Electronic Medical Record Security Sharing Model Based on Blockchain

Sihua Wu
College of Computer Science and Technology
Chongqing University of Posts and Telecommunications,
Chongqing, China
waalscquptedu@163.com

Jiang Du
School of Cyber Security and Information Law
Chongqing University of Posts and Telecommunications,
Chongqing, China
clouddu@163.com

ABSTRACT

The emergence of electronic medical records has provided great convenience for the storage and analysis of medical data. However, electronic medical records contain a large amount of personal privacy information, it is still very difficult to share medical information among various medical institutions. As the underlying technology of Bitcoin, blockchain technology has the characteristics of decentralization, security, trustworthiness, collective maintenance, and cannot be tampered, it is suitable for data protection and sharing. In this paper, data masking technology and Inter Planetary File System (IPFS) are introduced to build a safe and efficient electronic medical record sharing model based on blockchain. The model can not only guarantee the security of medical data, but also save resources in blockchain.

CCS Concepts

Security and privacy~Privacy protections

Keywords

blockchain; medical records; privacy protection; data masking

1. INTRODUCTION

Since electronic medical records are easy to store, operate and share, more and more medical institutions are attracted to them. As of 2015, at least 10% of medical records have been electronically stored, a 90% increase from 2008[1]. It is foreseeable that the electronic medical record will completely replace the paper medical record soon. Medical information has its particularity, including rich and detailed patient information and data. The reasonable use of these information and data will play a positive role in promoting the development of medical research. However, there are still many problems in the process of sharing medical information, such as, 1) The application of data sharing by medical institutions shall be subject to strict review.2) Third-party platforms that publish medical data are vulnerable to internal or external attacks that reveal patient privacy.3) Medical data is easily changed or lost during transmission. All these have seriously hindered the sharing of medical data and the development of medical big data.

Blockchain technology originated in a 2008 paper published by a cryptographic email group under the pseudonym satoshi nakamoto Bitcoin: A peer-to-peer electronic cash system [2]. Its features contains decentralization, time-series data, collective maintenance, programmability, security and reliability, etc. In the distributed system where nodes do not need to trust each other. Blockchain technology achieves point-to-point transactions, coordination, and operations. Thus it provides solutions to the problems of high costs, inefficiency, and insecure data storage that exist in centralized organizations.

Based on the characteristics of blockchain, this paper proposes an electronic medical record security sharing model based on blockchain (EMRSB), blockchain technology can ensure that published medical data will not be lost and tampered in the process of sharing, saving time for data verification. Data masking technology is introduced to solve the problem of privacy leakage at the expense of the accuracy of some data without damaging the statistical characteristics of data. Meanwhile, in order to store a variety of detailed medical record data, IPFS is used to store the detailed data of medical record, saving valuable storage resources on the blockchain.

2. RELATED WORK

The rapid development of blockchain technology has attracted wide attention from governments and financial institutions. Many countries have issued relevant white papers, including China's "White Paper on China's Blockchain Technology and Application Development (2016)", the UK's "Distributed Ledger Technology: Beyond the Blockchain", actively promoted The application of blockchain in financial and government affairs. In addition, blockchain technology has gradually been introduced into the medical field, including medical data applications, medical record sharing, personal privacy protection, medical payment and drug anti-counterfeiting, etc.

Because medical information has specific requirements for information security and privacy protection, blockchain technology, as a powerful tool for data integrity protection in information security, has great development space. Based on the advantages of blockchain, some related researches on blockchain combined with medical scenes have appeared at home and abroad. Ivan [3] proposed a method for securely storing patient medical records based on blockchain, which enhanced the transparency of medical information. Kuo et al. [4] used blockchain private chain technology to create an inter-agency medical health prediction

ICCSP 2019, January 19–21, 2019, Kuala Lumpur, Malaysia

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6618-2/19/01…$15.00

DOI: https://doi.org/10.1145/3309074.3309079

model. Xue et al. [5] provided a blockchain-based medical data sharing model, which provides a new idea for data sharing in medical institutions. Shrier et al. [6] pointed out to use the OPAL/Enigma encryption platform of the Massachusetts Institute of Technology to implement secure storage of medical data with blockchain technology and encryption technology. Ekblaw et al. [7] indicated a novel decentralized electronic medical record management system, which first proposed the use of smart contracts to achieve rights management.

