A Privacy-Preserving Medical Data Traceability System Based on Attribute-Based Encryption on Blockchain

Yujuan Zhao^(✉), Baojiang Cui, and Jie Xu

Beijing University of Posts and Telecommunications, Beijing, China
yjzhao192@bupt.edu.cn

Abstract. With the continuous development of distributed technology, blockchain has been widely applied to solve practical problems in various fields. Its advantages such as decentralization and non-tamperability perfectly meet the needs of permanent query of traceability scene information, and can solve the trust problem of multi-party participation in various links. At the same time, as data security has received more and more attention. The privacy protection and permission control of “on-chain” data need to be resolved. Medical data records are very valuable and sensitive data. Reliable storage and credible traceability are of great significance to the health of patients and the long-term development of medical undertakings. To issue the above problems, here, we propose a secure medical data traceability system that combines data desensitization and attribute-based encryption algorithms to achieve the traceability capabilities and ensure the reliable storage of medical data records on blockchain. Data desensitization provides a certain degree of security for sensitive data involving personal information in medical data. The introduction of attribute encryption has solved the problem of unauthorized access to data. Finally, we analyzed the security of the solution and Verify the efficiency of the program.

Keywords: Hyperledger Fabric · Reversible desensitization · Attribute encryption

1 Introduction

With the rapid development of information technology, medical data, as an important data asset, has received more and more attention. Research on medical data is also an important way to promote the development of the medical field. Medical data is of special importance and sensitivity, relating not only to information related to the medical treatment process, but also to personal privacy and other information, and even to industry development and national security. With the informatization of medical data, the risk of its leakage has been increasing. The methods of leakage are mainly divided into interactive leakage and non-interactive leakage. Interactive disclosure refers to the disclosure of medical information when it is released and shared among different institutions; Non-interactive disclosure refers to the disclosure of data by hospital internal systems or personnel, such as private reselling or misuse of information [1]. At present,

most of the medical data is stored in the hospital's local database. In the daily management process, the information department of many hospitals is not very aware of prevention, and is vulnerable to attacks in the process of multi-party interaction. This will undoubtedly lead to privacy leakage and cause serious consequences.

Blockchain technology, also known as distributed ledger technology, is an Internet database technology. As shown in Fig. 1, the blockchain composes data blocks into a chain data structure in chronological order, and uses cryptographic technology to ensure that the data on the chain cannot be tampered with or forged. Each node uses the Hash algorithm and Merkle tree to encapsulate the transactions received within a period of time into a block with a timestamp, and link it to the longest main chain to form the latest block [2].

Fig. 1. The structure of blocks

Blockchain is a distributed shared ledger and database, which has the characteristics of decentralization, non-tampering, openness and transparency, traceability, and programmable. The blockchain records important data in the circulation process of blood cell extraction, transportation, processing, and use in medical scenarios, and provides traceability for the entire process. However, the data privacy issues contained in medical data are still tricky. Data desensitization can process private information such as ID card numbers closely related to patients' personal information, while ensuring the standardization and reversibility of data, and attribute encryption can provide fine-grained access control while protecting data confidentiality. This type of technology provides support for the realization of data privacy and security sharing.

Block chain fusion cryptography principle guarantees non-tampering and data confidentiality, and realizes the function of data storage and traceability [3]. Aiming at a series of practical problems such as the incomplete credibility of the current medical data and the high risk of sensitive data leakage, this paper proposes a scheme based on the combination of blockchain and reversible desensitization and attribute encryption algorithms. First, reversible desensitization was performed on ID number, mobile phone number and other private information in the patient information. Then the other on-chain data is classified and processed into fully public data and restricted privacy data. After that the restricted privacy data is encrypted and uploaded. Finally, in the traceability

information query, access control through attributes is adopted to achieve hierarchical protection of data, greatly improving the security of data.

