A Blockchain Future for Secure Clinical Data Sharing

A Position Paper

Yan Luo
UMASS Lowell
Yan_Luo@uml.edu

Hao Jin
UMASS Lowell
Hao_Jin@uml.edu

Peilong Li*
Elizabethtown College
lip@etown.edu

ABSTRACT

In the digital healthcare era, it is utmost important to harness medical information scattered across healthcare institutions to support in-depth data analysis. However, the boundaries of cyberinfrastructure of healthcare providers place obstacles on data sharing. In this position paper, we firstly identify the challenges of medical data sharing and management. Then we introduce the background and give a brief survey on the state-of-the-art. Finally, we conclude the paper by discussing a few possible research directions to cope with the challenges in current medical information sharing.

CCS CONCEPTS

- **Security and privacy** → **Security services**; **Systems security**;
- **Networks** → *Cloud computing*;

KEYWORDS

Medical Data; Cloud Computing; Software Defined Networking; Blockchain; Smart Contract

ACM Reference Format:
Yan Luo, Hao Jin, and Peilong Li. 2019. A Blockchain Future for Secure Clinical Data Sharing: A Position Paper. In *ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization (SDN-NFVSec '19), March 27, 2019, Richardson, TX, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3309194.3309198

1 INTRODUCTION

Electronic medical records (EMR) are usually stored in local databases of healthcare providers without being shared among medical research institutes due to privacy and security reasons. Therefore, the interoperability of medical data from different providers becomes a big challenge nowadays, which greatly hinders the medical research that requires a large amount of data.

In the era of cloud computing and big data, there rises a demand on medical data to be shared in a large number of medical research institutes to support better healthcare service and emerging medical solutions. The medical records and clinical trials that scattered across nation-wide hospitals, if integrated in a holistic manner, will bring unprecedented opportunities on precise treatment plans and accurate clinical diagnosis, and further reduce the costs on repetitive medical tests. However, security and privacy compliance regulations such as Health Insurance Portability and Accountability Act (HIPAA)[2] and Health Information Technology for Economic and Clinical Health (HITECH) in United States, or General Data Protection Regulation (GDPR)[1] in Europe, require data to be stored and shared in a secure and privacy-preserving way, and may inflict severe penalties on privacy breach events.

Hence, how to efficiently integrate these segregated medical databases for comprehensive medical care without violating privacy regulations has been an continuously active and difficult task. Generally, existing approaches for healthcare data sharing mainly face the following obstacles.

Interoperability. The shift from traditional enclosed healthcare systems to a more holistic and shared healthcare infrastructure demands that medical data be securely shared among various care providers so that they can work collaboratively. Existing healthcare infrastructure built in an enclosed domain is facing the difficulty of managing the rapidly increasing silos of health information which is hard to be interoperated across multiple domains.

Security. Security should provide protection for medical information in transit and at rest, so that data confidentiality, integrity, and availability can be guaranteed. For data in transit, currently, Transport Layer Security (TLS) protocol can be used to guarantee the security of data transfer and network communication. For data at rest, cryptography primitives such as data encryption, digital signature, and access control mechanisms can ensure secure data access in a single domain. However, how to enforce cross-domain access control and secure sharing of medical data in a statewide or even nationwide scale still remains a challenging task.

Privacy. Privacy is a closely-related concept to security but has its own concentrations, i.e., it assures that personal information are collected, used and protected legally. The privacy compliance regulations require all electronic Protected Health Information (ePHI) related activities, across the entirety of data storage, transfer, and provision, to consistently abide by security and privacy rules.

*Peilong Li is the correspondent author.

SDN-NFVSec '19, March 27, 2019, Richardson, TX, USA

ACM ISBN 978-1-4503-6179-8/19/03…$15.00
https://doi.org/10.1145/3309194.3309198

Generally, the difficulty primarily lies in that security and privacy of healthcare information should be protected not only from external attackers, but also from unauthorized access inside the network or system, e.g., system or service administrators. According to a 2014 study[3], over 50% security breaches occur in the medical industry, and with up to 90% healthcare organizations having exposed their data. Therefore, new methods, architectures, or computing paradigms may be needed to address security and privacy problems in medical data sharing area.