By comparing existing electronic medical record sharing technologies and solutions, there may be problems that require complicated encryption algorithms to consume a large number of blockchain resources or vulnerabilities in smart contracts. This paper proposes a blockchain-based electronic medical record security sharing model.

3. KEY TECHNOLOGIES

In order to implement the electronic medical records security sharing model on blockchain（EMRSB）, we will use various key technologies. The next part of this chapter will give a brief introduction to hash function, blockchain, data masking, and IPFS file system.

3.1 Hash Function

The hash function maps input values of arbitrary length to binary values of relatively shorter fixed-length. For example, the more common SHA256 algorithm in Bitcoin maps an input of any length to a fixed-length output of 256 bits, which is called a hash value. Hash function generally has the following three typical characteristics [8]:

Collision resistance. A hash function H is said to be collision resistant if it is infeasible to find two values, x and y , such that $x \neq y$, yet $H\left(x\right)=H\left(y\right)$.

Hiding. A hash function H is hiding if: when a secret value r is chosen from a probability distribution that has high min-entropy, then given $H\left(\mathrm{r}\ \mathrm{P}x\right)$ it is infeasible to find x .

Puzzle friendliness. A hash function H is said to be puzzle-friendly if for every possible n-bit output value y , if k is chosen from a distribution with high min-entropy, then it is infeasible to find x such that $H\left(k\ \mathrm{P}x\right)$=$y$ in time significantly less than.

The blockchain is based on the above characteristics of the hash function, which has been widely applied. For example, the hash pointer in the blockchain header, the Merkle-tree structure that stores transaction records and the IPFS file system. These all use the hash function.

3.2 Blockchain

Blockchain is a kind of chain data structure that combines data blocks in sequence according to the time sequence, and guarantees non-tamperable and non-forgerable decentralized sharing general ledger by means of cryptography, it can safely store simple sequential data that can be validated within the system [9]. The data structure of the blockchain consists of two parts, namely, the inner block structure and the chained structure between blocks, as shown in Figure 1. A block structure contains header information and body information.

The header information is the metadata of the block, which is used to verify the block and establish the association with its predecessor and successor block. Typically, the header information contains timestamp, the hash value of the precursor block, nonce and Merkle-Tree root value. Block body information is a sequence of transactions.

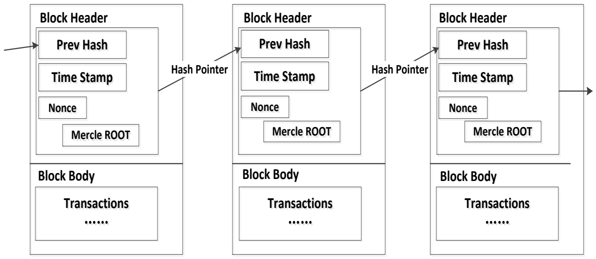

Figure 1. Block structure diagram

Transaction sequences in the blockchain are organized by the data structure merkle-tree, which is an important part of ensuring that data can't be tampered with in the block chain. Merkle-Tree's leaf node records the hash of the transaction, the value of each internal node is the hash of all its children, the value of the root node can be considered as the signature of the entire tree, any change in data will result in a different hash value for the root node. The structure and calculation process of Merkle-Tree is shown in Figure 2 [10]. Using the properties of Merkle-Tree can be used to implement data set equality testing, location modification and zero-knowledge proof. In the blockchain, merkle-tree is mostly used to detect whether the block copies are the same.