2 Related Work

2.1 Blockchain Technology

In the medical field, G. Zyshind et al. [4] proposed to store user access control permissions and data hash values in the blockchain, and encrypt the data to be hosted in a trusted third party, which did not achieve decentralization, while increasing the risk of data leakage. Azaria A [5] and others proposed to combine blockchain and public key encryption technology for access rights management, but the consensus mechanism of this scheme uses Proof of Work, which will consume a lot of computing power for data storage to ensure the authenticity of data on the chain. Wang et al. [6] proposed that the information owner distributes keys for data users and encrypts the data through the specified access rules of the Ethereum smart contract, and controls the access to the data through the attribute-based encryption mechanism.

2.2 Reversible Data Desensitization

Data desensitization technology [7] deforms sensitive data through deletion, masking, replacement and other means to achieve reliable protection of sensitive data. Data desensitization technology mainly has three types of methods: the first is a covering method, which is an irreversible solution to achieve data protection by covering the data; the second is to convert into a random sequence, using hash encryption Algorithms make data lose business attributes. This method is only suitable for scenarios with strong data protection requirements; the third is reversible desensitization, which uses reversible algorithms such as data table mapping and algorithm mapping to achieve data protection, which can guarantee the business attributes of data. It also has the characteristics of reversible reduction [8].

2.3 Attribute-Based Encryption Technology

Attribute-based encryption can refine the degree of access control to shared data to the attribute level. Li et al. [9] aimed at the centralization of existing cloud storage technologies, and proposed an attribute encryption access control based on the combination of blockchain and cryptographic accumulators to achieve privacy protection. Wang et al. [10] proposed a scheme based on Multi-Authority Attribute Based Encryption to achieve privacy protection and access control. Li et al. [11] proposed a decentralized storage solution that combines ciphertext-policy attribute-based encryption (CP-ABE), decentralized multi-authority attribute-based signatures (DMA-ABSs) and blockchain technology, but when the access policy is changed, data encryption needs to be re-encrypted and new key distribution is required.

3 System Model

3.1 Reversible Data Desensitization

Table 1 presents some of the notations used throughout the paper.

Table 1. Notations.

Notation	Description
M	Original coded data
N	Global secret integer
ID	ID of the desensitization task
g	Greatest common divisor of M and N
C	Coded data after desensitization
M'	Recovered coded data
ID'	Identification obtained by traceability
ID^{-1}	The multiplicative inverse element under N
$(M/g)^{-1}$	The multiplicative inverse element of M/g modulo N/g

The coded data reversible desensitization and traceability algorithm based on the Abelian group can generate a global security integer of the target order of magnitude according to the original coded data to be desensitized, and the size of the target order of magnitude is determined according to the data length of the data to be desensitized; The global secret integer is used to obtain the Abelian group; the desensitization task identifier is selected from the Abelian group; the original encoded data is encoded according to the desensitization task identifier and the global secret integer to obtain the desensitized encoded data:

$$C = M \cdot ID \pmod{N} \quad (1)$$

Restore the desensitized encoded data:

$$M' = C \cdot ID^{-1} \pmod{N} \quad (2)$$

Trace the source of encoded data according to the following formula:

$$ID' = C/g \cdot (M/g)^{-1} \pmod{N/g} \quad (3)$$

3.2 Access Control Based on Attributes

The core idea of attribute-based access control is to use attributes to express restricted information in the access control model, and generate corresponding access policies through certain logical relationships [12] (Fig. 2).

Fig. 2. Basic model architecture

An attribute-based encryption technology that can solve the problem of data access management through reasonable configuration of sharing strategies [13]. According to the different embedded objects, attribute based encryption (ABE) can be divided into Key-Policy Attribute-based encryption (KP-ABE) and Ciphertext-policy Attribute-based encryption (CP-ABE) [14].

CP-ABE uses a password mechanism to protect data [15]. The data owner specifies the strategy for accessing the ciphertext, and associates the attribute set with the access resource, and the data user can access the ciphertext information according to their authorized attributes [16].

