J Med Syst (2019) 43: 26
https://doi.org/10.1007/s10916-018-1144-x

TRANSACTIONAL PROCESSING SYSTEMS

Medical Data Management on Blockchain with Privacy

Haibo Tian[1] · **Jiejie He**[1] · **Yong Ding**[2]

Received: 28 February 2018 / Accepted: 18 December 2018 / Published online: 3 January 2019

Abstract

Medical data are important in diagnosis, treatment, recovery, and medical accident investigation. The integrity and availability of medical data are the basic guarantee for the smooth operation of these activities. The privacy of medical data is a natural demand from the sensitivity of medical data. At present, there are mainly two ways to protect the privacy of medical data. One way is to store medical data in a local database and set up an access control strategy of the database. The other way is to encrypt medical data with the patient's key and to share the key when needed. The problem with the first method is that the data in the local database may be modified or deleted. The problem with the second method is that the key cannot be shared when the patient dies during the diagnosis and treatment. These two problems will damage the availability of data. This paper proposes to establish a shared key that could be reconstructed by the legitimate parties before the process of diagnosis and treatment begins. The data in the diagnosis and treatment process is encrypted and stored in a blockchain using the shared key. The proposal meets the integrity, availability and privacy requirements of medical data. It uses the sibling intractable function families (SIFF) to establish a shared key, and uses the Hyperledger Fabric to store encrypted data. The simulation shows that the system has good efficiency. Additionally, it is the first time to introduce SIFF to a blockchain application.

Keywords Medical data · Privacy · Integrity · Availability · Blockchain

Introduction

Medical data, as the digital evidence of patients' diagnosis and treatment process, is very important. According to statistics in 2016 [1], about 17,000 malpractice cases are filed in the United States. In medical malpractice, the plaintiff has the burden of proof to prove all the elements by a preponderance of evidence [2]. The defendant obviously has a strong desire to delete or modify the adverse digital evidence.

Centralized medical data management system is difficult to ensure the integrity of medical data. In such a system, medical data are often stored in a database of a medical center. An attacker can delete or modify the data after obtaining the corresponding permissions of the database. What is more serious is that in case of a medical malpractice, the manager of the medical center may directly request the administrator of the database to delete or modify the data.

Decentralized blockchain technology can naturally ensure the integrity of medical data. The medical data in a blockchain are distributed in storage devices of different parties. It does not affect the data of other parties if the data of a few parties are modified or deleted. Under a consensus mechanism, the data in the blockchain is still intact. Omar et al. [3], Yue et al. [4] have proposed to put encrypted medical data in a blockchain.

The problem of blockchain technology lies in the privacy and availability of data. In order to ensure privacy, people often encrypt medical data. The encryption uses a patient's key. The key is obtained from the patient when encrypted data is used [3–5]. However, according to a report in 2013 [6], the most common consequence of medical negligence is the death of patients. A plaintiff in a medical malpractice

This article is part of the Topical Collection on *Blockchain-based Medical Data Management System: Security and Privacy Challenges and Opportunities*

✉ Haibo Tian
tianhb@mail.sysu.edu.cn

[1] Guangdong Key Laboratory of Information Security, School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, Guangdong, 510275, People's Republic of China

[2] Guangxi Key Laboratory of Cryptography and Information Security, School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, Guangxi, 541004, People's Republic of China

is often the executor or administrator of a deceased patient's estate [2]. The plaintiff often do not know the patient's key. This directly damages the availability of medical data.

This paper provides a new scheme to satisfy the integrity, availability and privacy of medical data at the same time. Considering the visible range of a patient's diagnosis and treatment data in real life, we propose to use the sibling intractable function families (SIFF) [7] to establish a shared key for a small number of entities and use the shared key to encrypt the data in a diagnosis and treatment process and store it in a blockchain. Thanks to the characteristics of SIFF and blockchain technique, each entity can decrypt the authorized medical data only by using its own password.