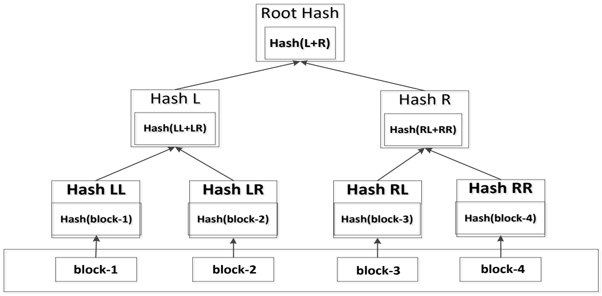

Figure 2. Merkle tree structure.

3.3 Data Masking

Data masking refers to the transformation and hiding of certain sensitive information into virtual data through data masking rules (generally specific algorithms). The purpose of data masking is to protect users' privacy information on the basis of preserving a certain degree of statistical value. Data masking can be achieved by a variety of techniques, such as substitution, generalization, numerical transformation, encryption, occlusion, control insertion/deletion and so on [11-12].

According to the specific application scenarios, data masking rules should be design differently for different objects. This paper introduces data masking to protect users' personal privacy in the application scenario of electronic medical records, by masking and generalizing sensitive information of patient, the privacy leakage problem in the process of electronic medical records sharing on the blockchain is solved.

3.4 IPFS

Inter Planetary File System(IPFS) is a peer to peer file system, which is a single Bitorrent cluster using Git repository for distributed storage. In order to connect all computer devices to the same file system and ensure that files are not tampered with, IPFS will be a good choice. IPFS can not only store files in various formats, but also return the hash value of the current file when the file is uploaded to the IPFS system. You only need to use the hash code as an index when accessing the same file for the next time. In this paper, IPFS is mainly used to store detailed information related to electronic medical records, saving valuable storage resources on the blockchain.

4. ELECTRONIC MEDICAL RECORD SECURITY SHARING MODEL BASED ON BLOCKCHAIN

we have proposed an electronic medical records security sharing model based on blockchain (EMRSB), which is mainly aimed at solving the problem of data lost, tampered and patient privacy leakage in the process of medical records data sharing. At the same time, EMRSB can save valuable resources on the blockchain. The overall framework of the model is shown in Figure 3, it shows a complete process for the patient (doctor) to upload and review the medical records.

Because of the particularity of the medical scenario, EMRSB adopts the framework of alliance chain to ensure the stable operation of the whole model. Where patients and doctors must be authorized before joining.

Data masking is a commonly used method of privacy protection. For the sharing of electronic medical records, we use generalization, occlusion and replacement techniques to protect patient privacy information.

① and ② in Figure 3 represent a process in which the user submits data that contains sensitive information and returns security data.

With the development of medical big data, information in electronic medical records presents many forms. If all the data, especially the pictures are shared in the blockchain, it will occupy a lot of resources and produce block expansion problem, which is extremely difficult for the growing mass medical data storage. Therefore, we upload excessive electronic medical records to IPFS file system and return the Hash value of current files, which can not only guarantee that the files stored in IPFS are not tampered, but also save precious storage resources on the blockchain. ③ and ④ in Figure 3 are processes for uploading the remaining various electronic medical records data to the IPFS file system and returning the Hash value of the uploaded file.

⑥ in Figure 3 shows that the blocks in EMRSB have a safe and efficient consensus-building process through an improved consensus algorithm to ensure stable operation of the model.

This model mainly serves patients and doctors, ⑤ in Figure 3 shows that patients (doctors) anchor the data of medical records after data masking and the Hash value of the file on the block to Merle root and store it, ⑦ in Figure 3 represents the process of returning data from a blockchain when a doctor (patient) submits a medical record query.

Over the next few sections, we will introduce the data masking, block storage and consensus algorithm in detail.

