Data Sharing Method for Heterogeneous Medical and Health Databases with Blockchain Technology

Lurui Wang, Zhongwei Jiang*, Yue Wang
Yamaguchi University, Yamaguchi, Japan

Abstract: The currently used medical and health systems have a problem with data islands. In this paper, a data sharing method for medical and health record system among heterogeneous databases is presented by applying the block-chain technology. Records are divided into the personal information part and health information part; Medical record part is stored in databases of institutions. Record indices are abstracted for each record and stored in the blockchain to link records together. Asymmetric encryption technology is applied to encrypt personal information and record indices. The medical information is encrypted by the user's private key, stored in a database; the personal information and record indices are encrypted by the user's public key and stored in the blockchain, all records are transformed into a unified form. JSON (JavaScript Object Notation) file was introduced as an intermedia to process records. Users can access all his records spread in different databases through blockchain with his private key. Experiments were carried out to show the feasibility and efficiency of this method.

***Key-Words**: Blockchain, Data Sharing, Asymmetric encryption, Distributed Database*

1. Introduction

Medical and health records usually stored across different organizations, when patients visited institutions, their records are scattered in different systems, and there is no unified way to access and share them. Blockchain is a tamper-proof, anonymous peer to peer network, where each node has a copy of the full ledger. The blockchain technology can be applied in the health and medical domain to provide a holistic, transparent, whole picture of scattered records [1]. This helps the patients to get a full picture of their health and sustain crucial trust in the medical system, meanwhile provides a secure method for users to protect privacy and share their records [2].

Some researchers have been done on blockchain technology to utilize its characteristics in the health and medical field to utilize its characteristics in the health and medical field. Asaph Azaria built a novel, decentralized system based on blockchain technology to deal with electronic medical records (EMRs). Patients have a comprehensive privilege to access to their EMRs that spread in various providers and organizations [3]. Qi Xia proposed a blockchain-based system to resolve the problem of medical data sharing among organizations in a trust-less environment, all actions have done to data were recorded in blockchain in universal format [4]. Alevtina Dubovitskaya proposed a framework based on blockchain technology for cancer patient care, to provide security and privacy-preserve access control over EMRs data [5]. Yi Chen designed a storage scheme to securely store EMRs on blockchain and cloud storage, and a service framework for sharing EMRs was introduced [6]. Hongyu Li proposed a block-chain based data preservation system to solve the EMRs sharing problem, users can store and share data with high security on blockchain framework [7]. Faisal Jamil proposed a novel drug supply chain management system based on Hyperledger Fabric and

Received: 2020/07/28, Accepted: 2020/08/17
*Corresponding author: Zhongwei Jiang
E-mail address:jiang@yamaguchi-u.ac.jp

blockchain technology, this system launched a smart contract to manipulate access control to electronic drug records and patient EMRs [8].

These researchers have studied the application of blockchain to EMRs systems from different aspects. However, most of the research use the blockchain as a functional module of the existing system or use blockchain to extend the system functionality, these systems are usually open to one institution, only the administrator can use them. Ordinary users do not have access to these data that spread in different institutions. User-oriented data sharing methods among institutions are of great significance to users and medical institutions. Not only can users have a comprehensive understanding of their situation, but it can also promote the sharing of data among different institutions and reduce repeated inspections and costs. In this paper, a new user-oriented blockchain-based data sharing method among different institutions is proposed. Users can access personal records stored in different databases through the semi-private blockchain with their private key. Heterogeneous records can be processed and retrieved in a unified format. User's privacy is protected by asymmetric encryption.

2. Method

The generation and storage of EMRs usually involve various stages and institutions. The flowchart of the system is shown in Fig.1. Users can record their daily health data with a smartphone or smart bracelet, these data usually stored in companies. When they go to a community hospital, doctors will examine their health and do the treatment, sometimes users need to be transported to a bigger hospital for further treatment, these data stored in hospital databases. Meanwhile, some of their tissue or sample may be sent to research institutions for further analysis, these data usually stored in research databases. Therefore, a user's EMRs usually separately stored in different databases. Databases are usually only open to particular administrators. In this paper user-oriented semi-private blockchain is built to link databases together and provide service for users. Each registered user or doctor is one node of the blockchain. Users can use their private key to retrieve all their records stored in a different database with the functionality of blockchain. All records were processed and presented in a universal form.

