

Research on electronic medical record access control based on blockchain

Yongbin Zhang¹, Meng Cui¹, Lijuan Zheng¹ , Rui Zhang^{2,3}, Lili Meng⁴, Dong Gao¹ and Yu Zhang¹

Abstract

For the medical industry, there are problems such as poor sharing of medical data, tampering, and leakage of private data. In view of these problems, a blockchain-based electronic medical record access control research scheme based on the role-based access control model is proposed in this article. First, the appropriate access control strategy is adopted to solve the leakage problem of the user's medical privacy information during the access process. Then, the information entropy technology is used to quantify the medical data, so that the medical data can be effectively and maximally utilized. Using the distributed general ledger characteristics of blockchain and its inherent security attributes, data islands can be eliminated, data sharing among medical systems can be promoted, access records can be prevented from being tampered with, and medical research and precise medical treatment can be better supported. Through this research, not only can user's medical privacy information protection be realized during the service process but also patients can manage their own medical data autonomously, which is beneficial to privacy protection under the medical data sharing.

Keywords

Medical data, access control, blockchain, privacy protection

Date received: 5 July 2019; accepted: 15 October 2019

Handling Editor: Liran Ma

Introduction

The rapid development and wide application of computer and information technology have brought about the informationization reform of medical institutions. The extensive use of electronic medical records (EMRs)¹ has brought great convenience to the medical field. It is said that a condition may accompany these various characteristics during the examination. When the doctor diagnoses the disease, it is common practice to ask the patients historical illness, physical condition, and the like. There are two disadvantages to this approach: (1) it is difficult to ensure that patients can accurately remember quantitative values of historical conditions, such as blood pressure history records. (2) Patients often have non-professional medical terms when describing the condition, which will affect the doctor's understanding of the patient's historical condition. Therefore, a precise and accurate medical record

file will undoubtedly provide a more reliable reference for a doctor. The application of medical big data is more and more extensive, and the accompanying problems are very prominent too. The most concern issue is the privacy. According to the 2015 China website security report released by 360 company, the number of medical privacy disclosure incidents is second only to the Internet privacy information disclosure incident, and the amount of personal privacy leakage caused by

¹School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, China

²University of Chinese Academy of Sciences, Beijing, China

³Georgia Institute of Technology, Atlanta, GA, USA

⁴Shandong Normal University, Jinan, China

Corresponding author:

Lijuan Zheng, School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, Hebei, China.
Email: zhenglijuan@stdu.edu.cn



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<http://www.creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work

without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

medical website vulnerabilities is the largest.² It can be seen that the harm of medical data leakage is very great. The disclosure of medical data is not only caused by external risks of medical institutions but also caused by internal leakage. Medical institutions are careless about data management, and access rights are not clear. Individuals who sacrifice the patient's private information in exchange for rich illegal interests sometimes occur and cannot trace the problem when privacy leaks occur.³ In February 2016, the Ohio Mental Health and Addiction Service Center mailed postcards of the words "Voice of Patients" to patients who had previously visited the clinic, inviting them to do online surveys, which is equivalent to telling them that they have been in the service center. In addition, the postcard contains basic information about the patient, such as name, address, and gender. The handlers can get basic information about the patient in the postcard and then contact the medical information of the Ohio Mental Health and Addiction Service Center to infer the patient's private information. Then, it causes the patient's private information to leak. In the 2018, there were 18 data breaches involving more than 100,000 medical records in the United States. Eight of these accidents even affected more than 500,000 medical records, and three more violations caused more than 1 million health care records to be accidentally exposed. Among them, Med Associates, a medical billing company based in Rasham, New York, is responsible for providing claims for more than 70 health care providers. They found that an employee's computer was accessed by an unauthorized individual, and the attacker could get personal medical information of up to 270,057 patients. Therefore, protecting the privacy protection of EMR sensitive data is a research hotspot and research trend.⁴ In addition, patients are not involved in the management of their own medical information—the patient may be unaware of who uses their data and for what purpose. In the traditional medical database system, the administrator can modify the access record of the medical data, which leads to the failure of the data, and it is impossible to determine whose the data are and when data are leaked, and it cannot be accurately accounted for. While improving privacy protection, it tends to reduce the utilization of medical data and fails to realize the sharing of medical data.

As an important means of protecting data security, access control manages user rights, so that legitimate users can only access corresponding data in the system according to the rights they have, and prohibit users from unauthorized access to data, thus ensuring data security and the normal operation of the business system. In today's era, access control technology will continue to be an important means of protecting data security and controllable sharing. However, as

management scenarios and security requirements become more complex, access control authorization management difficulty increases, access control object description difficulty increases, access control is difficult for personal privacy protection in data objects, and tamper resistance is poor. In this article, information entropy and blockchain are combined with access control. Information entropy can be used to quantify private information to prevent visitors from mastering too much medical data. The sharing and non-destructive modification of blockchain can well solve the problems of access control.

Research status

At present, various medical privacy protection technologies are constantly being updated. In terms of access control,⁵ the privacy protection research method based on access control mainly designs a secure authentication algorithm to restrict access rights of the access subject to the EMR system. Zhu et al.⁶ proposed a user-friendly and easy-to-manage attribute-based access control (ABAC) mechanism for cloud storage services in 2015. This mechanism defines priorities for attributes and refining the granularity of data access control in cloud environments. Somchart and Hiroyuki integrated ABAC, role-based access control (RBAC), symmetric encryption technology, and ciphertext policy attribute-based encryption system (CP-ABE) and proposed a new access control model C-CP-ARBE.⁷ The model defines the access control policy as a tree. It ensures security by continuously computing and distributing keys in the tree. The model not only implements fine-grained access control, but its efficiency is also much higher than that of traditional ABAC. Belaazi et al.⁸ proposed an ontology-based privacy protection access control architecture, which uses the self-established privacy ontology to verify the access control policy. The architecture performs redundancy elimination and consistency check on the policy. Imran-Daud et al.⁹ designed an ontology-based access control system, which is a combination of ABAC and ontology technology. It overcomes the problem of low interoperability between components in a distributed environment. These studies have proved that ontology technology can be applied in the distributed environment of big data to unify the description of roles, attributes, and so on. However, the current research in this area only focuses on the description of uniformity and does not consider how to further quantify the degree of privacy data leakage. These studies focus on how to refine the granularity of access control, improve efficiency, and adapt to the current big data environment. Little research has been done with privacy protection as the focus of discussion. According to the different degrees