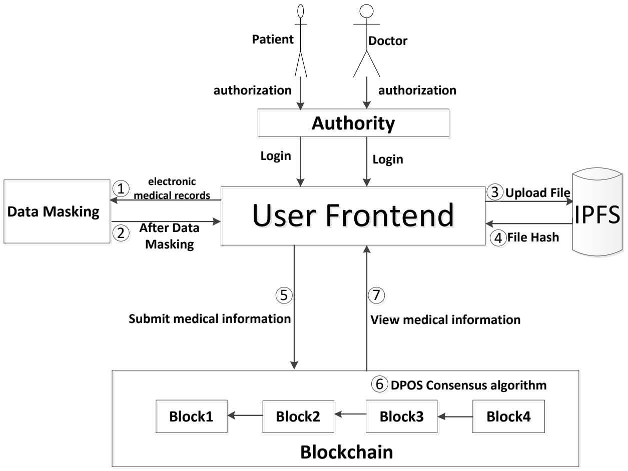

Figure 3. The framework of electronic medical record security sharing model based on blockchain

4.1 Data Masking Design

There are many algorithms for data masking. Thus, it is necessary to select appropriate algorithms for sensitive data fields in specific application scenarios, which can not only guarantee the value of data, but also prevent the leakage of personal privacy information. In the medical scene, the sensitive information field of electronic medical records mainly includes name, age, id number, mobile number, home address, doctor's advice, etc. This kind of data can be divided into three types of processing methods, namely, for fields with certain statistical value, such as age, home address, medical date, etc., this paper adopts a generalized design method. Mask out is used for fields that clearly identify the data of the individual patient. For instance, name and id number. Other fields are not processed and stored in clear text for doctors to view patient information easily.

Generalization refers to replacing a specific value in a data table with a more abstract value. For a numeric attribute, we could replace the value with a range containing the value, for a categorical attribute, we could replace it with a wider range of values. The generalization algorithm can be identified by the semantic tree structure. The age generalization tree is shown in Figure 4 and the home address generalization tree is shown in Figure 5 Generalization has the typical characteristics that the more specific the underlying node information, the greater the statistical value; the higher the layer's nodes contain more abstract data, the smaller the value; and the root node does not contain any meaning. For the age of the patient, a mid-level generalization can be used, while the home address is more important to the patient and a higher generalization result can be used.

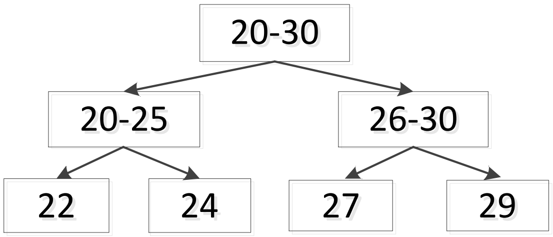

Figure 4. Age generalized tree

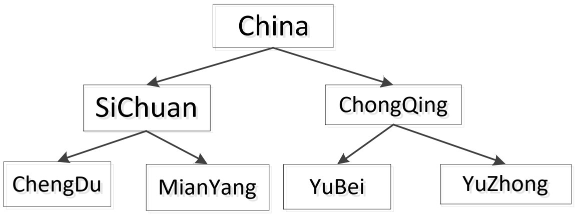

Figure 5. Home address generalization tree

Masking out is the partial replacement of sensitive content with masks (such as "&, #, *"), so that sensitive data remains partially disclosed. Normally, the data after masking out is not available. For example, the first six digits of China's ID card are the address code, the 7-14th digit is the birth date code, the 15th-17th is the sequence code, the 18th digit is the check code. To protect users' privacy, we would deal with the 7-17th with masking. Table 1 and Table 2 show the data comparison after the above data masking algorithm.

The images stored in the IPFS are generally medical images in the DCM format, and the medical images in the DCM format are matured by open source software for data masking. Therefore, the files uploaded to IPFS by default are processed by software.

