benefits in realizing security services for a BCbmds.

Future work in this area can include the following:

- Analyze the extent to which BC S&P features are met by going more into implementation details
- Look at the social aspect of security, where a malicious entity could get access to the BCbmds through social engineering and how BC handles this
- Perform a similar analysis as this paper did, but beyond the scope of solely medical data sharing. A look could be done at how IoT device [23] S&P can benefit from BC and SC technology
- Analyze how trade-offs in BC and SC implementations affect the overall S&P of the system.

References

[1] 2018 reform of eu data protection rules.

[2] Hyperledger fabric; security model.

[3] Hyperledger membership service provider (msp) implementation with identity mixer.

[4] Medibloc techinical whitepaper(eng)v0.3. Technical report.

[5] Why new off-chain storage is required for blockchains. Technical report.

[6] X.509 : Information technology - open systems interconnection - the directory: Public-key and attribute certificate frameworks. Technical report.

[7] Ehr incentive programs, 2014.

[8] Millions of anthem customers targeted in cyberattack, 2015.

[9] Ucla health victim of a criminal cyber attack, 2015.

[10] The biggest doctor-patient environment based on blockchain. Technical report, 2018.

[11] Clinicoin - blockchain powered global wellness. Technical report, 2018.

[12] Medicalchain. Technical report, 2018.

[13] Medx protocol - launch unstoppable medical. Technical report, 2018.

[14] What is an electronic health record (ehr)?, 2019.

[15] Elli Androulaki, Artem Barger, Vita Bortnikov, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Christopher Ferris, Gennady Laventman, Yacov Manevich, Srinivasan Muralidharan, Chet Murthy, Binh Nguyen, Manish Sethi, Gari Singh, Keith Smith, Alessandro Sorniotti, Chrysoula Stathakopoulou, Marko Vukolić, Sharon Weed Cocco, and Jason Yellick. Hyperledger fabric: A distributed operating system for permissioned blockchains. 2018.

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

[17] 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, 2016.

[18] Sotirios Brotsis, Nicholas Kolokotronis, Konstantinos Limniotis, Gueltoum Bendiab, and Stavros Shiaeles. On the security and privacy of hyperledger fabric: Challenges and open issues. In *2020 IEEE World Congress on Services (SERVICES)*, pages 197–204, 2020.

[19] Vitalik Buterin. A next-generation smart contract and decentralized application platform. Technical report, 2013.

[20] Erikson Júlio De Aguiar, Bruno S. Faiçal, Bhaskar Krishnamachari, and Jó Ueyama. A survey of blockchain-based strategies for healthcare. *ACM Comput. Surv.*, 53(2), March 2020.

[21] A. Dubovitskaya, Z. Xu, S. Ryu, M. Schumacher, and F. Wang. Secure and Trustable Electronic Medical Records Sharing using Blockchain. *AMIA Annu Symp Proc*, 2017:650–659, 2017.

[22] Uriel Feige, Amos Fiat, and Adi Shamir. Zero-knowledge proofs of identity. *Journal of Cryptology*, 1:77–94, 06 1988.

[23] Jigna J. Hathaliya and Sudeep Tanwar. An exhaustive survey on security and privacy issues in healthcare 4.0. *Computer Communications*, 153:311–335, March 2020.

[24] Hao Jin, Yan Luo, Peilong Li, and Jomol Mathew. A review of secure and privacy-preserving medical data sharing. *IEEE Access*, 7:61656–61669, 2019.

[25] Don Johnson, Alfred Menezes, and Scott Vanstone. The elliptic curve digital signature algorithm (ecdsa). *International Journal of Information Security*, 1(1):36–63, Aug 2001.

[26] Shafaq Naheed Khan, Faiza Loukil, Chirine Ghedira-Guegan, Elhadj Benkhelifa, and Anoud Bani-Hani. Blockchain smart contracts: Applications, challenges, and future trends. *Peer-to-Peer Networking and Applications*, April 2021.

[27] Atsushi Kogetsu, Soichi Ogishima, and Kazuto Kato. Authentication of patients and participants in health information exchange and consent for medical research: A key step for privacy protection, respect for autonomy, and trustworthiness. *Frontiers in Genetics*, 9, June 2018.

[28] Sunil Kumar and Maninder Singh. Big data analytics for healthcare industry: impact, applications, and tools. *Big Data Mining and Analytics*, 2(1):48–57, 2019.

[29] Iuon-Chang Lin and Tzu-Chun Liao. 2. A survey of blockchain security issues and challenges. international. *International Journal of Network Security*, 2017.

[30] Damiano Di Francesco Maesa and Paolo Mori. Blockchain 3.0 applications survey. *Journal of Parallel and Distributed Computing*, 138:99–114, April 2020.

[31] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. Technical report, 2008.

[32] AKM Iqridar 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, 2020.

[33] Thi Hoang Lan Nguyen and Thi Thu Hang Nguyen. An approach to protect private key using fingerprint biometric encryption key in biopki based security system. In *2008 10th International Conference on Control, Automation, Robotics and Vision*, pages 1595–1599, 2008.

[34] Abdullah Al Omar, Mohammad Shahriar Rahman, Anirban Basu, and Shinsaku Kiyomoto. MediBchain: A blockchain based privacy preserving platform for healthcare data. In *Security, Privacy, and Anonymity in Computation, Communication, and Storage*, pages 534–543. Springer International Publishing, 2017.

[35] Seyed Morteza Pournaghi, Majid Bayat, and Yaghoub Farjami. MedSBA: a novel and secure scheme to share medical data based on blockchain technology and attribute-based encryption. *Journal of Ambient Intelligence and Humanized Computing*, 11(11):4613–4641, January 2020.

[36] Tara Salman, Maede Zolanvari, Aiman Erbad, Raj Jain, and Mohammed Samaka. Security services using blockchains: A state of the art survey. *IEEE Communications Surveys Tutorials*, 21(1):858–880, 2019.

[37] Nick Szabo. Smart contracts: Building blocks for digital markets, 1996.

[38] Florian Tschorsch and Björn Scheuermann. Bitcoin and beyond: A technical survey on decentralized digital

currencies. *IEEE Communications Surveys Tutorials*, 18(3):2084–2123, 2016.

[39] Joachim Sandgaard Steve Wishstar. Medchain. Technical report, 2018.

[40] Qi Xia, Emmanuel Boateng Sifah, Kwame Omono Asamoah, Jianbin Gao, Xiaojiang Du, and Mohsen Guizani. Medshare: Trust-less medical data sharing among cloud service providers via blockchain. *IEEE Access*, 5:14757–14767, 2017.

[41] 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.

[42] Buket Yüksel, Alptekin Küpçü, and Öznur Özkasap. Research issues for privacy and security of electronic health services. *Future Generation Computer Systems*, 68:1–13, March 2017.

[43] Rui Zhang, Rui Xue, and Ling Liu. Security and privacy on blockchain. 52(3), 2019.

[44] Zibin Zheng, Shaoan Xie, Hong-Ning Dai, Xiangping Chen, and Huaimin Wang. Blockchain challenges and opportunities: A survey. *International Journal of Web and Grid Services*, 14:352, 10 2018.