of patients' needs for privacy protection of their own information, Hsu and Pan¹⁰ proposed an RBAC for the problem of increasing user authorization difficulty in the medical information system as the number of users and the amount of information increased. The method is capable of supporting authorization for different kinds of objects and a new authorization domain. On the basis of the RBAC model, Huo and Wu proposed a patient-oriented privacy protection access control (POP-PAC) model. In this model, users can define access control policies that match their own preferences according to their own needs and can solve the problem of passive leakage of privacy data.¹¹ However, the model does not make a detailed distinction between the patient's data. While the doctor obtains the patient's medical record, it is possible to obtain case information unrelated to the case and even obtain basic information of the patient. Shin et al.¹² proposed a personalized medical service platform based on RBAC for smart devices to intelligently manage personal health records. Hui et al.¹³ proposed a risk adaptive access control model for medical big data. The model can be dynamically controlled and meet certain data utilization needs. However, this method only considers the complexity of the doctor's access to the data, does not make a careful distinction between the patient's data, and does not consider the value of the data itself in the process of utilization. Chen and Lin¹⁴ proposed a new authorization access control model, which stores the patient's data according to the privacy level and obtains corresponding information according to different authorization modes. The privacy level is set according to the specific situation. However, this model only addresses the medical information access control problem of legally authorized users and does not involve other types of medical information leakage and security protection issues.

The above medical access control model can protect patient privacy to a certain extent, but they are poor in medical data interoperability and data are too concentrated. Centralized databases bring increased risk, increased costs, and limited node expansion. Once the central point has a problem or crash, not only will all nodes be unusable but also central data have a risk of disclosure. Moreover, once the central data are leaked, the disclosure of private information will be devastating. The database used in traditional medical systems can tamper with data and has poor traceability. Blockchain technology¹⁵ can effectively solve these problems.

Blockchain is a distributed, decentralized network database that emerges with the emergence of cryptocurrencies.¹⁶ The blockchain uses a time-stamped chained block structure to store data, adding a time dimension to the data. Each transaction on the block is cryptographically associated with two adjacent blocks, so any transaction is traceable.¹⁷ And, the blockchain stores

all transaction data since the system is running. Based on these non-tamperable log-type data, all historical operations can be easily restored and traced.¹⁸ Xue et al.¹⁹ proposed an electronic medical information sharing model based on blockchain technology, which helps to solve the problem of information sharing difficulties between medical units. Shae and Tsai²⁰ proposed a blockchain platform architecture to help medical clinical trials and precision medicine. Ivan²¹ proposed that make blockchain as a novel approach to protecting medical health data storage. It analyzes implementation barriers and a gradual transition from current technology to blockchain solutions. Azaria et al. proposed a medical data acquisition and rights management system based on blockchain technology by combining the OPAL/Enigma encryption platform of the Massachusetts Institute of Technology with blockchain technology.²² Kuo and Ohno-Machado²³ adopted a combination of privacy protection online machine learning and private blockchain technology. Witchey and Nicholas²⁴ introduced the medical transaction verification system and method. Xia et al.²⁵ believe that the patient's medical records may face various risks such as privacy leakage and economic loss in the process of transmission. To address these issues, Q Xia proposes a solution for sharing medical big data in a weak trust environment hosting problem. The system is based on blockchain and provides data traceability, data auditing, and shared medical data management and control. Study by Dubovitskaya et al.²⁶ is also based on the traceability of blockchain, and proposed a safe and trustworthy medical electronic record system. Ahram et al.²⁷ introduced a health chain-based health care application based on blockchain. The disadvantage of the various scenarios or solutions described above is that the medical data cannot be updated to the blockchain in time, and it is necessary to pay a certain amount of compensation and the cost is high.

Information entropy is an effective tool for measuring information. Information volume can be represented by information entropy. Privacy information can also be measured by information entropy.²⁸ Information entropy has many applications in location privacy protection and data anonymization. According to the traditional access control model, the method of information entropy used for access control of private information is relatively mature. In the access control of private information, the system intuitively understands the amount of private information held by the visitor to assist in the formulation of policies and the execution of decisions. Y Liu et al. proposed a privacy-based data access control and medical document sharing mechanism. Information entropy is used to calculate private information. It can identify an integrated mode with a large amount of information and uses the integrated mode to query distributed medical

documents.²⁹ However, privacy information has different sensitivity levels. In the process of data utilization, the use of data is often limited by the privacy information protection requirements, which makes the utilization rate greatly reduced.

Blockchain technology

Blockchain technology is a decentralized recording technique. The data on the blockchain can be maintained by all the nodes participating in the system. Each participating node can belong to different organizations and does not need mutual trust between nodes, but each node can obtain a complete record. Permissioned Blockchain, also known as Consortium Blockchain or Federated Blockchain, is a sort of private blockchain which allows participants to restrict the participation of other members. So, the autonomy is transferred from one in charge to more than one in charge. Only those predefined members have the right to manage certain actions at different levels. To improve security and responsiveness, this article uses the Permissioned Blockchain to build and store medical data, and access records on the Permissioned Blockchain.

Medical blockchain and medical block

Each user's EMR and data access record are stored in a separate chain. The medical blockchain is mainly composed of two parts: a block and a transaction. A blockchain consists of blocks that record the previous block ID, and each block contains several trade orders. These transactions are the carriers that store the blockchain data. For example, a blockchain just as a database. Each block that makes up the blockchain can be thought as a table in the database. The transaction can be regarded as a record of each table record. The added block contains the hash value of the previous block and the start owns the basic information about the patient, so the medical data obtained by the patient after each treatment is only required to link to the previous block. The structure is shown in Figure 1.

