PAPER SUMMARY

Secure and Trustable Electronic Medical Records Sharing using Blockchain

1. INTRODUCTION

“Secure and Trustable Electronic Medical Records Sharing using Blockchain” is a research paper published in August 2017 as part of the American Medical Informatics Association (AMIA) Annual Symposium 2017. The paper was authored by five researchers: Alevtina Dubovitskaya MS (University of Applied Sciences Western Switzerland, École polytechnique fédérale de Lausanne), Zhigang Xu MD (Stony Brook Medicine), Samuel Ryu MD (Stony Brook Medicine), Michael Schumacher PhD (University of Applied Sciences Western Switzerland) and Fusheng Wang PhD (Stony Brook University). The paper opens by discussing the problems surrounding existing solutions for storage and management of Electronic Medical Records (EMRs) and then provides background on the blockchain technology. Next three applications for blockchain technology within the healthcare field are discussed and the paper proposes a framework to make the storage, management and sharing of patient’s EMRs more secure, trustable and convenient. The proposed framework utilises a permissioned blockchain system in combination with local and cloud databases. The paper details a prototype based on the proposed framework developed in collaboration with the Department of Radiation Oncology at a major US hospital.

This paper aims to provide a summary of the key contributions of the above paper, as well as providing a critical evaluation.

2. SUMMARY

2.1. PROBLEM STATEMENT

The healthcare industry generates a vast quantity of data much of which is critical for positive healthcare outcomes but also highly sensitive and private. Patient EMRs are an example of such data. The data contained in an EMR is essential for accurate diagnosis and treatment. However, the information contained will need to be distributed, shared and updated frequently. It is a major challenge to ensure that a clinician always has access to the most up to date records, whilst also ensuring that the information is managed according to the patients wishes and secured from malicious actors.

During the course of their lifetime, a patient will likely visit many different medical facilities, all of which will require access to the most up to date records. According to legislation (in some territories), a patient has the right to manage their medical information and will need to provide consent to any institution requiring access to it. The consent process can be arduous, often requiring a patient to specify what data will be shared, who can receive it and among others, the period of time for which the data can be accessed. This can be difficult to co-ordinate, especially if the patient is unwell or does not know ahead of time which institute will need access to the data. Even after consent has been provided, the process of transferring the data can be very slow as most hospitals rely on paper records and these will need to copied and sent via the post.

Other than facilitating successful diagnosis and treatment, patient data is very useful for research purposes. However, the sharing of any patient data for research purposes also requires consent, unless the data can be anonymised. Research has shown that the independent release of locally anonymised data from multiple sources about the same patient can be used to identify the patient; which can be a violation of privacy. Therefore, the anonymisation process must be carefully considered. One possible solution is the use of a single centralised authority to store and manage patient data in accordance with the wishes of the patient. This can be achieved in two ways: the centralised storage of encrypted data, or choosing a fully trusted entity that has full access to sensitive information. Neither are ideal - the former requires massive amounts of memory in order to conduct operations over encrypted data and the latter requires the patients to put their trust in an organisation that could potentially misuse it. Both imply a single point of failure and performance bottleneck.

The discussed paper argues that a blockchain-based system could overcome many of the challenges discussed above. A distributed ledger could provide a shared, immutable and transparent history of all actions taken by all the participants of the network. These actions could include a patient altering permissions, a doctor accessing or uploading data, or sharing data for research purposes. The authors argue that blockchain technology guarantees data security, control over sensitive data, integrity, tamper resistance and convenient healthcare data management between the patient and various medical actors without relying on a single trusted party.

2.2. Blockchain Background

Blockchains are a class of peer-to-peer distributed ledger technologies that enable the creation of a distributed and verifiable database where all participants can be sure that they have identical data. Every transaction that occurs on a blockchain is executed and shared by every participant, as well as being permanently and verifiably recorded. Blockchains were originally used within the financial industry in order to facilitate digital currencies, however, the technology has since been applied in numerous fields. Blockchains can be broadly divided into two categories: permissioned and permissionless. In permissionless blockchains, the identities of the participants are pseudo-anonymous or anonymous, and all users have the same permissions. In a permissioned blockchain, the identities of the users are known, and users have different access rights and permissions based on their role as prescribed by an identity provider. In contrast to permissionless blockchains, participants in permissioned blockchain systems are required to trust an identity provider which acts as a central authority. As a result, there is less of an incentive for malicious behaviour as the identity of users is revealed to the identity server and participation is restricted to a known set of users.