Table 1. Original data

Name	Age	ID Number	Address	Disease
YangYang	60	533527195809210238	Chaoyang park, chaoyang district, Beijing	flu
Sihua Wu	24	500101199310020723	Chongwen road, nanan district, chongqing	stomach trouble
Qiang Zhang	43	620503197512200856	Mai ji district, tianshui city, gansu province	headache

Table 2. After data masking

Name	Age	ID Number	Address	Disease
* Yang	[60, 65]	533527**********0238	chaoyang district, Beijing	flu
** Wu	[20, 25]	500101**********0723	nanan district, chongqing	stomach trouble
*Zhang	[40, 45]	620503**********0856	tianshui city,gansu province	headache

4.2 Block Storage

Data publishing to block storage is an important part of the EMRSB, which mainly includes three procedures of data publishing and field design of storing data on block.

When the patient or doctor publishes the data, the following three processes are required:

1) First, the data is processed separately according to the agreed data masking rules, and the processed data is returned.

2) Upload files such as image data to the IPFS file system and record the hash values was returned.

3) Submit the processed data and the Hash value of the image to the leaf node of the Merkle-tree in the form of a transaction.

Since Bitcoin uses the POW consensus algorithm, it is stable to generate a block every 10 minutes, and the data is frozen every 10 minutes in the model. The Merkle root generated by the representative node in the consensus algorithm is submitted to the bitcoin blockchain. If so, we can ensure that the published data will not be lost or changed. For the public part data of the blockchain, which is security and will not reveal personal privacy.

Each block can store n pieces of electronic medical data. The composition of the medical data mainly includes three parts, including the public key field, metadata and the data summary part. The public key field mainly stores the corresponding patient and the doctor's public key as a unique representation. The metadata part stores the electronic resources of the data, such as the generated timestamp and other information. The data summary part is the data after data masking and the Hash value of image file are stored in the form of key-value pairs. The storage structure of the data block is shown in Figure 6.

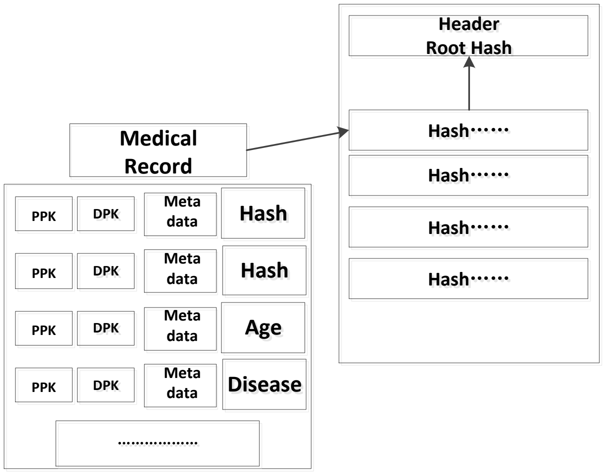

Figure 6. Medical record data block storage structure

4.3 Improved Consensus Algorithm

At present, the most successful consensus algorithm is the Proof of Work (POW) adopted by bitcoin blockchain. The core idea of POW is to guarantee data consistency and consensus security by introducing computational power competition of distributed nodes. In the Bitcoin system, each node competes with each other based on their respective computing powers to solve a complicated but easily verifiable SHA256 math problem, the fastest node to solve this problem will get the block bookkeeping rights and the bitcoin rewards automatically generated by the system. To make use of the POW consensus algorithm requires a lot of nodes to contribute their own computing power to obtain high returns, which is not applicable to the EMRSB model. Therefore, according to the application scenario, the EMRSB model introduces the delegated proof of stake (DPOS) [13] to design the consensus algorithm.

In the design of DPOS algorithm, three kinds of related roles are introduced:

Normal Role: This kind of roles is mainly doctors and patients. It adopts the architecture design of alliance chain. Each roles needs authorization to join.

Board Representation Role: This type of roles selects top 101 medical institutions nationwide, it is mainly responsible for recording requests submitted by normal roles into blocks and signing with its own private key

Monitoring Node: These roles are mainly medical government institutions. Responsible for checking in turn whether each signed block is valid or not.