A block is mainly composed of block header and others.³⁰ The block header contains the ID of the previous block, the public key of the block generator, the Merkle root hash value generated by the transaction ID, and the time slice of the generated block. The content outside the block header includes the digital signature of the block generator for the block header, number of trade orders ID, and all the transaction order IDs stored in the block. The digital signature is to ensure that the block content is not tampered with, and to ensure that the block generator cannot deny after generating the malicious block. And, the block only saves the ID of the transaction order, that is, not the transaction order itself is saved but the index

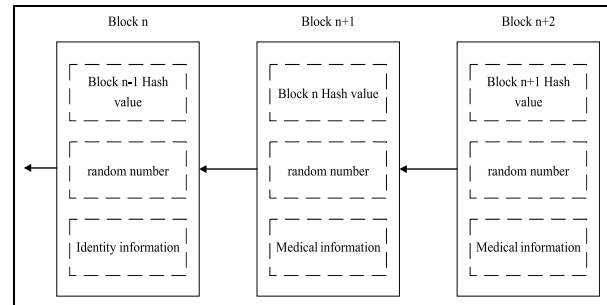


Figure 1. Chain structure of blockchain.

pointing to a certain transaction order, so that the capacity of each block can be reduced, which is convenient for synchronization and backup. Blocks and transaction orders are physically stored in the database and stored in the form of blockchains logically. On the transaction ticket design storage, just the transaction order ID, the transaction order type, the time stamp, the public key, the digital signature, and other transaction order field information are added to the data normally stored in the database, and the information to be stored is used as the transaction ticket content, forming a logical transaction order whose physical storage is not much different from general data storage.

Consensus algorithm

The practical Byzantine fault tolerance (PBFT) algorithm is used as the consensus algorithm in the medical blockchain because the PBFT algorithm is a consensus algorithm for the alliance chain. Its advantages and advantages are as follows:

1. The PBFT algorithm does not need a lot of computational power to avoid the "51% attack" like the POW algorithm, nor does it need to rely on tokens as a measure of voting rights like the POS algorithm or the Delegated Proof of Stake (DPOS) algorithm, it can allow less than $n - 1/3$ nodes in the system to go wrong (data loss, no work, etc.).
2. As a Byzantine fault tolerance (BFT) algorithm, the PBFT algorithm has less than or equal to $n - 1/3$ faults or malicious nodes in the system to ensure the normal execution of a distributed consensus process. This requires nodes in the system using the PBFT algorithm. There are at least $2n + 1/3$ normal nodes in each consensus process, so the environment in which these nodes operate must be relatively safe and stable.
3. The medical blockchain is a kind of alliance chain. The entities involved in the medical

blockchain have government endorsements, have certain credibility, and are strictly supervised by the health management department. The situation of malicious acts is far less than the blockchain such as Bitcoin system. At the same time, after years of information development, each hospital has a relatively complete network, server, and database system. Therefore, the existing medical system can provide a relatively safe and stable operating environment for the PBFT algorithm to operate normally. At the same time, there is no high voting rights because each node in the cluster running the PBFT algorithm has equal status. Therefore, the centralization of the transaction block or blockchain is avoided by the medical blockchain system, so the PBFT algorithm is very suitable for the medical blockchain.

Smart contract

Smart contract is a computer trading agreement that does not require intermediation, self-validation, or automatic execution of contract terms. In recent years, attention has been paid to the increasing popularity of blockchain technology. Smart contracts on the blockchain are decentralized, trusted, programmable, and non-tamperable. They can flexibly embed various data and protocols to help achieve safe and efficient information exchange, value transfer, and property management.

Smart contracts generally have two attributes, value and state. The code uses the If-Then and What-If statements to preset the corresponding trigger scenarios and response rules for the contract terms. The intelligent contract is submitted by the multi-party agreement, signed by each user and then submitted by the user. After being transmitted by the P2P network and verified by the miner, it is stored in a specific block of the blockchain. After the user receives the returned contract address and contract interface, the user can pass the information. Initiate a transaction to call the contract. The miners are motivated by the system's default incentives, which will contribute their own calculations to verify the transaction. The miner receives the contract creation or invoking the transaction and creates a contract or execution contract code in the local sandbox execution environment (such as the Ethereum Virtual Machine). The contract code automatically determines whether the current scene meets the contract trigger condition based on the trusted external data source (also known as the oracle) and the world state check information to strictly enforce the response rules and update the world state. After the transaction verification is valid, it is packaged into a new data block. The

new block is linked to the blockchain main chain after being authenticated by the consensus algorithm, and all updates take effect.

Program design and workflow

The current status of medical informatization shows that the traditional electronic data model has the following problems in medical information storage: (1) electronic data are poorly standardized and has low liquidity. (2) Electronic data security and privacy are vulnerable to threats. (3) Electronic data and access records are tamperable. Aiming at the problem of information storage security and sharing between various medical information systems, combined with the blockchain consensus mechanism, encryption mechanism, peer-to-peer network, and other technologies, this article applies the blockchain to solve the medical privacy problem in access control. This program has the characteristics of immutability and decentralization.

This research has two main purposes. First, it solves the protection of user medical privacy information during the use of EMRs. The other is to effectively ensure the effective and maximum utilization of medical data based on the protection of medical privacy information. This article introduces the concept of information entropy, and combines access control technology and blockchain technology to design this model. First, develop a dynamic access control strategy to assign authority to patients and other staff. Write access control policies to smart contracts to complete identity authentication for data visitors. Then, the patient's medical data are stored in chunks and deployed in a decentralized network. The scheme uses the theory of information entropy to quantify the information. According to the patient's requirements, scheme sets the conditions for accessing medical data. The first point is that the purpose of the visit is consistent with the purpose of intention. The second point is to quantify medical information and set the information tolerance, and the accessed information needs to be less than the information tolerance. If the purpose of the visit by the doctor or other data visitor does not match the intended purpose of the patient, or the amount of information is greater than the tolerance of the information set, the data access is not allowed, but the application can be made again. When the patient agrees, it will be allowed to view the medical data stored on the corresponding block. The information that is solved in this way is poorly interactive, the flexibility is not high, and the patient cannot participate in the management of the data. And, each time the data are accessed, the access behavior is recorded and stored in the blockchain, and the access record cannot be falsified. Thus, it solves the problem that tampering data existing in the

traditional database, tampering with the record, and so on, which increases the intensity of privacy protection.