Related works

Omar et al. [3] proposed a privacy preserving medical data platform based on blockchain. The data is encrypted and stored in a federation blockchain, and a data user needs to get the decryption key from a data owner. Yue et al. [4] proposed a health care data gateway architecture. The data is encrypted and stored in a private blockchain cloud. A patient downloads encrypted data, decrypts it and decides whether to share the data. When the data is shared, it is encrypted again, and the ciphertext and decryption key are sent to the receiver. Alevtina et al. [5] proposed a medical records sharing system based on blockchain. The patient's data is encrypted and stored in a cloud server. When the data is to be used, it is necessary to obtain the decryption key from the data owner. The blockchain here is responsible for enforcing access control policies.

Azaria et al. [8] proposed a medical data access and permission management system based on blockchain. Medical data is stored in a local database. Gem Health [9], an application on GemOS [10], builds medical ecology including patients, pharmacy, families, doctors, nurses, and payers. GemOS [10] can integrate local databases into a blockchain that provides data management services. Xia et al. [11] proposed a medical data sharing system based on blockchain, which explicitly requires that smart contracts contain secrete keys. Since blockchain smart contracts are run by blockchain nodes, the scheme requires that all blockchain nodes should have the same key. If a blockchain node is malicious, the privacy of data will be damaged.

Table 1 summarizes the above works in terms of availability, integrity and privacy. If patients are required to participate in the use of data, we believe that the availability of the scheme is limited. If data is stored locally or in a cloud server, we think the integrity of the scheme is weak. If a scheme needs to store a key in a smart contract, we think its privacy is weak. In this paper, the encrypted data is stored in a blockchain and is used without the participation of a patient, and a smart contract is used without any keys embedded, which balances the integrity, availability and privacy properties of medical data.

Table 1 Summary of related works

Proposals	Availability	Integrity	Privacy
Omar et al. [3]	Limited		
Yue et al. [4]	Limited		
Alevtina et al. [5]	Limited	Weak	
Azaria et al. [8]		Weak	
Gem Health [9]		Weak	
Xia et al. [11]			Weak

Preliminaries

Hyperledger Fabric

Fabric blockchain [12] is a project of the Hyperledger program led by the Linux Foundation. A Fabric blockchain consists of four types of entities: peer node, orderer node, Fabric certificate authority (CA) and client. Fabric CA issues certificates to other entities to build trust domains. Peer nodes can be functionally divided into endorser and committer nodes.

A client sends a proposal to some endorser nodes. The endorser nodes run a chaincode specified by the proposal and return endorsements back. Then the client packages the proposal and endorsements as a transaction to orderer nodes. The orderer nodes form a block under some consensus algorithm and sends the block to committer nodes. The committer nodes update the Fabric ledger. A client can subscribe custom events produced by a chaincode. Notifications of the events are obtained by the client when the ledger is updated.

SIFF

Zheng et al. [7] proposed the concept and construction method of SIFF. The property of SIFF is that given a function f in SIFF, a set of initial strings $\{s_1, \ldots, s_k\}$, satisfy $f(s_1) = \ldots\ldots = f(s_k)$, and it is computationally infeasible to find a new string $s^* \notin \{s_1, \ldots, s_k\}$ to satisfy $f(s^*) = f(s_j)$ for any $j \in \{1, ..., k\}$. The function f can be constructed from any one-way function h and polynomial u.

Given polynomial $u(\beta) = \alpha_0 + \alpha_1\beta + \cdots + \alpha_{k-1}\beta^{k-1}$. For each $s_j \in \{s_1, \ldots, s_k\}$, compute $\beta_j = h(s_j)$ for $j \in \{1, ..., k\}$, and form a equation set $\{u(\beta_j) = u_0||u_1\}$ where u_0 is a random part and u_1 is the collision. From the equation set, one can compute coefficients $\{\alpha_0, \ldots, \alpha_{k-1}\}$. The function $f(s_j) = u_1$ for $j \in \{1, ..., k\}$.