Once the Monitoring Role finds that the Board Representation Role has signed an invalid block or faked, the Board Representation Role will be kicked out of the current system and will be punished accordingly. This design guarantees the safe and stable operation of the consensus algorithm through the credibility of top medical institutions and government endorsements. The detailed process of the consensus algorithm is as follows:

step1: The patient (doctor) submits a record request and submits the doctor (patient) public key as the identification.

step2: A Board representation role accepts the request.

step3: The Board representation node broadcast has received a request.

step4: The current node adds the record to the transaction based on the patient (doctor) public key and simultaneously broadcasts the transaction.

step5: The Monitoring node checks the records, and the other nodes update the data according to the broadcast.

step6: Check the number of transactions in the block every 1 minute, reach the agreed quantity to form a data block, and calculate the Merkle root of the data block.

step7: After reaching the appointed time (e.g, 10 minutes for bitcoin) anchor the Merkle root of all newly generated blocks to the blockchain.

step8: Returen to setp1.

5. CONCLUSIONS

In this paper, we have proposed an electronic medical records security sharing model based on blockchain (EMRSB), which provides a safe and efficient solution for the sharing of medical records information. On the one hand, EMRSB solves the problem that the data is easily tampered and lost in the process of medical records sharing by the advantages of the blockchain technology. On the other hand, we store large files in IPFS file system and added the hash of the file to the blockchain, which saves valuable storage resources on the blockchain. Meanwhile, according to the characteristics of different format data in medical records information, we also select different data masking technologies to mask and replace sensitive information of users, especially for the privacy leakage problem that may occur in the process of medical records sharing. These can further increase the security of patient privacy information.

As a frontier technology in the field of information security, blockchain is still developing constantly. Due to its architectural advantages, blockchain has a very broad prospect in data preservation and data sharing. In the future, we will introduce the smart contract method to the blockchain, design a more detailed permission control structure, and combine the traditional cryptography method to better protect the privacy information of patients.

6. REFERENCES

[1] Charles D, Gabriel M, Furukawa M F. Adoption of electronic health record systems among US non-federal acute care hospitals: 2008-2012[J]. ONC data brief, 2013, 9: 1-9.

[2] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system[J]. 2008.

[3] Ivan D. Moving toward a blockchain-based method for the secure storage of patient records[C]//ONC/NIST Use of Blockchain for Healthcare and Research Workshop. Gaithersburg, Maryland, United States: ONC/NIST. 2016.

[4] Kuo T T, Hsu C N, Ohno-Machado L M C. Decentralized Privacy-Preserving Healthcare Predictive Modeling Framework on Private Blockchain Networks[C]//ONC/NIST Use of Blockchain for Healthcare and Research Workshop. Gaithersburg, Maryland, United States: ONC/NIST. 2016.

[5] Xue T F, Fu Q C, Wang C, et al. A Medical Data Sharing Model via Blockchain[J]. Acta Automatica Sinica, 2017, 43(9):1555-1562.

[6] Ackerman A, Chang A, Diakun-Thibault N, et al. Blockchain and Health it: Algorithms, Privacy and Data[J]. 2016.

[7] Ekblaw A, Azaria A, Halamka J D, et al. A Case Study for Blockchain in Healthcare:“MedRec” prototype for electronic health records and medical research data[C]//Proceedings of IEEE open & big data conference. 2016, 13: 13.

[8] arayanan A, Bonneau J, Felten E, et al. Bitcoin and cryptocurrency technologies: a comprehensive introduction[M]. Princeton University Press, 2016.

[9] Yong Y, Feiyue W. Development status and prospect of blockchain technology[J]. J. Autom, 2016, 42(4): 481-494.

[10] Merkle R C. Protocols for public key cryptosystems[C]//Security and Privacy, 1980 IEEE Symposium on. IEEE, 1980: 122-122.

[11] Datamasker.Data Masking:What You Need to Know[J].A Net 2000 Ltd. White Paper,2016.

[12] Choudry B.Masking the Data on Cloud[J].International Journal of Advances in Computing,2012,1(04):388-390.

[13] Larimer D. Delegated proof-of-stake white paper [Online]available: http://www.bts.hk/dpos-baipishu.html, 2014