Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Medical Data Sharing Model Based on Blockchain

To cite this article: Wanghu Chen *et al* 2019 *J. Phys.: Conf. Ser.* **1267** 012014

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 213.184.92.107 on 17/07/2019 at 16:31

Medical Data Sharing Model Based on Blockchain

Wanghu Chen, Yuxiang Mu*, Xiaoyan Liang and Yaqiong Gao

College of Computer Science and Engineering, Northwest Normal University, Lanzhou, 730070, China

*E-mail: 1078927193@qq.com

Abstract. In the traditional medical system, individual medical data is managed by hospitals rather than individual patients. It is difficult to exchange effectively with fragmented storage, and large amounts of data are difficult to realize their potential value. With the rapid development of medical informatization, centralized storage of fragmented medical data has been unable to meet the relevant needs of the medical industry. To solve the difficulty of sharing and the complexity of confirming rights in the medical system, this paper proposes a medical data sharing model based on blockchain. The model provides reliable storage with IPFS file system, uses Proxy re-encryption to realize data sharing and ensure data proprietary rights, and uses Token economic system to measure the contribution in the sharing process, which stimulates the enthusiasm of sharing. At last, based on the existing sharing problem of medical data, the paper shows the potential solution.

1. Introduction

With the rapid development of medical informatization, the areas focusing on electronic medical records and clinical information systems have achieved comprehensive and rapid development [1] ,in the core of interconnection, information sharing and business collaboration. The extensive use of Eletronic Medical Records (EMR) has brought great convenience to the Medical field, making patients' diagnostic information (including prescriptions, laboratory lists, pathological results, MRI images, etc.) become valuable data assets [2]. If a credible inter-agency data sharing platform can be established to provide comprehensive, timely and accurate medical data for patients, doctors and relevant scientific research institutes, it can not only help patients better understand family history to make early prevention, but also help medical personnel develop optimal medical plans to achieve the best therapeutic effect, which will provide researchers accurate data sets to make more efficient and reliable disease prevention and accelerate the development of biomedicine.

However, in the existing medical system, there are still many obstacles to form a credible data sharing platform. On the one hand, patients' personal medical data is managed by different hospitals and enterprises in a decentralized manner, patients cannot control their own medical data, so data storage is discrete and fragmented, making it difficult to exchange effectively. A large amount of useful information is difficult to exert its potential value, forming an "island of informationization. On the other hand, for hospitals and enterprises stored medical data, once their data centers are damaged, it will be difficult to repair those damaged data. Moreover, companies may be motivated by their own interests to buy, sell or modify data. Therefore, the reliability of storing data in the centralized link is also a problem worthy of attention. To sum up, the current centralizing storage of fragmented medical data is no longer sufficient to meet the relevant needs of the medical industry.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

Blockchain is a technical solution for collectively maintaining a reliable database in a distributed environment through decentralization and de-trusting [3]. The rise of blockchain technology provides a new idea for data sharing in the medical industry [4]. With the characters of difficulties to tamper and de-trusting, blockchain technology can ensure the reliability of stored data in the medical system. Decentralized peer-to-peer transmission can solve the problem of barriers to access and non-circulation of information between medical institutions. Cryptography-based Privacy protection avoids the disclosure of privacy of personal medical records. Therefore, this paper proposes a medical data sharing model that can ensure the authenticity and reliability of data, avoid privacy disclosure and stimulate the initiative of sharing based on blockchain.

2. Related work

At present, blockchain technology has been widely applied in finance, Internet and other fields by taking advantage of its features of decentralization, non-tampering and distributed storage. With the gradual improvement of blockchain technology, its application in the medical field has also made rapid development. Many people believe that the medical and health field is the second largest research field besides the financial field. Shrier and Chang et al. used MIT's OPAL/Enigma encryption platform combined with blockchain technology to create a secure environment for storing and analyzing medical data [5]. Based on the blockchain private chain, Kuo T et al. constructed a cross-agency medical health prediction model [6]. Literature [7,8] aims at the current problems of severe fragmentation of medical data, low sharing efficiency, insecure transmission process, lack of data integrity verification and insufficient privacy information protection, the access and sharing of medical data are realized through the Ethereum platform with smart contract. The MedRec framework mentioned in literature [9,10] combines smart contract with access control for automatic permission management, which realizes the integration and permission management of distributed medical data in different organizations. Since the MedRec framework uses the PoW consensus mechanism to maintain blockchain consistency, the computational overhead required is too large. Therefore, the MDSN[11] framework innovates the consensus mechanism using the DPoS consensus mechanism to reduce resource consumption, and the framework adopts the proxy re-encryption method to control access to medical data, which effectively improves the efficiency of data sharing while protecting privacy, but it also has the deficiency of limited data storage capacity.