System model

The system participants include patients, family members, doctors, nurses, pharmacists, insurers, lawyers, medical experts etc. Each participant has a Fabric client that interacts with the Fabric blockchain. The entities of Fabric are maintained by hospitals, healthcare centers, pharmacies, insurance companies, law firms and other institutions to form a consortium blockchain.

An adversary can corrupt any clients or Fabric entities, and control the network. In order to define meaningful security properties, we require that the adversary cannot corrupt the Fabric CA, and cannot corrupt too many orderer and peer nodes so that the Fabric consensus fails. In addition, the adversary should deliver messages of honest participants within a bounded delay.

We define the security properties of the system as follows:

- Integrity: the medical data of honest participants can be written into Fabric, and it is infeasible for an adversary to modify or delete data in the Fabric blockchain.
- Availability: the medical data of a deceased patient can be obtained by authorized participants.
- Privacy: an adversary is infeasible to reveal the medical data of honest participants without authorization.

The system

Setup

We give Table 2 to describe the symbols and their meanings in the system. Each participant in the system has a Fabric client. A participant x uses the client to manage password pw_x and certificate $cert_x$, to execute encryption algorithm enc and Diffie-Hellman protocol DH, to generate and submit proposals and transactions, to monitor custom events, and to make and scan quick response (QR) code. When the participant $x \in \{p, r\}$, its registration to Fabric CA can be anonymous. For example, a traceable anonymous certificate [13] could be issued to such participants.

The chaincode run by Fabric peer nodes includes two functions of $Store$ and Get. The $Store$ function parses the input parameter of a proposal. After verification, the input parameter is stored as a key-value pair in the blockchain. The key is the hash value of the input parameter and the value is the byte sequence of the input parameter. The $Store$ function also sets events according to the input parameter. The Get function parses the input parameters of the proposal, queries the blockchain, and returns the query result to the client.

Table 2 Symbols and their meanings in the system

Symbols	Description
p	A patient
r	A family member of a patient or a legal agent
d	A doctor
m	A pharmacist
a	An insurer
i	A lawyer or medical expert
x	A participant
$cert_x$	A certificate issued by Fabric CA to participant x
(vk_x, sk_x)	The public key in $cert_x$ and the corresponding private key
ind_x	The index of participant x defined as $vk_x \bmod 2^{64}$
pw_x	The password of participant x
pk_x	A temporal Diffie-Hellman public key of the participant x
msg_x	The messages generated by the participant x
Tx_x	The Fabric proposal of the participant x
Tx_{tc}	A token collection proposal
Tx_{tr}^x	The responses of the participant x to Tx_{tc}
enc	A symmetric encryption algorithm
sk	An encryption key of the symmetric encryption algorithm
DH	A Diffie-Hellman key agreement protocol

At the system setup phase, if the participant $x \notin \{p, r\}$, $cert_x$ is stored through the $Store$ function, and the key is ind_x. Any participant can get the public key of another participant by scanning a QR code.

Diagnosis and treatment initialization

A patient triggers a process of diagnosis and treatment. The initialization process establishes a shared key for the patient, a doctor and other informed participants.

1. The patient chooses the participants who should know the diagnosis and treatment process through their client among patients' families, doctors, pharmacists, insurers, lawyers etc. The client can obtain the information of the participants to be selected by performing the Get function or scanning the QR code. The patient's selection result form a QR code with the patient's public key, which is provided to the doctor later.
2. Before the selected doctor gives the patient a diagnosis, the client of the doctor should obtain the information of the patient by scanning their QR code to form a Tx_{tr} proposal. The function in the proposal is $Store$, and the input parameter is

   $$(type, vk_p, [vk_r], [ind_a], [ind_i], [ind_m], ind_d, pk_d)$$

   where $type$ = "$txtc$" meaning a transaction for transaction collection. We use "[ ]" to indicate that