2 BACKGROUND

2.1 Regulatory Compliance Requirements on Security and Privacy

Table 1: Technical Safeguard Requirements

Standards	Implementation Specifications
Access Control	Unique User Identification: identify and track user identity
	Emergency Access Procedure: procedures for obtaining necessary ePHI during an emergency (privilege endorsement)
	Encryption and Decryption: a mechanism to encrypt and decrypt ePHI
Audit controls	record and examine activity that contains ePHI
Integrity	Mechanism to protect ePHI from unauthorized alteration
Person/Entity Authentication	Implement procedures to verify that a person or entity seeking access to ePHI is the one claimed
Transmission Security	Integrity Controls: the security measures to ensure that electronically transmitted ePHI is not improperly modified without detection until disposed of
	Encryption: a mechanism to encrypt ePHI whenever deemed appropriate.

HIPAA and HITECH Act [2] extended security and privacy requirements to business associates. These guidelines stipulate that all necessary measures are in place to keep patient data secure whenever it is accessed, saved, or shared. Lack of compliance to the HIPAA security standards could lead to significant fines and, in some cases, loss of medical licenses.

Table 1 lists a collection of technical safeguard standards along with implementation specifications. From the table, we can see that the HIPAA regulation covers almost every aspect of security. Besides basic requirements such as confidentiality, integrity and authentication in traditional information security, access control with identity tracking and emergency access, and activity audit are also included. This implies that the secure management of healthcare data is a hybrid approach which requires various mechanisms and technical means to be adopted to implement different security and privacy targets.

2.2 Blockchain and Smart Contract

Since the emergence of Bitcoin[14] in 2009, blockchain technology wins a wide reputation in decentralized computing. In essence, blockchain can be viewed as a decentralized, public and immutable ledger where transactions are stored in chained blocks without the existence of a trusted central authority. Many cryptographic means, e.g., Merkle hash tree, chained hash, and digital signatures, are adopted in blockchain to guarantee its security. Moreover, consensus protocol runned behind the peer-to-peer network guarantees its immunity to single-point-of-failures.

Blockchain can be categorized into two types: permissionless and permissioned. Permissionless (also known as public) blockchains allow every user to participate in the network by creating and verifying transactions and adding entries (blocks) to the ledger. While permissioned (also known as consortium) blockchains act like a closed ecosystem, which maintain an access control layer to allow certain actions to be performed in a central controlled manner. In essence, permissioned blockchains trade decentralization to some extent for access control based central governing and the consensus protocols define the trade-offs of decentralization between these two types of blockchains.

Bitcoin is the most well-known example of permissionless blockchains. It applies a Proof of Work (PoW) algorithm to ensure network consensus [14]. While Ethereum, the successor of Bitcoin, uses a combination of Proof of Work and Proof of Stake [7]. Both strategies require participating nodes to add blocks at a certain cost, either at the expense of computation or capital. The script language embedded in Bitcoin is not Turing-complete, hence it is difficult to extend Bitcoin to support various applications. It was not until 2015 when Ethereum pioneered to instantiate the "Smart Contract" concept that it becomes a reality to build various decentralized applications upon blockchain. Smart contracts are small-size computer programs running atop blockchain that automatically execute whenever certain conditions are met.

Hyperledger [6] is an open source collaborative effort aimed at advancing cross-industry blockchain technologies. Hyperledger Fabric is one of the most influential projects in Hyperledger that is widely adopted. Hyperledger Fabric adopts BFT-SMART state machine replication algorithm, a variant of practical byzantine fault tolerance (PBFT) consensus algorithm, as its consensus protocol. Hyperledger provides the opportunity to broaden the scope of blockchain technology beyond cryptocurrency transactions to other fields including the healthcare data management focused on by this paper.

3 BRIEF SURVEY ON MEDICAL DATA SHARING

3.1 Cloud Based Approaches

Since the emergence of cloud computing, secure data sharing in a distributed setting has long been a hot and challenging topic. Considering the fact that users and cloud providers usually belong to different administrative or security domains, the difficulty of cloud based data sharing lies in how much trust users can place on cloud service providers. As for cloud based medical data management, the problem becomes even more complicated.