Therefore, this paper designs a model for sharing medical data in blockchain mode, which proposes a storage scheme for real data transmission off-chain and operation records on-chain, and carries out data transmission off-chain with the IPFS file system. By the means of proxy re-encryption, the shared access to medical data is realized on the basis of protecting patients' privacy. Token economy system is also adopted to measure the contribution of sharing process and improve the sharing enthusiasm of participants.

3. Medical data sharing model

The medical data sharing model consists of three main modules. The first is the data storage module, which combines the InterPlanetary File System (IPFS) on the basis of the blockchain to ensure the reliability of the shared data from the source and to reduce the data users' doubts about the authenticity of the data. In the second part, Proxy re-encryption technology is used to realize the effective transmission of data among medical institutions and to achieve the purpose of data sharing based the premise of ensuring patient privacy through effective access control. Finally, in order to stimulate the enthusiasm of the sharer, Token economy system is introduced to measure the contribution of the sharer uniformly. The overall schematic of the model is shown in Figure 1.

3.1. Storage model of medical data

In the blockchain mode of operation, the lowest level of data interaction will be treated as a transaction, and exists in the blockchain environment in the form of Merkle root through the Merkle hash tree. Therefore, we regard the operation of medical data as a transaction. When data access, copy and other operations occur, we regard the operation record as a transaction and upload it to the blockchain operating environment, which realizes the running mode that the data operation record is on-chain and the real data transmitting is off-chain. Based on the characters of difficulties to tamper and de-trusting, the data operation records on-chain can accurately correspond to the corresponding data changes off-chain.

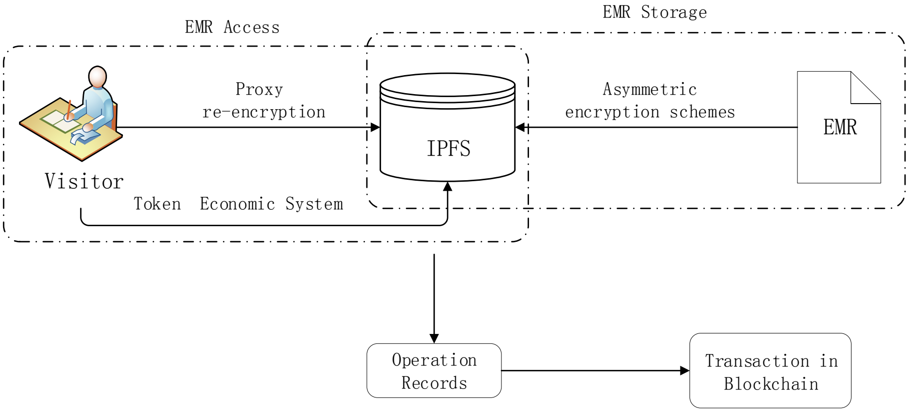

Figure 1. Schematic of the whole model

Since the operational record of the data is running on the chain, the actual data transfer takes place in the IPFS off the chain, and IPFS identifies files by generating a unique Hash value from the contents of the file, so we can divide the data operation into two types. One is the query. During the query process, the data hash value will not change, and we only need to record the data operation, including the querier and the agent (3-2 will be described in detail). The other is modification. In the process of modification, besides the operation records involved in the query, the hash of the corresponding data before and after the operation should be also added to update the data status in the IPFS file system. The corresponding data operation structure is shown in figure 2. Each data operation is accompanied by two state changes. First, an operation record is generated according to the corresponding operation type, which acts as a transaction in the blockchain. Second, after the data operation, it will update the EMR status, which retains the latest data of EMR. The operation record after each data change has a one-to-one correspondence with the data status.