A popular example of a permissionless blockchain is the Ethereum chain, which acts as a global, decentralised and distributed virtual machine that allows any user to create and execute arbitrary code within the Ethereum Virtual Machine (EVM) specification. This code forms “smart contracts” that users and other smart contracts can interact with, enabling a wide variety of applications. Users must pay for their code execution within the virtual machine using the chain’s native cryptocurrency: Ether. The amount of ether that users must pay is based on the computation complexity of the operations they are seeking to perform and this acts as an incentive to limit complexity. Consensus across the network is achieved through a process called Proof of Work (PoW) in which participants compete to find a “nonce” which satisfies certain requirements. Participants of the consensus mechanism are known as “miners” who also perform the computation necessary to update the state of the virtual machine. Their services are rewarded with a variable “mining reward” as well as the Ether used to pay for code execution. PoW blockchains have been criticised for their power efficiency and alternative consensus mechanisms such as Proof of Stake (PoS) have been proposed, however, the suitability of such mechanisms is still an open question.

A popular example of a permissioned blockchain system is Hyperledger which is an open-source project developed by the Linux Foundation. Hyperledger has a modular architecture which enables the use of a variety of consensus mechanisms. The services provided by Hyperledger can be divided into three categories: Membership, Blockchain and Chaincode. Membership services manage identity, privacy and confidentially on the network. Every user is assigned a username and password and issued a corresponding Enrolment Certificate (ECert). Every transaction is associated with a Transaction Certificate (TCert). Users can generate new TCerts at will to preserve anonymity and the mapping between TCerts and ECerts is only known by the membership service. Blockchain services manage the distributed ledger through a peer-to-peer HTTP/2 protocol. Chaincode services are used to create smart contracts, which are able to interact with data on the blockchain, users and other smart contracts.

2.3. Applications of Blockchain Technology in Healthcare

2.3.1. Primary Patient Care

The authors claim that the use of blockchain technology in primary patient care can help address the following problems in current healthcare systems:

- A patient will often access multiple unrelated healthcare institutions during the course of a lifetime. The patient will need to maintain a complete medical history and if they are unable to do so, some of the information required by clinicians may not be available.
- In the case that laboratory results are unavailable, a patient may have to repeat tests that they have already taken as it is quicker than obtaining the information from another institution. This is a waste of time and resources.
- Despite the fact that healthcare data is sensitive and its management complex, a convenient system to allow patients to maintain access control policy in an efficient and privacy-preserving manner does not yet exist.
- Sharing data between healthcare providers may be difficult and time consuming.

The authors propose two ways blockchain technology could be utilised in order to solve the problems discussed above:

- *Institution-based:* A network of peers would be formed where each peer/node is a healthcare institution such as a hospital or general practice. The peers maintain a distributed ledger secured through a consensus algorithm. The patient (or his representative) will be able to access and manage all their data through an application connecting to one of the nodes. If one of the nodes is offline, the patient can still access the data through any other online node. The access control policy will be encoded in chaincode ensuring the security and privacy of the data stored on the ledger.
- *Case-specific*: A case specific ledger would be created during a patient's stay in a hospital. The network would connect all involved clinicians (doctors, nurses, specialists) with each other and the patient they are caring for in order to improve efficiency and transparency relating to the treatment. This would help to eliminate human-made errors and to ensure consensus in the case of a disagreement between practitioners.

2.3.2. Data Aggregation for Research

For data to be useful for research purposes, it is very important to verify that the sources of data are trusted institutions and the data is authentic. However, this must be done in a way that balances the need for patient privacy. Due to the current lack of mechanisms to do this, patients are often unwilling to take part in data sharing. The use of a shared ledger between various medical institutions would facilitate the process of collecting patient data for research purposes whilst guaranteeing patient privacy and transparency of the aggregation process.

2.3.3. Connecting Providers for Better Patient Care

"Connected Health" is an area of active research and aims to maximise healthcare resources and improve collaboration between different healthcare players whilst providing more opportunities for consumers to engage with caregivers and improve self-management of a health condition. Providing access to the shared ledger to additional entities such as insurance companies and pharmacies can lead to more efficient cost management and provide a more complete picture across the whole treatment process. For example, a pharmacy with access to the ledger can provide more efficient logistics, whilst an insurance company could provide real time cost estimates for treatments.