Information entropy for quantitative processing of privacy data

Medical data relates to health data and non-health data. Health data are data onto the physical condition of the individual, such as medical information, and non-health data are information that is not directly related to the health of the individual. Different data have different degrees of privacy relative to patients. After the visitor obtains medical data, the use value of different private information is different. At this time, the amount of private information obtained by the visitor cannot be measured by the amount of information. Different weights should be set for different privacy information, and different proportions should be set according to the stakes in the patient's private information.

In this study, patient privacy information is divided into three levels according to the sensitivity of privacy protection. The sensitivity of the three types of privacy information is different. The first type of privacy information has the highest sensitivity, and the corresponding weight should be the largest. The weights of the second and third types of private information are sequentially reduced. The weight set can be set according to different patients, but the weights are equal to 1.

The three types of privacy information protection weight settings are shown in Table 1.

The first type of private information includes basic information of the patient, such as name, ID number, address, and contact information, and the level is expressed as L1; the second type of private information includes the patient's medical record, and the level is expressed as L2; the third type of private information includes patient detection and laboratory data, and grade is expressed as L3.

The first type of private information is the information that is directional to the patient. Such information will require a higher degree of privacy protection. The second type of private information is the patient's medical record, which is related to the diagnosis and treatment of the disease, including the history of the patient's illness, illness, and treatment. The third type of private information includes the test records of the

patients. These kinds of information are pure medical data. They are not directional to the patient, but they are helpful for the analysis and diagnosis of the disease. They have research value and do not require high-level privacy protection sensitivity.

The first type of private information weight is defined as q_1 , the second type of private information weight is defined as q_2 , and the third type of private information weight is defined as q_3 .

The access information is defined as $\text{access} = \text{id}, a_1, a_2, a_3, \dots, a_n$; a_i is the access information entry, and the number of access information is n . When the weight is not included in the calculation, the amount of information for each request is calculated according to the definition of entropy as follows

$$E_{a1} = -\frac{1}{n} \log_2 \frac{1}{n} \quad (1)$$

$$E_s = -q_1 \sum_{i=1}^{s_1} \frac{1}{n} \log_2 \frac{1}{n} - q_2 \sum_{j=1}^{s_2} \frac{1}{n} \log_2 \frac{1}{n} - q_3 \sum_{k=1}^{s_3} \frac{1}{n} \log_2 \frac{1}{n} \quad (2)$$

E_{a1} is the amount of information corresponding to the level in the access request; a_1 represents an access request; $1/n$ is the proportion of individual patient privacy information in all EMR privacy information entries. E_s is the amount of information that will be obtained for the entire access request. After calculating the amount of information of each access information item, it classifies according to its private information, calculates various types of private information entropy, and calculates the entire access according to the weight of each type of private information. The amount of information that the entire access request will receive is then calculated based on the weight of each privacy message. The system's information tolerance for each access request is set to E_t , which can be set on a case-by-case basis depending on the system.

Access control strategy

In order to achieve fine-grained privacy protection access control for data, user role designs and authority allocation are performed based on the RBAC model. The patient sets the intentional purpose of the data. When the visitor accesses the data, the visitor needs to indicate the purpose of the visit, and then compares it with the intended purpose.

Visitors to the EMR system have diverse identities and different needs. Different levels of access rights need to be set according to different visitors. The hospital's medical information system not only has internal department visits but also has external medical insurance interfaces, community health service interfaces, and telemedicine consultation system interfaces.

Table 1. Privacy information weight setting table.

Privacy information category	Weight
L1	$q_1 = 0.6$
L2	$q_2 = 0.3$
L3	$q_3 = 0.1$

Different visitors have different needs for medical information. Patients should have full access to their own EMRs, without restrictions. Doctors access medical information mainly using medical information to help medical diagnosis and medical research. Access to medical information should have certain privacy restrictions. The data administrator manages the medical data, has great authority on data operations, and human intervention for other people's access rights to meet the special needs of medical access events, but the administrator should perform data reading privacy protection, so that medical information content viewing is restricted. External visitor access has very low privileges and gets less medical information.

Medical data access behavior refers to operations such as querying, processing, and utilizing medical information. The main operational objects are the patient, the doctor, the external visitor, and the data manager. According to different visitors, different behaviors are divided, and different actors are assigned different permissions. Prevent visitors from having unauthorized access. The user has full access to his or her medical information. The doctor's private information about the patient is mainly written, queried, and modified. The data administrator mainly assigns the authority of other users and manages the data system. The data are encrypted.

The established access control policy is implemented with smart contracts, so that no third party is required to verify the identity of the visitor. Only when the visitor meets the requirements and passes the identity authentication can the request for access to the medical information be made.

Access control method: when a visitor wants to access medical data, the following rules are required to implement access control:

1. The visitor authenticates, and if the authentication fails, the visit ends; if the authentication passes, submits a medical information access request.
2. Receiving an access request, extracting a patient id and a specific request entry ai ;
3. Sorting the request items according to L1, L2, and L3; then, recording the number of various private information items s_1 , s_2 , and s_3 ;
4. Calculating the access request information entropy E_s ;
5. The purpose of the visit and the purpose of the intention, E_s and E_t , two-to-two comparison, if the purpose of the visit is consistent with the intended purpose, and $E_s < E_t$, then access is allowed; if the purpose of the visit is different from the intended purpose, or $E_s > E_t$, access is not allowed.