Table 3 Proposals and events in a diagnosis and treatment process

Participant	Proposal	Function in the proposal	Stored data	Event receivers	Event parameter
Doctor	Tx_d	$Store$	(κ_2, ν_2)	ind_p, ind_m	κ_2
Pharmacist	Tx_m	$Store$	(κ_3, ν_3)	ind_p, ind_a	κ_3
Insurer	Tx_a	$Store$	(κ_4, ν_4)	ind_p	κ_4

the field is optional. When patients choose some participants as the insiders, the input parameter of the proposal includes the indexes or public keys of the participants.

3. The peer node receives the proposal and runs the chaincode. The $Store$ function parses the input parameter and stores the input parameter in the key-value pair (κ_0, ν_0). Since the type of the input parameter is "$txtc$", the chaincode sets events to the patient, doctor, and the selected participants. The event parameter includes the key κ_0.
4. When the client of a participant x catches an event, the client gets the input parameter of the proposal indexed by κ_0 through the Get function in the chaincode. Since the input parameter type is "$txtc$", the client computes $s_x = h(pw_x||\kappa_0)$, $sk_{xd} = DH(pk_x, pk_d)$, $c_x = enc(sk_{xd}, s_x)$, and creates the input parameter of a new Tx_{tr} proposal as

   $(type, \kappa_0, pk_x, c_x)$

   where $type$ = "$txtr$" meaning a transaction for transaction collection response. The function in the proposal is $Store$.
5. The peer node receives the proposal and runs the chaincode. The $Store$ function parses the input parameter and stores the input parameter in the key-value pair (κ_1, ν_1). Since the type of the input parameter is "$txtr$", the chaincode sets events to the doctor. The event parameter includes the key κ_1.
6. When the doctor's client receives all the events of the patient and the selected participants, the input parameters of the corresponding proposal are obtained through the parameter of the received events. For each input parameter with c_x, the client computes $sk_{xd} = DH(pk_x, pk_d)$, decrypt c_x with sk_{xd} to obtain s_x. The client then computes $s_d = h(pw_d||\kappa_0)$ to form an initial string set of SIFF as $\{s_x|x \in \{p, d, m, a, i, r\}\}$. After the function f is computed, the client sets $tag = \alpha_0||...||\alpha_{k-1}$ and $sk_a = u_1$ where "||" denotes bit string concatenation.

Diagnosis and treatment process

Suppose that a patient selects a doctor, a pharmacist, an insurer and a lawyer at the initialization phase.

1. After the doctor has established a shared secret sk_a, the doctor can create a message msg_d containing information about diagnosis and treatment, encrypt and store it in the blockchain.
2. The chaincode notifies the pharmacist and the patient through the event mechanism of Fabric.
3. The pharmacist can make a message msg_m including information about charges and medicines, encrypt and store it in the blockchain.
4. The chaincode notifies the patient and the insurer through the event mechanism.
5. The Insurer can form a message msg_a consisting of information such as insurance compensation, encrypt and store it in the blockchain.
6. The chaincode notifies the patient through the event mechanism.
7. Further, if the patient or the patient's relatives or their legal agent doubt the diagnosis and treatment process, the lawyer may be asked to investigate the diagnosis and treatment data in the blockchain. The lawyer can decrypt the related data to find medical evidence.