Fig. 3. System model

As shown in Fig. 3, there are three types of users in this system, namely hospital users, transportation users and factory users. Hospital users are mainly blood collection doctors, examiners, quality monitors, etc., transportation users are mainly transport managers, quality inspectors, etc., and factory users are mainly quality managers, cell processing managers, etc. The user submits the responsible part of the key information data, uploads the data to the Fabric through the smart contract for storage, and saves the personal privacy-related data submitted by the hospital user about the patient after reversible desensitization. The system process includes extracting the patient's blood cells, transporting them to the factory for Chimeric antigen receptor T (CAR-T) cell immune treatment after inspection, and transporting them back to the hospital for patient use after the treatment. Because the immunotherapy is still in clinical trials, the entire process needs to be preserved for future research and analysis. The CP-ABE strategy is used for data access, and the access authority is controlled through the attribute set and the access control tree, so as to realize the hierarchical protection of the data.

4 Scheme

We design a CP-ABE-based blockchain medical data traceability system. The scene is the whole process of CAR-T cell therapy, including information such as blood collection and testing in the hospital, cold chain transportation information of blood bags, cell separation and other processing information. The system process is as follows (Fig. 4).

The hospital collects the patient's blood and puts it into the blood bag after the collection is completed. Each blood bag corresponds to a unique QR code. The corresponding doctor submits the patient information, blood information, etc., and formulates the attribute encryption strategy and then encrypts the encrypted data. Store the blood bags on the chain, and then transport the blood bags to the factory for processing. The temperature and humidity are monitored by sensors to ensure the normal progress of the cold chain transportation process. After arriving at the factory, the corresponding factory responsible personnel will check and receive the blood bags. The processing process uses key information such as reagents and operators to be encrypted and then chained. After the preparation is completed, the blood is transported back to the hospital for patients to use. By scanning the QR code, doctors who meet the attribute policy on the hospital side can view the complete information of the blood bag, and the operators who meet the attribute policy on the factory side can view the relevant operation information of the factory and the transportation status. The personnel on the transportation side can only view the transportation-related information. A blood bag quality test must be carried out at the handover of each link, and the results and the information of the test personnel shall be recorded to ensure the quality of the blood bag.

Fig. 4. System workflow

Different organizations correspond to different data collection and uploading work. First, during the upload process, the system will first perform data classification. Reversible data desensitization of ID card number, mobile phone number and other data will ensure the privacy and standardization of data. Encrypt the data through the symmetric key, and then use the attribute access control strategy to encrypt the symmetric key. When the encrypted data is on the chain, consensus, and traceability access to the data, the access application is first submitted, and then the attribute verification is performed by the smart contract. If the attribute is valid, the corresponding key will be sent, and the encrypted symmetric key can be decrypted by this key, so as to decode the ciphertext and access the data.

Algorithm : Decryption

Input: Ciphertext C , Attribute collection Ai , Decryption Key DK , Public Key PK

Output: Invalid Signature or Unencrypted Data

```

1 if  $Ai$  Satisfy (Access strategy ) then
2   | Return "1: Unencrypted Data"
3 else
4   | Return "0: Invalid Signature"

```

5 Performance and Safety Analysis

We implemented a prototype to analyze the feasibility and performance of the program. The specific configuration of the experimental platform and experimental environment is: Intel Core i7-8565U@1.80 GHz processor, 8 GB RAM, and the system is ubuntu 16.04LTS. Fabric version is 1.4.2.

Fig. 5. Operation time

As shown in Fig. 5, the data size is set to 128B, and the number of access attributes is set to 4, 8, 12, 16. This scheme implements the Decentralized Ciphertext Policy Attribute Based Encryption schema described in [17]. The code depends on the jPBC library [18] and the Bouncy Castle library.

In terms of security, in this solution, only when the attributes of the visitor match the attributes in the access policy formulated by the data owner can the key be obtained, so that the corresponding data can be accessed. If there is no match, the key cannot be obtained. In the design of this solution, even if the visitor's identity is legal, the data can be accessed, but the private data such as the ID number is still non-real data. Therefore, this program is safe.