Overall design

The establishment of this blockchain is mainly used to solve several problems such as dispersal of medical data, slow access, poor interoperability between data, the need to improve the quality and quantity of data in medical research, patients lack ownership and management of their own data, and patients do not participate in the management of their own medical data. Doctors may obtain additional information about patients not related to this treatment when accessing data. The patient's medical records are placed in the blockchain, and the patient's medical information is stored in the form of ciphertext. The EMR administrator cannot obtain the patient's plaintext, and the database is completely invisible to the patient's private information. Each patient's medical record is scored, and then split, such as basic information of the user, medical diagnosis, medical report, and medical experiment data, for block storage. When a piece of data needs to be accessed, after obtaining the consent, only the data of the block are taken, which improves the data security. Through the smart contracts on the Ethereum blockchain, we record the relationship between patients and providers, which correlate medical records with viewing permissions and data for execution on the database. Records on the blockchain use cryptographic hashing to prevent tampering, thereby tracking data integrity. The database administrator can add new records associated with a particular patient, and the patient can authorize sharing of records between the visitors. The party receiving the new message will receive an automatic notification and can verify the record before accepting or rejecting the data. This allows participants to understand the situation and participate in the evolution of their records. Map an already existing and widely used id (e.g. name or user account) form to one of the person's addresses. After confirming the authority, data exchange between the patient database information and the visitor is performed. Figure 2 shows general picture of the study, which is divided into two parts for detailed introduction. Part of it is to add medical data to patients, detailing how to protect the privacy of patients' medical data during the data storage phase. The other part is the interaction process between data, describing the techniques and specific steps used in the interaction of data. Next, we will introduce the functions and principles of the specific modules one by one.

Adding records to patients. In the initial treatment of patients, it is necessary to store the user's basic data, medical diagnosis, medical reports, and other data. This work is handled by the EMR manager.

Steps 1–4 in Figure 2 show store each medical record of the user in blocks according to the sensitivity of

privacy protection. Use asymmetric cryptography to encrypt private information with the patient's public key. When there are many processes, they can be placed in the local database for caching. After waiting for data storage, delete the information of the local database. If the patient's medical record information has a large amount of information and the degree of privacy is not high, an index can be established on the blockchain without storing the information on the blockchain.

Interaction of privacy information. This section contains two nodes, the patient node and the medical information access node. The patient node belongs to the user end; the medical information access node can be considered as a data demand provider, can be a medical institution, and so on, and the database is a local database. This part mainly realizes the process of the patient's request and the privacy of the corresponding information submitted by the visitor, and selectively extracts the information of the corresponding block from the blockchain and returns it to the visitor. Step 8 in Figure 2 shows that when the visitor submits the access request, the access control policy formulated by

the smart contract is triggered to authenticate the identity. If the authentication does not pass, the access request cannot be made. If the authentication is passed, the patient information request required by the visitor is sent to the EMR manager of the access node.

After receiving the request, Step 9 shows that the EMR manager first checks whether the corresponding storage content is stored in the local database. There are three situations at this time: existence, partial existence, and non-existence. The EMR manager needs to modify the request corresponding to the content to be correct and needs to be updated. If there is no such content, the request modification is not required, and the subsequent operation can be continued. (Because the request may involve part of the patient's information, it is necessary to use the patient's public key for encryption to prevent it from being compromised.)

Step 10 shows that the EMR manager of the patient node sends the patient's public key to the EMR manager of the access node.

Step 11 shows that the EMR manager transmits the data request encrypted by the patient public key and the public key of the visitor to the patient node.

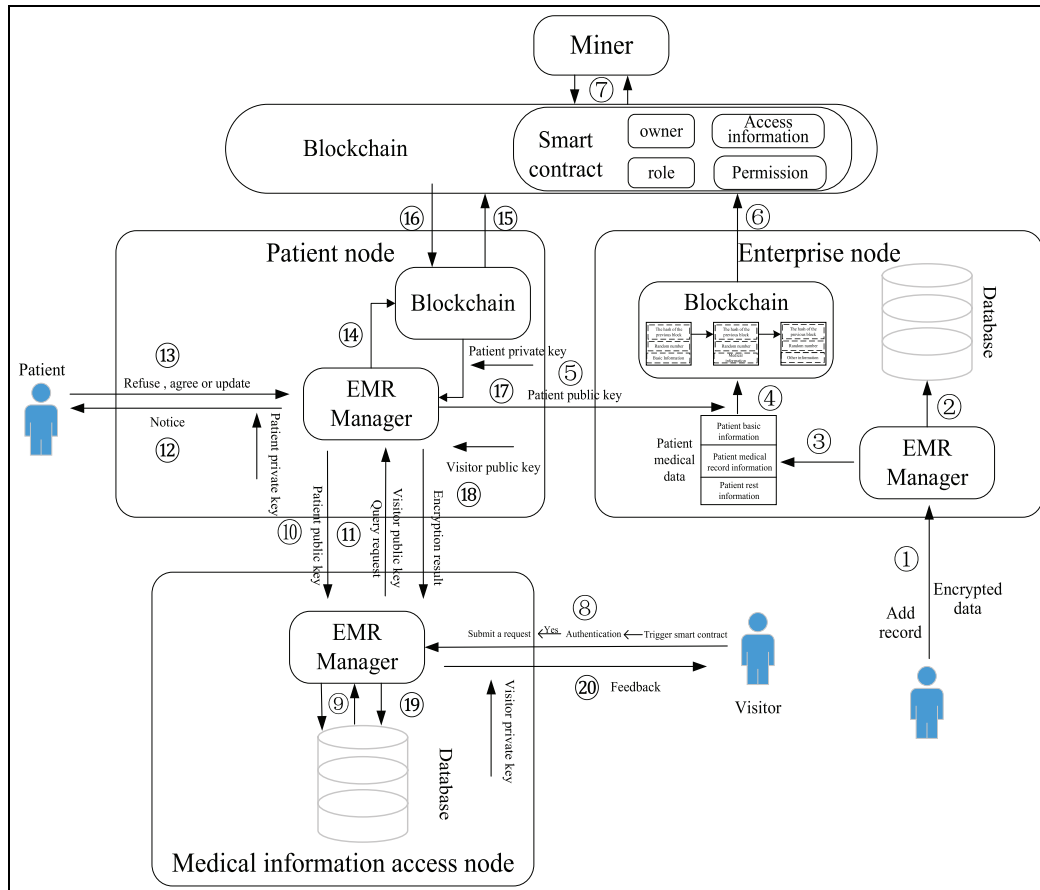


Figure 2. Design picture and workflow.