2.4. Proposed Solution

2.4.1. Use Case

The authors aimed to produce a framework to support EMR sharing for primary patient care. They focused on cancer patients receiving care in the Department of Radiation Oncology of a major US hospital.

Cancer is a serious medical condition that often requires long-term treatments and monitoring over the lifetime of the patient. Due to this, accurate medical histories and the availability of them to clinicians is extremely important. A patient may not use the same hospital for all treatments and so the sharing of data between hospitals is also very important. However, at present, this can be a very cumbersome process.

The authors base their solution on an existing local oncology information system called ARIA. ARIA combines radiation, medical and surgical information and enables clinicians to manage different kinds of oncological data, develop cancer-specific care plans, and monitor patient dosage. The different types of data stored in the system can be organised according to the needs of the clinician and then exported to PDF format. The most important and widely used documents are those that contain medical history, physical exams, laboratory results and delivered radiation doses.

Currently, a time consuming and arduous process must be followed if data is to be transferred from one hospital to another. Firstly, the patient must sign a consent form, then, the data must be printed and posted to the recipient hospital. Consent management and data transfer can become complicated and inconvenient: a patient may need to contact a member of staff and sign a consent form at a hospital they no longer receive care at. The data transfer can take a long time, especially if there are difficulties in posting it and once a hard copy of the data has arrived it will need to be introduced to the system at the recipient hospital. The authors argue that it is very difficult for a patient to maintain any fine grain access control of their data or to obtain a complete view of the data in light of the complexity of this process. They believe that a blockchain-based solution will make the consent process more convenient and speed up the transfer of data whilst offering the patient more fine-grained and easier to use access control.

2.4.2. Permissioned vs Permissionless

The authors justified their decision to use a permissioned blockchain system with the following reasons:

- The anonymity of users and impossibility to verify the identity of users of a permissionless blockchain system could facilitate impersonation and data misuse.
- Healthcare data is highly sensitive, even just monitoring the communication between patients and clinicians could reveal valuable information and therefore its best to restrict the participants of the network to avoid privacy violations.
- Fast response and availability of the system is important as clinicians must make crucial decisions quickly.
- The need to pay for a transaction could limit the usefulness of the system.

2.4.3. System Architecture

Figure 1, taken from the discussed paper, shows the architecture of the proposed oncology-specific system. The system consists of the membership service, databases for storing data off-chain, nodes which manage the consensus process and APIs for different user roles. The authors note that whilst currently only the doctor and patient roles are considered, these could be extended based on the scenario.

The job of the membership service is to register new users with different roles. The roles define the functionality of the chaincode available to the user. When a new user with a privileged role (such as a doctor) is registered, it is important to ensure that that the user is qualified for the role. To verify this in the case of Doctors in the US, the membership service could interrogate the National Practitioner Data Bank. The membership service is responsible hosting a certification authority involved in the generation of a signing key pair ($SK^S{}_U$, $PK^S{}_U$) and encryption key pair ($SK^E{}_U$, $PK^E{}_U$) for every user (U).

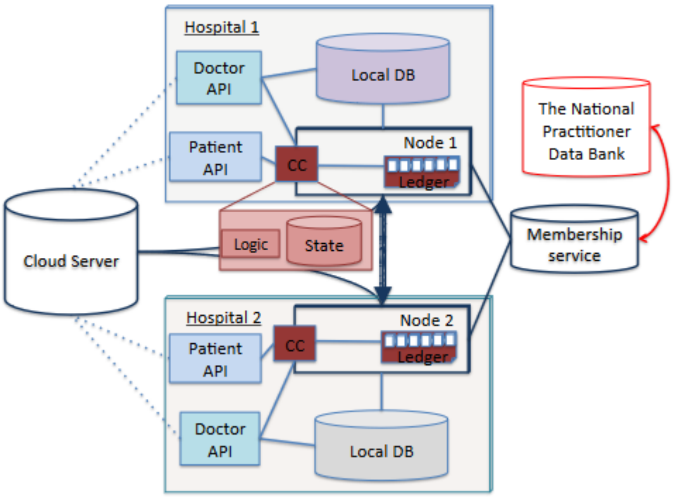


FIGURE 1 THE SYSTEM ARCHITECTURE OF BLOCKCHAIN BASED DATA MANAGEMENT AND SHARING FOR RADIATION ONCOLOGY.