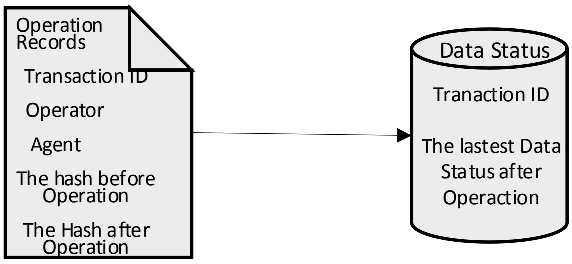

Figure 2. Data Operation Structure

The operation record and data operation structure in the blockchain can theoretically guarantee the accuracy of corresponding EMR, but the reliable storage of real EMR is still a problem. The InterPlanetary File System (IPFS) is a kind of permanent, decentralized distributed File System, which is a distributed protocol that can be used for point-to-point hypermedia through content addressing and versioning [12]. The IPFS file system has the following three characteristics: 1) Content addressability: the file is identified by generating a unique Hash value based on the file content, rather than by the file storage location. For saving storage space, the same content in the system will exist only once; 2) Versioning: trace the modification documents history; 3) Peer-to-peer hypermedia: saves various types of data through P2P. IPFS generates Hash based on content to identify files, which perfectly matches the Hash value of data in data operation structure; IPFS can trace the historical version of files, which well fits the update process of data in EMR; IPFS peer-to-peer transmission mode just fits the data transmission mode off-chain. In general, IPFS is a feasible scheme for storage data off-chain.

As shown in figure 3, the data operation record acts as a transaction, which is processed by Merkle tree and packaged into the blockchain environment in the form of Merkle root, and the corresponding data state is stored in the IPFS file system through Merkle DAG. The two are connected by a data manipulation structure. Data operation records running in the blockchain environment are guaranteed to be true and unforgeable with the guarantee of the blockchain characteristics such as non-tampering and traceability. The data operation record has a one-to-one correspondence by the data status. In this case, the real patient-related information stored in the IPFS is also guaranteed.

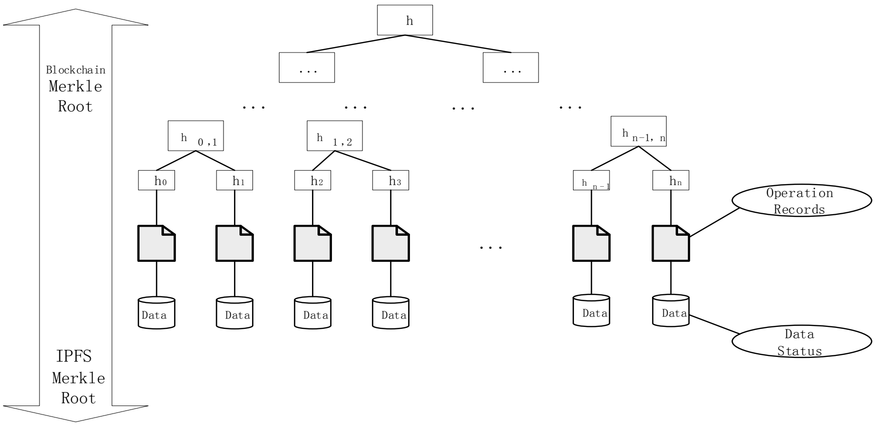

Figure 3. Schematic of Data Storage

3.2. Access control for medical data

The medical industry is very sensitive to data privacy and has a unique set of regulatory requirements for privacy protection. For example, the medical service industry in the United States must comply with the health insurance privacy and liability act (HIPAA) issued by the government of this country in 1996 [13]. Therefore, the application of blockchain in the medical industry needs to overcome the difficulties in privacy protection. In the traditional medical field, the storage of cases is managed by other institutions such as hospitals. Patients cannot have real ownership of their own cases, nor can they participate in EMR management in real time. In the shared model proposed in this paper, patients are the core, they have absolute ownership of their own cases, and other institutions such as hospitals are only third parties that operate on EMR data. All operations on the case must be authorized by the patient.