Steps 12 and 13 are to judge the data request sent by the patient to the EMR manager and determine the content request shared by the visitor in combination with the privacy degree of the corresponding information, and send it to the EMR manager. Since the data request was previously encrypted with the patient's public key, it is necessary to decrypt the patient's private key before reviewing it.

Step 13 shows that when the EMR manager receives the request confirmed by the patient, the part agreed to in the request and the part to be updated are sent to the system. If there are no consents or updates in the received requests, then Steps 14 to 17 are not required.

The system implements all the functions required to join and participate in the blockchain network. This can handle a large number of tasks, such as connecting to a peer-to-peer network, encoding and sending transactions, and maintaining a verified local copy of the blockchain. Steps 15 and 16 show that the Ethereum client obtains the transaction key according to the hash, block height, and block hash of the required information block, thereby querying relevant information.

Step 17 shows that the information obtained by the system from the blockchain is the data encrypted by the patient's public key, so it needs to be decrypted by the patient's private key to obtain the plaintext and sent to the EMR manager.

Step 18 shows that the EMR manager of the patient node. When sending the information required by the visitor or the result of the complete rejection of the patient to the EMR manager of the access node, it needs to be encrypted with the public key of the previously received visitor to achieve the purpose of secure transmission. Step 19 shows that the EMR manager of the medical information access node first stores the encrypted result information and then stores it in the local database, and keeps the backup. Step 20 shows the encrypted result received is first decrypted by the visitor key and then sent to the visitor.

Performance analysis

Example analysis

When the user requests data, the user strictly follows the access control method. Next, the model will be explained in conjunction with the corresponding examples:

Example 1. Dermatologist Cary proposes to visit the patients Bob's bronchitis (respiratory medicine) examination data for the purpose of treatment. First of all, Cary wants to carry out identity authentication. For the established access control strategy, the identification of doctor Cary is not passed, and

there is no corresponding authority to view the medical data of patients in non-undergraduate departments. In this way, access control controls the user's identity and permissions to achieve data protection.

Example 2. Cardiovascular physician Mark proposed to visit the patient's Mary's cardiovascular disease medical data for the purpose of disease research, and the medical data of the patient's Mary cardiovascular disease is intended for treatment. doctor Mark passed the corresponding identity authentication, and the purpose of the access is inconsistent with the intended purpose. Therefore, it is not necessary to calculate the information entropy E_s of the access request, and the access request is rejected, and the medical data cannot be viewed. This reduces the risk of accessing data and prioritizes the patient's requirements; if doctor Mark believes that the disease study is a preventive effect on the disease and is conducive to the patient's health, then doctor Mark can apply again to explain the purpose of the visit. Then, E_s is calculated and compared it with E_t , and the result is sent together with the public key to the EMR administrator. After receiving the request, the EMR manager will notify the patient whether Mary agrees to the request. If the patient still refuses the access request, the doctor will not receive medical data and the visit will be terminated.

If the patient agrees to the request, the patient's gastrointestinal medical information will be decrypted and sent to the EMR administrator. The administrator will use Mark's public key for encryption and send it to Mark, as Decrypt your own private key to view the data. This enhances the flexibility and real-time interactivity of data access while protecting patient privacy.

Example 3. Clinical surgeon Jack proposed to visit the patient's gastrointestinal health status for the purpose of further treatment. First of all, the role of doctor Jack is the clinical attending doctor. The corresponding identity is authenticated. The purpose of the interview is the same as the intention. The E_s and E_t are compared. If $E_s < E_t$, access is allowed. Doctor Jack sends his request and public key to the EMR administrator, and after receiving the request, the EMR manager will notify the patient Tom that the patient Tom will decrypt his gastrointestinal medical information and send it to the EMR administrator. The administrator encrypts it with Jack's public key and sends it to Jack. The transfer of the information is completed. After each visit, the EMR administrator will record the access process and store it on the blockchain to resolve the issue of data corruption.

Security analysis

1. File storage security: the characteristics of the blockchain are used as a time stamp series of books. Once the consensus mechanism is confirmed, no modification is made. If an attacker wants to modify the data stored in the blockchain system, it must mimic a main chain like the source chain, which requires a lot of computation, which is almost impossible. In addition, the data stored in the blockchain is divided into block sequences and stored in the system. These data are obtained and spliced in a certain sequence order to form a source file. The probability is very low, and it is difficult to synthesize these files in order. It is assumed that the attacker can obtain the data saved by the patient in the system by some means, but the data will not be viewed, deleted, or modified; so, the data are safe.
2. Data anti-tampering: the encrypted file is stored in the blockchain. In the case where the patient's private key is not available, the file in time cannot be decrypted as the source file, that is, the attacker cannot view the patient's medical data to ensure patient privacy. Suppose an attacker can get a fragmented file by some means and stitch it together in a certain order to get the same file as the source file. The attacker wants to view the contents of the file and needs to pass the patient's private key to decrypt the file. With asymmetrically encrypted data files, it is difficult to decrypt without obtaining a private key.
3. Data theft prevention: the attacker attempts to replace the real file stored in the system with a fake file by some means. In the case where the source file exists, it is very difficult. In this article, a hash value repetitive check is required for a file that executes a smart contract, when the attacker intends to use a fake file F' to execute the smart contract, the hash obtained by the hash algorithm is $hashF'$. But, the hash obtained by the hash algorithm for the source file F to execute the smart contract is $hashF' \neq hashF$. According to the hash rule, the hash values obtained by hashing two files that are not identical are different, that is, $hashF'/hashF$. Such a fake file F' is not able to execute the contract. Therefore, it can ensure that the user's source file cannot be replaced by the fake file used by the attacker, thereby ensuring the security of the user's medical data file.
4. Privacy protection: transactions are signed by the initiator and stored on the block by verification,

so the privacy of the private key can guarantee the security of the transaction, and each transaction is attended anonymously. The access control protocol allows the patient to have absolute control over the medical record. Only authorized users can view the required medical record data without seeing sensitive information unrelated to this visit, so the access control protocol and private key are very secure. The privacy of medical records is well protected.