Currently, there are some cloud service providers (CSP), e.g., Amazon, Google, and Microsoft, proposed HIPAA compliant cloud service [4] for medical information management. A basic requirement of these HIPAA cloud solutions is storing

encrypted medical information. However, the dilemma is the key management problem. Leaving the key management to users will certainly enhance data security, but it can also be a troublesome burden for users and limits the scalability of data sharing among a large scale of research institutes. On the other hand, asking cloud providers to control the keys will potentially increase the risks of data leakage since cloud administrators have the chance to tamper the keys and even decrypt the data.

Other research focus on adopting advanced cryptographic primitives to secure medical data sharing based on cloud storage platforms. Li et al. [11] proposed to use attribute-based encryption (ABE) for secure sharing of personal health records stored in semi-trusted cloud servers. They divide security domains into public domains (physicians and medical researchers) and personal domains(family members and friends), where two kinds of ABE schemes, e.g., a revocable key-policy ABE scheme and a multi-authority ABE scheme, are adopted to address data sharing in above mentioned domains respectively. However, despite patients' full control of their medical information, the scheme poses too much burden on patients, since the patient side applications have to generate and distribute corresponding keys to authorized users.

Guo el al. [9] proposed to combine blockchain with a multi-authority attribute-based signature scheme to secure the storage and access of electronic health records. However, their scheme encapsulates and stores health records in on-chain blocks, which limits its scalability since the size of on-chain stored data has a great impact on the network throughput.

Finally, it should be noted that data interoperability is a big issue in cloud environments due to the incompatibility of various HIPAA cloud providers. Considering a medical data sharing scheme in a regional or nationwide scale, security and privacy protection mechanisms among the participating cloud providers will face tremendous challenges to interface with each other.

3.2 Blockchain Based Approaches

Recently, with blockchain technology being a widespread trend in distributed computing, many researchers consider to use blockchain to secure medical data sharing and management.

Zyskind et al. [19] proposed to use blockchain to provide secure and privacy-preserving data sharing among mobile users and service providers, where two types of transactions are designed, i.e., transaction T_{data} is used for data storage and retrieval, and transaction T_{access} is used for access control. MedRec [5] firstly proposed a decentralized EMR management system based on blockchain technology and provided a functional prototype implementation. It designed three kinds of Ethereum smart contracts to associate patients' medical information stored in various healthcare providers to allow third-party users to access the data after successful authentication.

Yue X. et al. [18] proposed a heealthcare data gateway, a blockchain based architecture integrate with a purpose-centric access control policy to let patients to own, control and share their medical information without violating privacy. Q. Xia el al. proposed BBDS [17], a high-level blockchain based framework that permits data users and owners to access medical records from a shared repository. But their sharing are limited to invited and verified users. Then the authors proposed MedShare [16], a blockchain based framework for medical data sharing that provides data provenance, auditing and control in cloud repositories among healthcare providers.

K. Fan et al. proposed MedBlock [8], a hybrid blockchain based architecture to secure electronic medical records by storing encrypted summary data and data hashes on-chain to enforce access control and preserve data privacy, where a variant of Practical Byzantine Fault Tolerance (PBFT) protocol is adopted as the consensus protocol. However, their proposal of using asymmetric encryption algorithms to encrypt medical information may degrade system performance considering the encryption/decryption overhead of asymmetric encryption.

Is it worth noting that these research focus on storing plain-text medical data on local databases and rely on the blockchain network and smart contracts to implement the business logic of data management. However, storing plain-text medical records in a hospital's database undoubtedly will increase leakage risks, rendering the prior approaches ineffective. An obvious fact is that an internal IT staff member can easily compromise the data, which makes data confidentiality difficult to guarantee. In this context, we believe that encryption of medical data and secure key storage are necessary to enhance the security and privacy of medical information.

3.3 Software-Defined Networking Based Healthcare

Software-defined networking (SDN), with its capability of decoupling data and control planes, can provide centralized network provisioning and management, accelerate service delivery and provide more agility. Thus it wins wide attention in network based data management systems.

P. Li et al. proposed CareNet[12, 13], a regulation compliant framework for home-based healthcare, where software-defined infrastructure are adopted at the network edge to filter and secure health information from home nodes, and further to enable a hybrid home-edge-core cloud architecture with high performance and real-time responsiveness for home-based healthcare services. L. Hu et al.[10] proposed a smart health monitoring method with software-defined networking, where a centralized smart controller are designed to manage all physical devices and provide interfaces to data collection, transmission and processing.