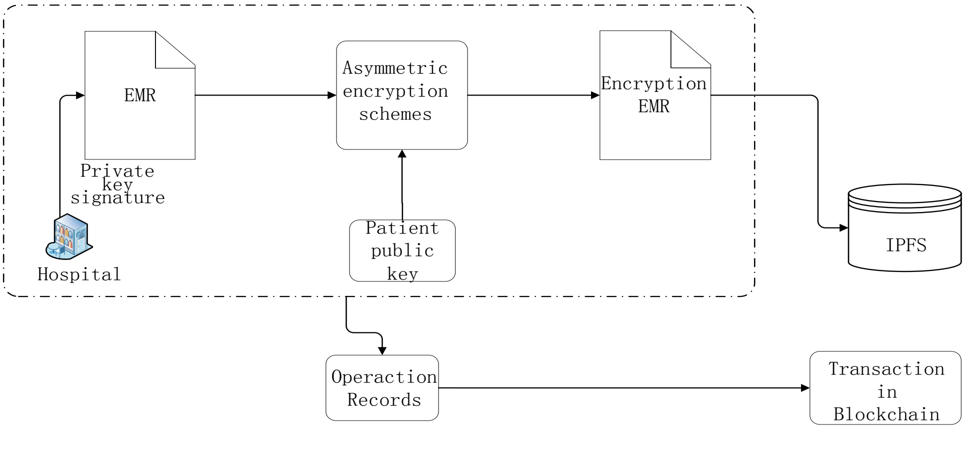

Figure 4. EMR Storage Process

As shown in Figure 4, the electronic medical record stored in the IPFS file system first needs to go through the digital signature of the hospital. The hospital signs with the private key, which avoids the possibility of fraud from the source and provides convenience for future accountability. Then, the patient public key is used to encrypt the EMR with hospital signature using asymmetric cryptography, and the final encrypted EMR is stored in the IPFS. The data operation records generated during the whole process are packaged in the blockchain operation environment in the form of blockchain transactions.

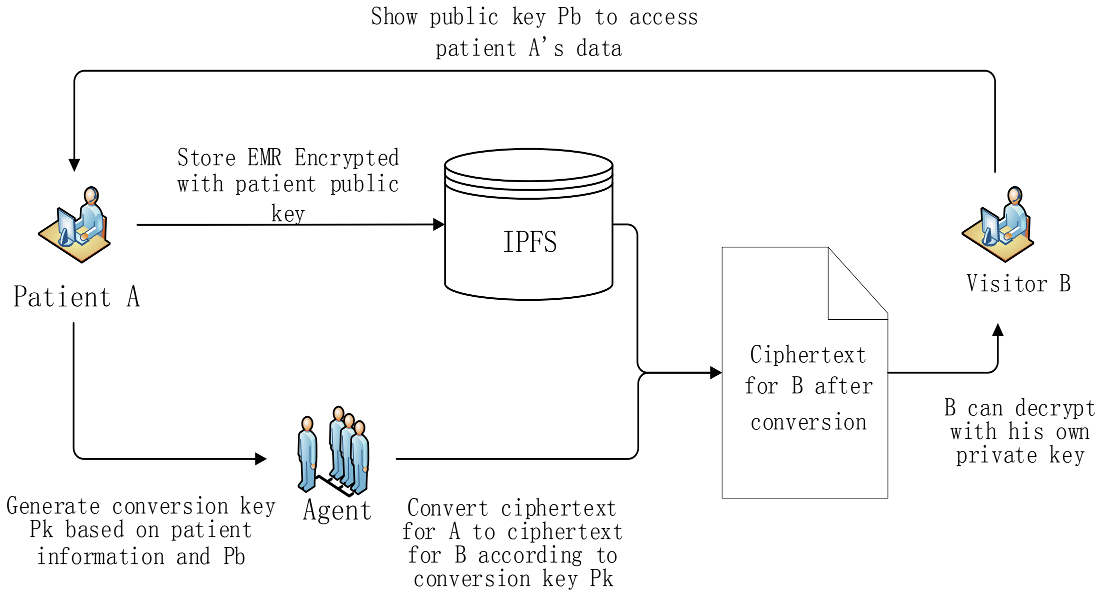
Show public key Pb to access patient A's data
Store EMR Encrypted with patient public key
IPFS
Patient A
Visitor B
Ciphertext for B after conversion
B can decrypt with his own private key
Generate conversion key Pk based on patient information and Pb
Agent
Convert ciphertext for A to ciphertext for B according to conversion key Pk

Figure 5. Proxy re-encryption Process

EMRs stored in the IPFS file system are ciphertext encrypted by the patient's public key and they can only be viewed after the patient authorizes it. In this case, the patient's ownership is guaranteed and the privacy issue is avoided. However, there are still many hidden dangers for patients to directly interact with visitors in the no-center link of the blockchain. Proxy re-encryption is a key conversion mechanism among ciphertexts. In the proxy re-encryption scheme, the semi-trusted agent can transform the ciphertext of the authorized person into the ciphertext of the receiver by using the re-encryption key granted by the authorized person. And in the process of transformation, neither external attackers nor agents can obtain the corresponding plaintext, thus it can reduce the risk of data leakage [14]. There is no trusted third party in the blockchain no-center environment and there is no unique proxy role in the traditional re-encryption scenario. Any node in the medical institution can act as an agent to complete the re-encryption operation and get the corresponding credit points as a reward. Therefore, it is feasible to use the proxy re-encryption scheme for access control. The specific case sharing acquisition process is shown in figure 5. This model uses the identity-based proxy re-encryption design protocol proposed by Jiang Mingming et al. [15] as follows:

1) Master key generation: Generate random matrix A and small norm matrix T, randomly select l + 1 linearly independent vectors $u_0, u_1, \cdots u_l$. Then the master public key MPK=(A, $u_0, u_1, \cdots u_l$)and the master private key MSK=T.
2) User private key extraction: Enter the primary public key MPK, the primary private key MSK, and the user identity id=($id_1, id_2 \cdots id_l$). Calculate u $= u_0 + \sum_{i=1}^{l} id_i u_i$, use the original image sampling algorithm to generate the vector e to satisfy Ae $= \mathrm{u} = u_0 + \sum_{i=1}^{l} id_i u_i$. Therefore, the private key of the user id is e.
3) Proxy re-encryption key generation: For the users id_1 and id_2 , use the corresponding private key $e_{id_1}$and e_{id_2} to calculate $rk_{id_1 \leftrightarrow id_2} = eid_1 - eid_2$ as a proxy re-encryption key.

4) Encryption: Enter the main public key MPK, user identity $id_1 = (id_{11}, id_{12}, \cdots id_{1l})$ and a message bit μ to calculate $\mathrm{u} = u_0 + \sum_{i=1}^{l} id_i u_i$. Select the random vector S to calculate $\mathrm{y} = A^T s + x, c = u^T s + x + \mu\lfloor q/2 \rfloor$. The output is ciphertext (y, c).
5) Re-encryption: Enter the re-encryption key $rk_{id_1 \leftrightarrow id_2} = eid_1 - eid_2$ and the ciphertext $(\mathrm{y}, c_{id_1})$of user $id_1 = (id_{11}, id_{12}, \cdots id_{1l})$. Agent uses proxy re-encryption key to calculate $c_{id_2} = c_{id_1} - rk_{id_1 \leftrightarrow id_2}^T \mathrm{y}$, Output the ciphertext$(\mathrm{y}, c_{id_2})$of the user $id_2 = (id_{21}, id_{22}, \cdots id_{2l})$.
6) Decryption: Enter the private key e_{id_2} of user $id_2 = (id_{21}, id_{22}, \cdots id_{2l})$ to calculate $c_{id_2} - e_{id_2}^T y$. If the result is close to 0, it outputs 0, and if the result is close to $\lfloor q/2 \rfloor$, it outputs 1.

3.3. Measurement of value in the Shared pattern

In the medical sharing model, patients, hospitals, and third-party research institutes are independent from each other, and there is no direct interest to make them mutually beneficial in the EMR sharing process. As a result, the participation enthusiasm of all parties in data sharing is not high [16]. Therefore, Token economy system is adopted in this paper to make measure the contribution in the process of data sharing, which is displayed in the form of credit score. In the end, it will form a win-win whole that offers preferential treatment to patients, enhances the hospital's popularity and promotes the development of scientific research institutes.

When a third-party access agency, such as a research institution, requests to view the data, the patient is authorized to an agent, which can increase the patient's own credit score. When the patient's credit score reaches the corresponding threshold, the next diagnosis can be carried out corresponding preferential treatment. Since the EMR has hospital's signature, the hospital's credit score will be increased after the review, which has an impact on hospital rankings. In this case, patients will actively share their own cases, and hospitals can also benefit from sharing. There will form an organic whole of mutual benefit, cooperation and symbiosis.

4. Model Evaluation

This paper adopts the method of comparative analysis to evaluate the proposed medical data sharing model. On the one hand, compared with the existing sharing scheme, we analyze the advantages and disadvantages of this model, the result is shown in Table 1. On the other hand, combined with the existing problems, the solution to the analytical model is shown in table 2.

Table 1. Comparisons with existing sharing schemes

	Third parties	Private chain	Incentives	Storage	Consensus
ABE program [16]	YES	—	NO	LARGE	NO
MedRec[9][10]	NO	NO	NO	SMALL	PoW
MDSN[11]	NO	YES	NO	SMALL	DPoS
The Model	NO	NO	HAS	LARGE	Decide with actual condition

Table 1 compares the model with the existing research results from five aspects. Overall, the model has certain advantages. On the basis of the blockchain, the IPFS file system is combined to expand the storage capacity, and the incentive mechanism has been introduced to increase the enthusiasm for participation.