Comparative analysis

The comparative analysis of the existing medical blockchain system, the medical solution of Blockchain combined with access control, and this study is given: medical data sharing model via blockchain (MDSM),¹⁹ MedRec,³¹ ModelChain,³² literature [33],³³ and literature [34].³⁴ The results of the comparison with the existing solutions are shown in Table 2.

1. Compared to the literature 31–34, this study uses information entropy to quantify medical information, so that the information obtained by visitors has clear quantitative control within the system. The system knows the amount of private information that the visitors have and prevents visitors from inferring other information about patients through the privacy information they have. In addition, this article adopts an access control strategy to strictly authenticate the identity of the visitor and prevent the occurrence of unauthorized behavior, which is not available in the literature 31–34.
2. Compared to MDSM using the DPOS algorithm, the number of startup nodes required is much less than MDSM, and MDSM needs to artificially set whether each hospital has the right to vote and the proportion of votes in the final result.
3. Compared with MedRec using the POW algorithm, the number of nodes required to maintain the blockchain system is far less than that of MedRec, and it is not required to pay the consensus of the blockchain system to participate in the node remuneration, and does not require a lot of computational power to maintain the blockchain system.
4. ModelChain adopts the form of private blockchain, and the number of required nodes is uncertain. However, because the work proof consensus mechanism is vulnerable to “51% attack,” the node has the ability to successfully tamper with and forge blockchain data by

Table 2. Comparison of security performance.

System	Quantification of information	Dynamic access control	Number of nodes	Voting weight setting	Pay
MDSM	No	No	121	Yes	No
MedRec	No	No	Many	No	Yes
ModelChain	No	No	Many	No	Yes
Literature [34]	No	Yes	Many	No	Yes
Literature [35]	No	No	Less	No	No
This	Yes	Yes	Less, at least four	No	No

mastering more than 51% of the computing power of the entire network. Therefore, more nodes are needed to “average” the power to prevent such attacks. Therefore, compared to the ModelChain using the POI algorithm, there is no need to pay the consensus participation node reward, and the number of nodes required is also small, and the POI algorithm is based on the POW algorithm, so the required computing power is also large.

5. Compared with the literature 34–35, blockchain is combined with access control to solve the leakage problem and sharing problem of medical data, which protects the patient’s medical privacy to a certain extent. However, patients cannot manage their own medical information autonomously and do not quantify the medical data. When the visitor has certain data, the background attack can be used to obtain more private information of the patient. In this article, by grading and quantifying medical data, using information entropy to measure the amount of private information held by visitors to EMR information, the information obtained by visitors can be clearly quantified within the system. Preventing excessive access causes leakage of EMR privacy information.

Therefore, it can be seen that this medical blockchain system does not need to pay remuneration, requires fewer startup and running nodes, can be extended later, has less computing power, and does not need to artificially set the proportion of voting rights. Moreover, the privacy data can be quantified, the dynamic access control strategy can be formulated, and the effective management of the rights can be realized. These are the unique features and advantages of the solution.

This model first uses access control technology to divide the permissions and roles of patients, doctors, and other personnel, and set the first barrier for protecting medical data. Next, using the blockchain technology, the patient’s basic data, medical diagnosis, medical reports, and so on are stored in blocks. What

kind of case data does the visitor need when the next treatment is performed, and only the corresponding information is required on the premise of the patient’s consent. Block data, the data are more strictly managed to prevent doctors or other personnel from obtaining too much medical data for illegal operations. And, adding records to each interaction process, the traditional database system can modify the stored data and access records, which leads to the problem when the data leaks. Blockchains is immutable, and access to the records and data cannot be tampered with on the blockchain, thus effectively solving the problem.

Conclusion

Medical privacy data have always lacked interoperability and sharing. The management of medical data centering deprives patients of ownership of data, making it impossible for patients to participate in the management of their own data. The design of this program uses access control technology, information entropy technology, and blockchain technology to further enhance the protection of medical data, improve the integrity of data, promote the exchange of trusted data, and decentralized management of medical data, so that patients can control data sharing and improve privacy protection. Blocking the patient’s medical data can also effectively solve the data leakage problem in the data mining process. It only needs to analyze the data related to the condition, which can predict the disease and prevent it in advance, better helping patients stay healthy and make the entire medical industry flourish. The information on the blockchain is immutable, which is based on the security of the private key. If the private key is lost, the secure storage will not exist, and the next research focus will be around this issue.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors acknowledge the support from the National Key R&D Program of China under grant no. 2017YFB1400700; Hebei Education Department under grant no. QN2015231; National Natural Science Foundation of China under grant nos 61472414, 61772514, and 61602061; and Higher Education Scientific Research Project of Hebei Higher Learning Society under grant no. GJXH2017-250; and School-Level Graduate Innovation Funding Project under grant no. YC2019069.