Recently, the continuing research is this field has been focusing on seamlessly integrating attribute-based encryption, privacy level classification, blockchain technology, and software-defined networking (SDN) to achieve secure and privacy-preserving sharing of clinical information.

4 FUTURE RESEARCH WORK

4.1 Pros and Cons of Blockchain Based Approaches

From the analysis in 3, it can be understood that cloud computing alone can not solve all security and privacy problems pertaining to clinical information sharing and management. The inborn multi-domain and multi-tenancy characteristics of cloud computing determines its trust lacking situation, which makes the outsourcing of medical information to clouds a difficult choice.

Blockchain, with its appealing characteristics of decentralization, trustlessness, tamper-evidence and traceability, can mitigate the trust loss in cloud scenarios. For blockchain based healthcare data sharing, there are following advantages:

- Agreement of an event or action can be reached without a trusted mediator.
- Due to the fault tolerance nature of underlying p2p network, it is immune to single point of failure.
- Data stored on chain is complete, consistent, tamper-evident and traceable.
- Blockchain can securely connect various geo-scattered healthcare providers and help to bridge the gap between healthcare domains and make data interoperability an easy task to implement.

However, blockchain is not a panacea to address all existing problems in medical data area. Secure sharing of medical data involves patients, healthcare-providers, HIPAA compliant clouds, and third-party researchers. Besides providing strict access control and secure data storage and provision required by privacy and security regulations of HIPAA, medical data sharing should further consider how to cope with the compatibility problem caused by different or even contradictory data privacy laws in various states and nations. Moreover, when adopting blockchain for healthcare data sharing, some key features need further investigation.

- Once stored on the blockchain, data can not be altered or deleted. However, GDPR in Europe has strengthened the rights of patients to erase their personal information.
- Most data has its life cycle, which makes it unnecessary to store these data permanently. This is also enforced by many data privacy protection laws.
- Blockchain originally is designed to record small size transactions, e.g., in Bitcoin, the block size is limited to one megabyte, which is insufficient for medical data storage such as X-ray images.
- Medical records are usually organized in relational databases for query, while blockchain stored information is in a format of linked blocks, which is inconvenient for relational query.

In this sense, storing all medical information on-chain is not an optimal solution. Instead, on-chain storage should only contain critial metadata of small size while the large amount of clinical information can be stored off-chain in secured local medical databases. The approaches in [5, 8, 16] chose to store medical information off-chain so that they could be secured, modified and deleted as necessary; while data query strings and hashes were stored on-chain for authenticity and integrity verification.

4.2 Possible Future Research

Blockchain based medical information sharing is an ongoing area, which entails a vast amount techniques to cooperate to achieve the security and privacy requirements implied by HIPAA regulations. It is not possible to rely on blockchain technology itself to solve all security problems. Instead, new access control policies and privacy protection mechanisms may be designed and seamlessly integrated with blockchain. Here, we briefly discuss the possible research and technologies that may appear in the future.

Advanced access control schemes are needed. Medical records usually contain multiple fields with different sensitivity. Hence, traditional all-or-nothing encryption schemes are insufficient to address such a requirement. Instead, we need fine-grained access and privacy control policies with rich semantics. Attribute-based encryption is a feasible solution. However, current ABE schemes cause significant overhead on user revocation. How to improve this problem has a great impact on its applications.

Search efficiency on encrypted medical data needs improvement. For encrypted medical information, a challenge would be the query efficiency considering the multi-domain collaboration in medical industry. Searchable symmetric encryption (SSE) [15] can enforce keyword search on encrypted data, which avoids the decryption process thereby enhances query efficiency. However, how to efficiently make a query and get a aggregated query result from scattered and independently managed databases still remains a question.

Software-defined networking is needed to facilitate domain management and collaboration. The SDN controller provides a central point of control to distribute policy information. However, centralizing control into one entity has the disadvantage of creating a central point of attack. Moreover, the programmability associated with the SDN platform adds its security risks. Therefore, how to properly and securely implement a SDN controller without bringing too much complexities is of much significance to SDN based healthcare data systems.