Table 2. Existing problems and solutions

Existing Problems	Solution
Consensus	Submit the data operation records to the public chain in the form of transactions, including Bitcoin, Ethereum and EOS
Security privacy	Use an asymmetric encryption scheme to store, use proxy re-encryption to access
Participation enthusiasm	The inclusion of Token economy makes the value in the sharing process be measured uniformly, so as to ensure that each participant gets what he needs
EMR ownership	The EMR is encrypted according to the patient's public key using the asymmetric encryption system, and the ownership always belongs to the patient
EMR fabrication and accountability	With the help of blockchain and IPFS, it can achieve traceability and accurate positioning can for hospitals that falsify data

Table 2 analyzes the solution of the model for existing problems. For the patient, it can not only ensures privacy, but also protects the ownership of their own cases. For the hospital, it not only ensures the authenticity of its own medical information, but also lays a foundation for the hospital to raise awareness. But there are still many flaws and areas that need to be optimized, such as no private medical chain, no predictive ability and so on.

5. Conclusion

In this paper, a medical data sharing model is designed in the blockchain operation mode. The paper shows an operation mode of real data transmission off-chain and operation records on-chain, which uses the IPFS to realize the efficient storage, takes asymmetric encryption system and proxy re-encryption to control data sharing access, and introduces Token economy system to measure the contribution in the sharing process to improve user engagement. However, with the continuous development of the blockchain technology, how to realize the patient's fine-grained access to the diagnosis information and the prediction of the patient's condition in the case sharing process is still an area worthy of investigation.

ACKNOWLEDGEMENT

This work is supported by the National Natural Science Foundation of China (No. 61462076).

References

[1] Huang Jian-Hua, Jiang Ya-Qin, Li Zhong-cheng, et al. Application Prospect of Blockchain in Medical Industry [J]. *Journal of Medical Intelligence*, 2018, 39(2):1-8, 13.

[2] MEI Ying. The Utilizing Blockchain-Based Method of the Secure Storage of Medical Records [J]. *Journal of Jiangxi Normal University(Natural Sciences Editio*n), 2017, 41(5):484-490.

[3] HE Pu, YU Ge, ZHANG Yan-feng BAO, et al. Survey on Blockchain Technology and Its Application Prospect [J]. *Computer Science*, 2017, 44(4):1-7.

[4] YUAN Yong WANG Fei-Yue. Blockchain: The State of the Art and Future Trends [J]. *Acta Automatica Sinica*, 2016, 42(4): 481-494.

[5] Shrier A A, Chang A, Diakun-thibault N, et al. Office of the national coordinator for health information technology us department of health and human services[J]. 2016.

[6] Kuo T T, Ohno-Machado L. ModelChain: Decentralized Privacy-Preserving Healthcare Predictive Modeling Framework on Private Blockchain Networks [J]. arXiv preprint arXiv:1802.01746, 2018.

[7] Dagher G G, Mohler J, Milojkovic M, et al. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology [J]. *Sustainable Cities and Society*, 2018, 39(1): 283-297.

[8] McFarlane C, Beer M, Brown J, et al. Patientory: A Healthcare Peer-to-Peer EMR Storage Network v1 [J]. 2017.

[9] Azaria A, Ekblaw A, Vieira T, et al. MedRec: Using Blockchain for Medical Data Access and Permission Management[C]// International Conference on Open & Big Data. 2016.

[10] 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.

[11] Xue T F, Fu Q C, Wang C, et al. Study on Medical Data Sharing Model Based on Blockchain[J]. *Acta Automatic Sinica*, 2017, 43(9): 1555-1562.

[12] Benet J. IPFS-content addressed, versioned, P2P file system [J]. arXiv preprint arXiv:1407.3561, 2014.

[13] U U. Health Insurance Portability and Accountability Act of 1996. Public Law 104-191[J]. *United States Statutes at Large*, 1996, 110(1):1936-2123.

[14] Ge ChunPeng. Research on Several Issues of Proxy Re-encryption [D]. Nanjing University of Aeronautics and Astronautics, 2016.

[15] JIANG Ming-ming, GUO Yu-yan, YU Lei, et al. Efficient Identity-based Proxy Re-encryption on Lattice in the Standard Model [J]. Journal of Electronics & Information Technology, 2019, 41(1):61-66.

[16] Xhafa F, Li J, Zhao G, et al. Designing cloud-based electronic health record system with attribute-based encryption [J]. Multimedia Tools and Applications, 2015, 74(10):3441-3458.