ORCID iD

Lijuan Zheng  <https://orcid.org/0000-0002-9925-6836>

References

1. Fang S, Cai Z, Sun W, et al. Feature selection method based on class discriminative degree for intelligent medical diagnosis. *Comput Mater Continua* 2018; 55(3): 419–433.
2. Cai J, Zhang T and Zong W. Challenges and considerations in medical big data. *China J Health Inform Manag* 2013; 4: 292–295.
3. Bi D, Dong K, Luning X, et al. Analysis and prospects of health care big data industry. *Big Data Era* 2017; 4: 6–20.
4. Zhang J, Li H, Liu X, et al. On efficient and robust anonymization for privacy protection on massive streaming categorical information. *IEEE T Depend Secure Comput* 2017; 14(5): 507–520.
5. Wang M, Wang J, Guo L, et al. Inverted XML access control model based on ontology semantic dependency. *Comput Mater Continua* 2018; 55(3): 465–482.
6. Zhu Y, Huang D, Hu CJ, et al. From RBAC to ABAC: constructing flexible data access control for cloud storage services. *IEEE T Serv Comput* 2015; 8(4): 601–616.
7. Fugkeaw S, Sato H, Fugkeaw S, et al. Privacy-preserving access control model for big data cloud. In: *Proceedings of the 2015 international computer science and engineering conference (ICSEC)*, pp.1–6, Chiang Mai, Thailand, 23–26 November 2015. New York: IEEE.
8. Belaazi M, Boussi Rahmouni H and Bouhoula A. An ontology regulating privacy oriented access controls. In: *Proceedings of the international conference on risks and security of internet and systems*, pp.17–35, Mytilene, 20–22 July 2015. New York: Springer.
9. Imran-Daud M, Sanchez D and Viejo A. Ontology-based access control management: two use cases. In: *Proceedings of the 8th international conference on agents and artificial intelligence*, vol. 1, pp.244–249, Rome, 24–26 February 2016. Setúbal, Portugal: SciTePress.
10. Hsu WS and Pan JI. The secure authorization model for healthcare information system. *J Med Syst* 2013; 37(5): 9974.
11. Huo C and Wu Z. Patient-oriented privacy information access control model for medical information system. *Comput Appl Softw* 2014; 11: 75–77.
12. Shin MS, Jeon HS, Ju YW, et al. Constructing RBAC based security model in u-healthcare service platform. *Sci World J* 2015; 2015: 937914.
13. Hui Z, Li H, Zhang M, et al. Risk-adaptive access control model for medical big data I. *J Commun* 2015; 36(12): 190–199.
14. Chen Y and Lin Y. Research on authorized access control in medical information privacy protection. *China J Health Inform Manag* 2018; 15(3): 288–291.
15. Lin G, Liu B, Xiao P, et al. Phishing detection with image retrieval based on improved texton correlation descriptor. *CMC-Computers Materials Continua* 2018; 57(3): 533–547.
16. Delmolino K, Armet M, Kosba A, et al. Step by step towards creating a safe smart contract: lessons and insights from a cryptocurrency lab. In: *International conference on financial cryptography and data security*, pp.79–94, Christ Church, Barbados, 26 February 2016. Berlin, Heidelberg: Springer.
17. Liu A, Du X, Wang N, et al. Blockchain technology and its research progress in the field of information security. *J Software* 2018; 29(7): 2092–2115.
18. Shao Q, Jin C, Zhang Z, et al. Blockchain technology: architecture and progress. *Chin J Comput* 2018; 41(5): 969–988.
19. Xue TF, Fu QC, Wang C, et al. A medical data sharing model via blockchain. *J Automat* 2017; 43(9): 1555–1562.
20. Shae Z and Tsai JJP. On the design of a blockchain platform for clinical trial and precision medicine. In: *Proceedings of the 2017 IEEE 37th international conference on distributed computing systems (ICDCS)*, pp.1972–1980, Atlanta, GA, 5–8 June 2017. New York: IEEE.
21. Ivan D. Moving toward a blockchain-based method for the secure storage of patient records. *ONC/NIST*, 2016, https://www.healthit.gov/sites/default/files/9-16-drew_ivan_20160804_blockchain_for_healthcare_final.pdf
22. Shrier AA, Chang A, Diakun-Thibault N, et al. Blockchain and health IT: algorithms, privacy, and data. *ONC/NIST*, 2016, https://www.healthit.gov/sites/default/files/1-78-blockchainandhealthitalgorithmsprivacydata_whitepaper.pdf
23. Kuo TT and Ohno-Machado L. Modelchain: decentralized privacy-preserving healthcare predictive modeling framework on private blockchain networks. arXiv:1802.01746.
24. Withey NJ. *Healthcare transaction validation via blockchain proof-of-work, systems and methods*. US14/711,740 Patent, 2015.
25. Xia Q, Sifah EB, Asamoah KO, et al. MeDShare: trustless medical data sharing among cloud service providers via blockchain. *IEEE Access* 2017; 5: 14757–14767.
26. Dubovitskaya A, Xu Z, Ryu S, et al. Secure and trustable electronic medical records sharing using blockchain. *AMIA Annu Symp Proc* 2017; 2017: 650–659.
27. Ahram T, Sargolzaei A, Sargolzaei S, et al. Blockchain technology innovations. In: *Proceedings of the IEEE technology & engineering management conference (TEMS-CON)*, pp.137–141, Santa Clara, CA, 8–10 June 2017.
28. Peng C, Ding H, Zhu Y, et al. Information entropy model of privacy protection and its measurement method. *J Software* 2016; 27(8): 1891–1903.

29. Liu Y, Wang Z and Wang W. A purpose-based privacy-aware access control mechanism for distributed medical documents sharing. In: *Proceedings of the international symposium on information technology in medicine and education*, pp.636–640, Hakodate, Japan, 3–5 August 2012. New York: IEEE.
30. Jiang X, Liu M, Yang C, et al. A blockchain-based authentication protocol for WLAN mesh security access. *CMC-Comput Mater Continua* 2019; 58(1):45–59.
31. Azaria A, Ekblaw A, Vieira T, et al. MedRec: using blockchain for medical data access and permission management. In: *Proceedings of the 2nd international conference on open and big data (OBD)*, pp.25–30, Vienna, 22–24 August 2016. New York: IEEE.
32. Yuan Y and Wang FY. Blockchain: the state of the art and future trends. *Circuit Syst Signal Pr* 2016; 42(4): 481–494.
33. Liu Y, Du X and Wang N. Big data access control mechanism based on blockchain. *J Software* 2019; 1: 1–18.
34. Wang X, Jiang X and Li Y. Data access control and sharing model of application blockchain. *J Software* 2019; 30(6): 1661–1669.