We summarize the proposals, their functions, stored data, events and event parameters of the above process in Table 3. For proposal Tx_d, the input parameter is stored in ν_2 including

$(type, \kappa_0, ind_a, ind_i, ind_m, ind_p, tag, enc(sk_a, msg_d))$

where $type$="txd" meaning a transaction of a doctor. The input parameter of Tx_m is stored in ν_3 containing

$(type, \kappa_0, ind_a, ind_i, ind_d, ind_p, tag, enc(sk_a, msg_m))$

where $type$ ="txm" meaning a transaction of a pharmacist. The input parameter of Tx_a is in ν_4 consisting of

$(type, \kappa_0, ind_m, ind_i, ind_d, ind_p, tag, enc(sk_a, msg_a))$

where $type$ = "txa" meaning a transaction of an insurer. κ_j is the hash value of ν_j for $j \in \{2, 3, 4\}$. A participant uses the event parameter to get the input parameter of the corresponding proposal. The tag is used to construct the function f in SIFF. A participant x uses its password pw_x to compute $s_x = h(pw_x||\kappa_0)$, and computes $sk_a = f(s_x)$ to decrypt the ciphertext in the input parameter. The $type$ and κ_0 can be combined to get the input parameter of a related proposal of the diagnosis and treatment process by the lawyer.

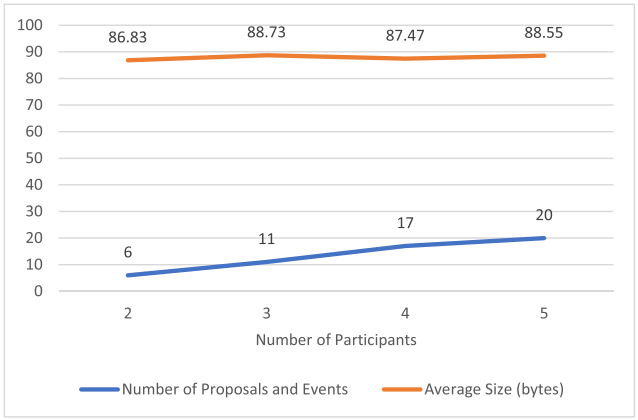

Fig. 1 The communication cost of the system

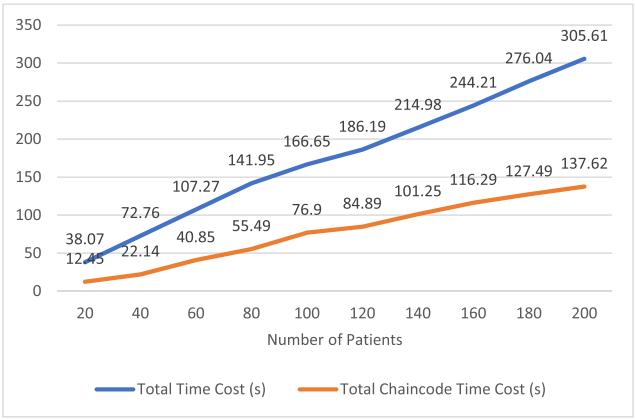

Fig. 2 The time cost of the system

Security analysis

Claim 1 If the digital signature algorithm in Fabric is secure, and the Fabric blockchain has the integrity property, the medical data of honest participants in the system has the integrity and availability properties.

Proof Sketch: The medical data of this system is embedded in the input parameter of a proposal. The digital signature algorithm in Fabric ensures the integrity of the proposal. The Fabric blockchain ensures the integrity of the input parameters. The property of the SIFF guarantees the availability of medical data.

Claim 2 If the Diffie-Hellman protocol, the SIFF and the symmetric encryption algorithm in the system are secure, the medical data of honest participants in the system has the privacy property.

Proof Sketch: The medical data of this system is encrypted by a symmetric encryption algorithm. The encryption key is generated by a function of SIFF. The input of the function is encrypted by a temporary secret key produced by the Diffie-Hellman protocol. If these cryptographic components are secure, the privacy of medical data in the system can be guaranteed.

Performance

We implemented a prototype of the system to analyze its performance. The hardware platform consists of Intel Xeon E3-1241 CPU and 4GB memory. The operating system is Ubuntu 16.04. The symmetric encryption algorithm is AES-256. The DH protocol uses the ECDH version with a SECP 256r1 curve. We set up the Fabric version 1.0 with 4 peer nodes and one orderer node. The throughput is about 50TPS under the default configuration. The client is implemented by the Java language. The chaincode is implemented by the Go language.