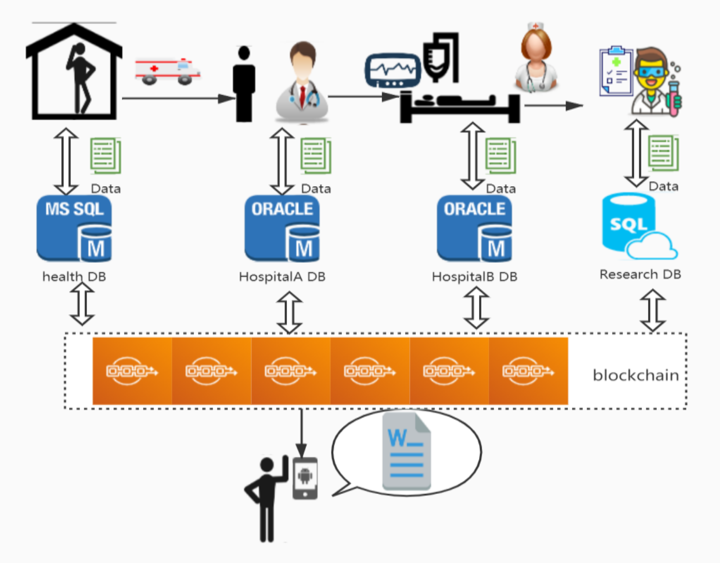


Fig.1 Flowchart of System

To facilitate users' retrieval and improve system security, the semi-private blockchain is implemented in this paper. It has the following advantages of using semi-private blockchain compared with the public blockchain:

(1) Launching a semi-private blockchain-based application most closely resembles how a company or organization runs a commercial website.

(2) Each node has been certificated so the risks of being attacked and business failure were low. This simplified implementation and deployment.

(3) The semi-private blockchain use PBFT (Practical Byzantine Fault Tolerance) algorithm as the consensus protocol. It consumes fewer resources to reach a consensus state than public blockchain.

2.1 Data Structure

EMRs can be divided into two parts: personal information and medical information. Personal information refers to the part of EMRs that involves personal privacy, such as name, age, and contact information. Medical information refers to the part of EMRs other than personal information, mainly including user symptoms, allergy records, medication records, etc. The data storage of the system is shown in Fig.2. In order to allow users to have the privilege to access their personal EMRs, personal information is stored in the blockchain, and medical information is still stored in the database of each institution. To link the personal information part and medical information part together, an index is abstracted as a pointer. The index consists of two parts: record address and the hash value of the record. Record address points to the access address of the record. The hash value is a verification code of record; it changes whenever the record is modified. Personal information and index are stored in the semi-private blockchain; medical information is stored in databases which spread in different organization. These two parts were linked together by the unique index.

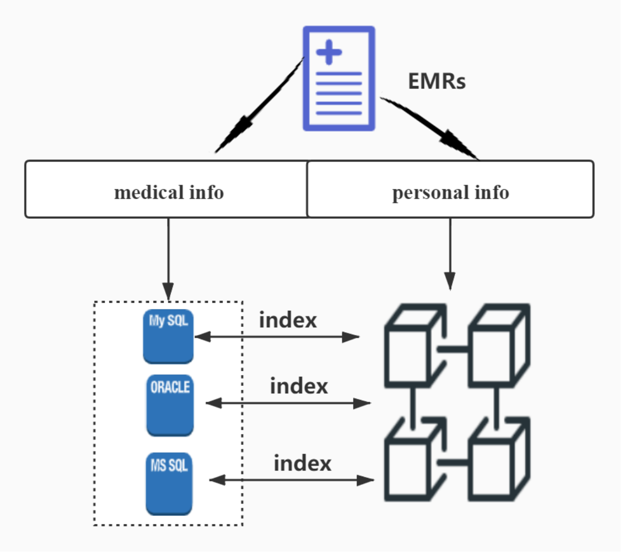


Fig.2 Data Storage of System

Each node is equal in semi-private blockchain, so the data on blockchain was open to every registered node. While EMRs contain privacy information, Asymmetric encryption is introduced to encrypt private information. Asymmetric encryption provides a pair of keys for each user: the public key and private key. One can be used as an encryption key and only the other key can decryption. Public key is stored in the database and open for administrator, medical information was encrypted by the user's private key, the administrator can decrypt medical information with the user's public key. The private key is kept secret by the user. Personal information is encrypted by the user's public key, only the user himself can decryption. The hash value of record will be checked to make sure records were untampered during this process.

2.2 Blockchain Structure

The structure of the block is represented by a list of blocks in a particular order. The structure of blocks is shown in Fig.3. Two vital data structures used in the blockchain are pointers and linked lists. Pointers are parameters that hold the information about the location of another variable, it points to the address of the previous block. Linked lists are a sequence of blocks where each block links to the previous block with the help of the pointer pointing to the former block. The header address was calculated by the hash algorithm to assure security. The main body of the block is personal information and record index. There is a bounty part at the bottom of each block. Users can use their bounty as a fortune to pay for advanced service.