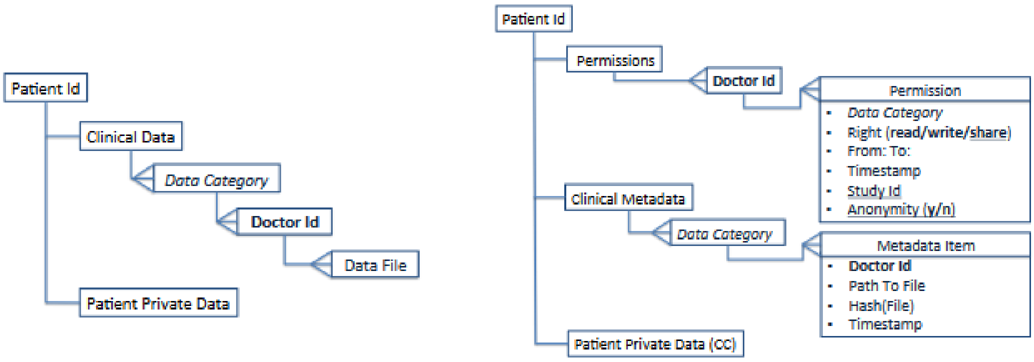
If the new user has the patient role (P), then they also generate a symmetric encryption key ($SK^{AES}{}_P$) that is used to encrypt and decrypt data related to the patient P. This key will also be used to generate pseudo-anonymous identities for the user so that only authorised users can verify if the ledger stores any information about the patient. When a patient P wants to share information with a doctor D, the patient shares their key $SK^{AES}{}_P$ using the encryption public key of the doctor $PK^E{}_D$. If a patient key $SK^{AES}{}_P$ is compromised, they can generate a new key, run a proxy re-encryption algorithm and then share the new key with clinicians according to the desired access policy.

The patient medical data is stored off-chain in two databases. A local database at each hospital stores data in whichever format the hospital prefers (in the case of the oncology system this would be ARIA). A database provided by a cloud service stores the patient's data organised into categories and encrypted by the patients symmetric key $SK^{AES}{}_P$. A clinician can view or upload data to the cloud database based on the access policy defined by the patient and implemented by the chaincode.

Each node acts as a Hyperledger validator and receive all transactions submitted by users through the role-based API. One of the nodes is selected as the leader and organised the transactions into a block and then initiates the PBFT consensus protocol. Transactions are executed by all nodes according to the chaincode logic. The blockchain records the system state and is used to store information about patients in a key-value format. The key is a patient ID and is a pseudonym of the patient that is generated by hashing a combination of the patient's symmetric key $SK^{AES}{}_P$ and Uniquely Identifiable Information about the patient (UII_P): $H(SK^{AES}{}_P || UII_P)$. Any combination of Social Security Number (or equivalent), date of birth, name, address or post code could be used as a patient's UII_P.

2.4.4. Data Structure and Functionality

Figure 2, taken from the discussed paper shows how a patient's data and metadata is organised. The patient's medical data is stored locally at the hospital and in the cloud organised into categories as shown in figure 2(a). The categories can be defined based on the scenario. For their oncology scenario the authors defined three categories: history and physical exams, laboratory results and delivered radiation dose. Data files are further categorised under the clinician that uploaded them. Patients may also store private data encrypted using their encryption public key $PK^S{}_p$.



(a) The structure of the patient's data stored in the cloud.

(b) The structure of the patient's metadata stored on the chaincode.

FIGURE 2 THE DATA STRUCTURE OF A PATIENT RECORD.

Figure 2(b) shows how the patient's metadata is stored on chain. Metadata is divided into Permissions, Clinical Metadata, Patient Private Data (optional). For the permissions block, every Permission corresponds to a doctor's ID registered in the system, specifies a timeframe for which the right is valid and specifies which of the rights (read, write or share) that the clinician has. If the right is share, then the patient can also specify which study they are allowing their data to be shared with and whether their data should be anonymised first.

Clinical metadata contains information about the data files uploaded to the cloud service. The metadata is categorised based on the semantics of the corresponding data files. Every item contains the ID of the clinician who uploaded the file, a path to where the file is stored in the cloud, a hash of the data file to ensure that the data hasn't been modified and a timestamp of when the file was uploaded. The patient again has the option to store optional data on the chain.