6 Summary

CP-ABE is a widely used solution that can achieve fine-grained access control. However, the direct use of traditional CP-ABE in medical information systems will still have certain access strategy leaks and high algorithm complexity. The solution in this article combines data reversible desensitization with the improved CP-ABE, and uses the characteristics of the blockchain to have greater advantages in function and efficiency for the actual scenarios of the actual CAR-T preparation process. In the next step, we will consider issues related to attribute revocation and make a more optimized plan.

Acknowledgements. This work is supported by CNKLSTISS and the National Natural Science Foundation of China (Grant No. 61802025).

References

1. Hu, R., He, Y., Fan, X.: Research on the security technology of medical privacy protection. *J. Beijing Electron. Sci. Technol. Inst.* **26**(3), 46–54 (2018)
2. Zou, J.: *Blockchain Technology Guide*. Machinery Industry Press, Beijing (2016)
3. Venkatesh, J., Aksanli, B., Chan, C.S., et al.: Modular and personalized smart health application design in a smart city environment. *IEEE Internet of Things J.* **5**(2), 614–623 (2018)
4. Zyskind, G., Nathan, O., Pentland, A.: Decentralizing privacy: using blockchain to protect personal data. In: 2015 IEEE Security and Privacy Workshops (SPW), San Jose, pp. 180–184. IEEE Computer Society (2015)
5. Azaria, A., Ekblaw, A., Vieira, T., et al.: MedRec: using blockchain for medical data access and permission management. In: 2016 2nd International Conference on Open and Big Data (OBD), Vienna, pp. 25–30. IEEE (2016)
6. Wang, S., Zhang, Y., Zhang, Y.: A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems. *IEEE Access* **6**(7), 38437–38450 (2018)
7. Xing, Y., Zhang, B., Mao, Y.: Application of data desensitization in massive data systems. *Telecommun. Sci.* **33**(S1), 8–14 (2017)
8. Hu, R., He, Y., Zeng, P., Fan, X.: Design and implementation of a medical privacy protection scheme in a big data environment. *Inf. Netw. Secur.* **09**, 48–54 (2018)
9. Li, X.: Research and implementation of blockchain technology in secure cloud storage. PLA Strategic Support Force Information Engineering University for the Degree of Master of Engineering, Zhengzhou (2020)
10. Wang, J., Xie, Y., Wang, G., et al.: A method of privacy preserving and access control in blockchain based on attribute-based encryption. *Netinfo Secur.* **20**(9), 47–51 (2020)
11. Li, G., Sato, H.: A privacy-preserving and fully decentralized storage and sharing system on blockchain. In: Proceedings of the IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), WI, USA, pp. 694–699 (2019)
12. Jemel, M., Serhrouchni, A.: Decentralized access control mechanism with temporal dimension based on blockchain. In: Proceedings of the IEEE International Conference on e-Business Engineering, Shanghai, China, pp. 177–182 (2017)

13. Li, X.: Application research and implementation of blockchain technology in secure cloud storage. Information Engineering University of Strategic Support Forces, Zhengzhou (2020)
14. Cao, L., Liu, Y., Dong, X., Guo, X.: User privacy-preserving cloud storage scheme on CP-ABE. *J. Tsinghua Univ.* **58**(2), 150–156 (2018)
15. Wang, X., Jiang, X., Li, Y.: Data access control and sharing model using blockchain Type. *J. Softw.* **30**(6), 1661–1669 (2019)
16. Tian, Y., Yang, K., Wang, Z., et al.: Algorithm of blockchain data provenance based on ABE. *J. Commun.* **40**(11), 101–111 (2019)
17. Lewko, A., Waters, B.: Decentralizing attribute-based encryption. In: Paterson, K.G. (ed.) *EUROCRYPT 2011*. LNCS, vol. 6632, pp. 568–588. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-20465-4_31
18. <http://gas.dia.unisa.it/projects/jpbc/>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.