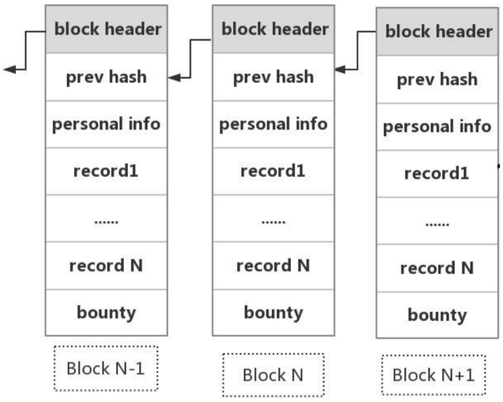


Fig.3 Blockchain Structure

3.2 Retrieving Process

The user can access all his medical records which were stored separately through a private key. The retrieving process is shown in Fig.4. These records are stored in the heterogeneous database with different data types. A transform method is introduced to process records in a unified format. Retrieved records are transformed into {key, value} pairs. SON is introduced as an intermedia. According to the characteristics of semi-private blockchain, the entire traversal process from the first block to the last block is required to retrieve the information stored in it. According to the user's private key, the traversal process finds the record belonging to the user and decrypt the record with the private key to obtain the index and the hash value of the record. After verifying that the hash value is correct, the records are query out from the database according to the index, the transform algorithm is used to process records into JSON.

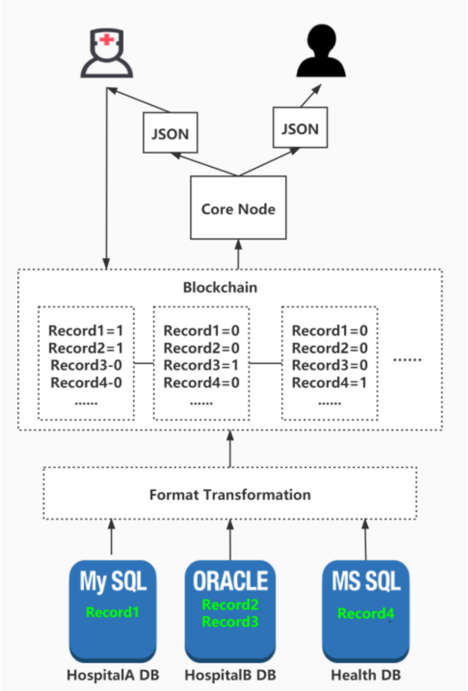


Fig.4 Record Retrieve Process

All record fields are changed into 5 types: empty, string, number, date, Boolean based on their original types. The date and timestamp are changed into yyyy/mm/dd format. The number type is changed into float type. The string type is changed into text type. The Boolean type is changed into True/False type. As different databases use different encoding methods, such as Shift_JIS and UFT-8, Unicode is used as a uniform encoding method to make sure all records were stored and showed correctly on the Internet without garbled.

4.Experiment and Result

Experiments are carried out on Hyperledger Fabric with Python 3.7.3. Hyperledger Fabric is an open-source enterprise-grade distributed Permissioned blockchain technology platform. It supports smart contracts authored in general-purpose programming languages.

Medical and health records are downloaded from the official website of the Department of Health & Human Services(USA) as a test dataset. For the sake of brevity, the records have been simplified. One user (IdentificationID=999-99-9999) is chosen as an example. This user has two records stored in two different databases. The records have a different structure and data format. The record one was created in a community hospital, the record two was created in a bigger hospital after the user was tested in the community hospital. The record one stored in the MySQL database was shown in table 1. Record two that stored in the PostgreSQL database is shown in table 2. The user uses the private key as a decryption key to decrypt personal and record indexes, then the indexes are used to get access to records. These two records were retrieved by index from the blockchain. The hash code is checked in this retrieving process to make sure records have not tampered.

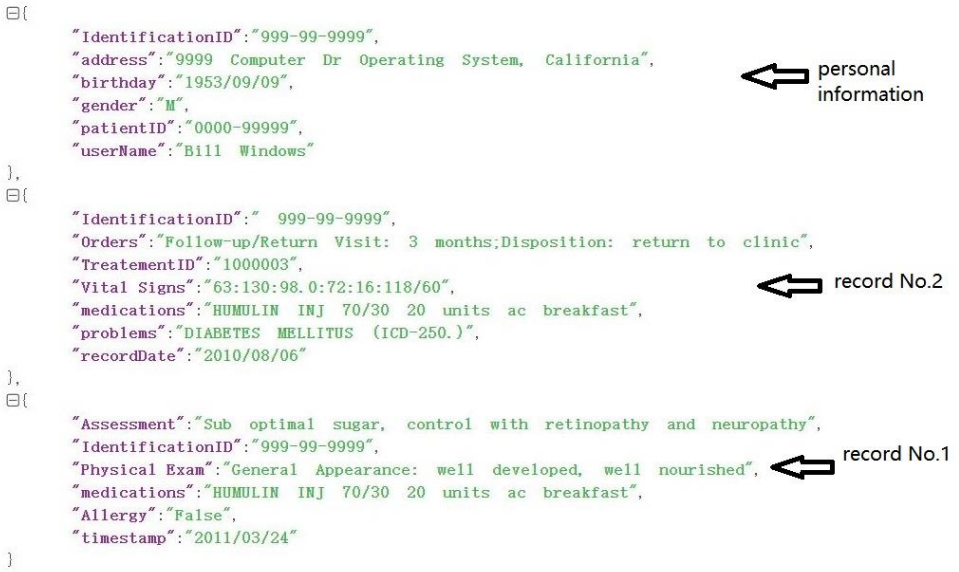
The retrieved result in JSON file is shown in Fig.5. The result consists of three parts. The first part of the JSON file is the user's personal information, the second part is record one and the third part is record two. The minus symbol is used to fold sections in JSON file.