2.4.5. Prototype

The authors succeeded in creating a prototype of the proposed solution consisting of a membership service and four nodes. They deployed chaincode written in the Go programming language to each node and tested both "invoke" transactions (that add new records to the databases and chain) and "query" transactions (that view data stored in the system). Patients are able to create metadata records, add permissions and retrieve up-to-date records. Doctors are able to upload, access and share data in the cloud based on the permissions specified by the patient. Doctors are only able to access a patient record if they have the correct access right and it has not expired. They are also not able to share the data for research purposes unless the patient has agreed to this. These rules are all guaranteed by the chaincode implementation.

3. Critical Reflection

The prototype system described and delivered by the authors meet the author's goals of providing a secure and trustable system. Immutability and transparency are inherent to the blockchain, and these properties are passed onto system facilitating a trustworthy and convenient record system.

A patient's privacy is guaranteed by the fine-grain access control offered by the chaincode implementation as long as the consensus mechanism continues to operate correctly. The consensus mechanism could only fail in the scenario that a fraction of the nodes become dishonest actors. The integration of a membership service means that the identities of all the participants are known, discouraging malicious behaviour and making Sybil attacks impossible. Additionally, as the system is not fully decentralised, there is the option to restrict malicious nodes. Furthermore, the use of the membership services and, in particular its integration with the National Practitioner

Data Bank ensures that all users with the doctor role are indeed qualified to hold that role. Patient privacy is further enhanced by the inability of the membership service to access the clinical information of a patient. This is because (through the use of pseudo-anonymous identities) all clinical data is linked to the patient's identity through their secret key SK^{AES}_{P}, known only to them. Finally, the system offers the ability to recover data in the case that the patient has lost (or compromised) their secret key through proxy re-encryption.

Confidentially is guaranteed through the use of symmetric encryption across all patient data files where the key is known only to the patient. Only the patient is able to share this key, giving them the sole ability to manage access control to such data. Data integrity is offered through the hash of every data file stored immutably on the blockchain. This hash is further signed by the patient signing secret key SK^{S}_{p}, further ensuring data integrity and unforgeability.

The authors claim that data availability is guaranteed through the use of cloud services to store the data. Whilst cloud services do avoid a single point of failure and most offer high uptime guarantees, there is always the risk of a service outage which could prevent a user accessing data at a critical time. At the time of writing no cloud providers offer a 100% uptime guarantee and there have been incidents of service outages in the past such as [1,2,3]. This issue is alleviated somewhat by the presence of local databases at each hospital; however, this is only helpful if clinicians have already retrieved the data they need from the cloud providers.

Scalability of the system is vital if it is to see widespread adoption. The authors assert that the PBFT consensus protocol used by the proposed system provides excellent scalability in terms of the number of users but admit node scaling has not yet been explored sufficiently. They also note that performance tweaks such as adjusting the frequency of block creation and number of transactions per block could be explored. System load is minimised through the hybrid approach of using the blockchain for metadata but storing the heavy data on cloud services.

The system described is a critical system as the consequences of failing to obtain a record on demand could have severe consequences on patient outcomes, potentially leading to injury or even death. For instance, a doctor unable to access a patient's medical records due to a service outage could administer a drug that the patient is allergic to, or begin a treatment that has already proven ineffective. The authors have taken care to address the issues of patient privacy, security, confidentiality, data integrity and unforgeability; however, they have not focused on availability and scalability with the same rigour. In the authors' defence, they have said that they would like to make scalability the focus of a future work. The notions of reliability and dependability have not been directly addressed.

References

[1] David Ramel, Visual Studio Magazine 04/13/2020 "Microsoft Confirms March Azure Outage due to COVID-19 Strains": https://visualstudiomagazine.com/articles/2020/04/13/azure-outage.aspx

[2] Donna Goodison, CRN 26/03/2020 "Google Cloud Outage Attributed To 'Infrastructure Components' Issues": https://www.crn.com/news/cloud/google-cloud-outage-attributed-to-infrastructure-components-issues

[3] Frederic Lardinois, Techcrunch 10/06/2020 "IBM Cloud suffers prolonged outage": https://techcrunch.com/2020/06/09/ibm-cloud-suffers-prolonged-outage