Because the scale of participants who know a diagnosis and treatment process is usually small, Fig. 1 shows the average size of the input parameters and events, and the number of proposals and events for a diagnosis and treatment process considering 2 to 5 participants. Assume that each participant generate a message less than 32 bytes. And the lawyers or experts do not generate new proposals. Figure 1 shows that with the increase of the number of participants, the total number of proposals and events increases linearly. The average communication cost of the system keeps almost constant, and the total communication cost increases linearly.

Assuming that there are 3 participants, we measured the running time of the system and chaincode time in 20 to 200 patients during a sequential trigger process. As shown in Fig. 2, patients spend an average of 1.66 s on completion of the system, in which the average time spent on the chaincode is 0.69 s.

Conclusion

In this paper, we introduce the SIFF to enable the sharing of medical data with the granularity of a diagnosis and treatment process among related participants. The underlying ledger is the Fabric blockchain. The final system could guarantee the privacy, availability and integrity of medical data.

Acknowledgments This work is supported by the National Key R&D Program of China (2017YFB0802500), Supported by Guangxi Key Laboratory of Cryptography and Information Security (No.GCIS201711), Natural Science Foundation of China (61672550), Fundamental Research Funds for the Central Universities (No. 17lgjc45).

References

1. Bazemore, N., Does your doctor have malpractice claims? How to find out, https://www.forbes.com/sites/amino/2016/04/19/does-your-doctor-have-malpractice-claims-how-to-find-out/#54189b955a64, Accessed July 2018.
2. Wiki, Medical malpractice in the United States. https://en.wikipedia.org/wiki/Medical_malpractice_in_the_United_States. Accessed July 2018.
3. Omar, A. A., Rahman, M. S., Basu, A., and Kiyomoto, S., MediBchain: a blockchain based privacy preserving platform for Healthcare Data. In: *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*, pp. 534–543, 2017.
4. Yue, X., Wang, H., Jin, D., Li, M., and Jiang, W., Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. *J. Med. Syst.* 40(10):1–8, 2016.
5. Dubovitskaya, A., Xu, Z., Ryu, S., Schumacher, M., and Wang, F., Secure and trustable electronic medical records sharing using Blockchain, arXiv preprint arXiv:1709.06528, 2017.
6. Wallace, E., Lowry, J., Smith, S. M., and Fahey, T., The epidemiology of malpractice claims in primary care: a systematic review. *BMJ Open* 3(6):e002929, 2013.
7. Zheng, Y., Hardjono, T., and Pieprzyk, J., Sibling intractable function families and their applications. In: *International Conference on the Theory and Application of Cryptology*, pp. 124–138, 1991.
8. Azaria, A., Ekblaw, A., Vieira, T., and Lippman, A., MedRec: using blockchain for medical data access and permission management. In: *International Conference on Open and Big Data*, pp. 25–30, 2016.
9. Gem, Gem Health, https://gem.co/, Accessed February 2018.
10. Kannan, S., and Smith, M., GemOS Platform Whitepaper, https://211hzr1wjznm2pclk01j90ly-wpengine.netdna-ssl.com/wp-content/uploads/2016/10/GemOSPlatformWhitepaper.pdf, Accessed July 2018.
11. Xia, Q., Sifah, E. B., Asamoah, K. O., Gao, J., Du, X., and Guizani, M., MeDShare: trustless medical data sharing among cloud service providers via blockchain. *IEEE Access* PP(99):1–1, 2017.
12. Hyperledger, A blockchain platform for the enterprise, http://hyperledger-fabric.readthedocs.io/en/release-1.2/, Accessed July 2018.
13. Network Working Group, Traceable Anonymous Certificate, https://tools.ietf.org/html/rfc5636, Accessed August 2018.