Table 1 The Record No.1

Column Name	Value
DoctorID	12000
IdentificationID	999-99-9999
Physical Exam	General Appearance: no acute distress
Medications	HUMULIN INJ 70/30 20 units ac breakfast
Assessment	Sub optimal sugar, control with retinopathy
Timestamp	3/24/2011 12:00:00 AM

Table 21 The Record No.2

Column Name	Value
DoctorID	10000
IdentificationID	999-99-9999
Problems	DIABETES MELLITUS (ICD-250.)
Medications	HUMULIN INJ 70/30 20 units ac
Vital Signs	63:130:98.0:72:16:118/60
Orders	Follow-up/Return Visit: 3 months
RecordDate	8/6/2010



```
⊟{
        "IdentificationID":"999-99-9999",
        "address":"9999  Computer  Dr  Operating  System,  California",
        "birthday":"1953/09/09",
        "gender":"M",
        "patientID":"0000-99999",
        "userName":"Bill  Windows"
},
⊟{
        "IdentificationID":"  999-99-9999",
        "Orders":"Follow-up/Return  Visit:  3  months;Disposition:  return  to  clinic",
        "TreatementID":"1000003",
        "Vital Signs":"63:130:98.0:72:16:118/60",
        "medications":"HUMULIN  INJ  70/30  20  units  ac  breakfast",
        "problems":"DIABETES  MELLITUS  (ICD-250.)",
        "recordDate":"2010/08/06"
},
⊟{
        "Assessment":"Sub  optimal  sugar,  control  with  retinopathy  and  neuropathy",
        "IdentificationID":"999-99-9999",
        "Physical Exam":"General  Appearance:  well  developed,  well  nourished",
        "medications":"HUMULIN  INJ  70/30  20  units  ac  breakfast",
        "Allergy":"False",
        "timestamp":"2011/03/24"
}
```

Fig 5 Records Retrieved from Different Databases

5. Conclusion

A blockchain-based method is proposed to share records stored in different databases are. Records stored in different databases are divided into personal information and medical information part. The record index is abstracted for each record of heterogeneous medical and health information. Personal information and record index are encrypted and stored in block to link records together. Records can be retrieved by the user with private key through blockchain and transformed into a unified form. JSON is introduced as intermedia to process data. The multimedia file like the medical images sharing method will be considered in future research.

References:

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

[2] Al Omar, A., Rahman M. S., Basu A. and Kiyomoto S., Medibchain: A blockchain based privacy preserving platform for healthcare data,International conference on security, privacy and anonymity in computation, communication and storage, 2017, pp.534-543.

[3] Azaria A., Ekblaw A., Vieira T. and Lippman A., Medrec: Using blockchain for medical data access and permission management, 2016 2nd International Conference on Open and Big Data IEEE, 2016, pp.25-30.

[4] Xia Q. I., Sifah E. B., Asamoah K. O., Gao J., Du X. and Guizani M., MeDShare: Trust-less medical data sharing among cloud service providers via blockchain. IEEE Access, 2017, pp.5: 14757-14767.

[5] Dubovitskaya A., Xu Z., Ryu S., Schumacher M. and Wang F., Secure and trustable electronic medical records sharing using blockchain,AMIA annual symposium proceedings, 2017, 2017: pp.650.

[6] Chen Y., Ding S., Xu Z., Zheng H.and Yang, S., Blockchain-based medical records secure storage and medical service framework,Journal of medical systems, Vol.43, No.1, 2019, pp.5.

[7] Li H., Zhu L., Shen M., Gao F., Tao X.and Liu S., Blockchain-based data preservation system for medical data,Journal of medical systems, Vol.42, No.8, 2018, pp.141.

[8] Jamil F., Hang L., Kim K.and Kim D., A Novel Medical Blockchain Model for Drug Supply Chain Integrity Management in a Smart Hospital. Electronics, Vol.8, No.5, 2019, pp.505.