5 CONCLUSION

Medical information sharing without violating security and privacy regulation has long been a challenging topic. This paper reviews related solutions in this area, including cloud computing based approaches, blockchain based approaches, and SDN based approaches. We observed that security and privacy protection of medical information covers security in transit, confidentiality and integrity of data at rest, identity authentication and access control. Therefore, a practical approach for medical data sharing may need to integrate many different schemes and techniques to achieve its design goals.

As a new computing paradigm, blockchain has its advantages over traditional technologies. However, as we have analyzed in this paper, it is important to choose the right type of blockchain (permissoned or permissonless) for medical data sharing. This paper points out some possible future improvements for medical data management based on blockchain approaches.

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for their reviews and insightful suggestions to improve this paper. This work is partially supported by the National Science Foundation of USA (Award No. 1547428, No. 1738965 and No. 1450996).

REFERENCES

[1] 2016. General Data Protection Regulation. https://eugdpr.org/the-regulation/.
[2] 2017. Summary of the HIPAA Security Rule. https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/.
[3] 2018. https://www.healthcareitnews.com/projects/biggest-healthcare-data-breaches-2018-so-far.
[4] 2018. Architecting for HIPAA Security and Compliance on Amazon Web Services. https://d1.awsstatic.com/whitepapers/compliance/AWS-HIPAA-Compliance-Whitepaper.pdf.
[5] Asaph Azaria, Ariel Ekblaw, Thiago Vieira, and Andrew Lippman. 2016. Medrec: Using blockchain for medical data access and permission management. In *Open and Big Data (OBD), International Conference on*. IEEE, 25–30.
[6] Christian Cachin. 2016. Architecture of the hyperledger blockchain fabric. In *Workshop on Distributed Cryptocurrencies and Consensus Ledgers*, Vol. 310.
[7] Ethereum. 2014. Proof of Stake FAQ. https://github.com/ethereum/wiki/wiki/Proof-of-Stake-FAQ.
[8] Kai Fan, Shangyang Wang, Yanhui Ren, Hui Li, and Yintang Yang. 2018. MedBlock: Efficient and Secure Medical Data Sharing Via Blockchain. *Journal of medical systems* 42, 8 (2018), 136.
[9] Rui Guo, Huixian Shi, Qinglan Zhao, and Dong Zheng. 2018. Secure attribute-based signature scheme with multiple authorities for Blockchain in electronic health records systems. *IEEE Access* 776, 99 (2018), 1–12.
[10] Long Hu, Meikang Qiu, Jeungeun Song, M Shamim Hossain, and Ahmed Ghoneim. 2015. Software defined healthcare networks. *IEEE Wireless Communications* 22, 6 (2015), 67–75.
[11] Ming Li, Shucheng Yu, Yao Zheng, Kui Ren, and Wenjing Lou. 2013. Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. *IEEE transactions on parallel and distributed systems* 24, 1 (2013), 131–143.
[12] Peilong Li, Chen Xu, Yan Luo, Yu Cao, Jomol Mathew, and Yunsheng Ma. 2017. Carenet: Building a secure software-defined infrastructure for home-based healthcare. In *Proceedings of the ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization*. ACM, 69–72.
[13] Peilong Li, Chen Xu, Yan Luo, Yu Cao, Jomol Mathew, and Yunsheng Ma. 2017. CareNet: building regulation-compliant home-based healthcare services with software-defined infrastructure. In *Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2017 IEEE/ACM International Conference on*. IEEE, 373–382.
[14] Satoshi Nakamoto. 2009. Bitcoin: A peer-to-peer electronic cash system. http://bitcoin.org/bitcoin.pdf.
[15] Geong Sen Poh, Ji-Jian Chin, Wei-Chuen Yau, Kim-Kwang Raymond Choo, and Moesfa Soeheila Mohamad. 2017. Searchable symmetric encryption: designs and challenges. *ACM Computing Surveys (CSUR)* 50, 3 (2017), 40.
[16] Qi Xia, Emmanuel Boateng Sifah, Kwame Omono Asamoah, Jianbin Gao, Xiaojiang Du, and Mohsen Guizani. 2017. MeDShare: Trust-less medical data sharing among cloud service providers via blockchain. *IEEE Access* 5 (2017), 14757–14767.
[17] Qi Xia, Emmanuel Boateng Sifah, Abla Smahi, Sandro Amofa, and Xiaosong Zhang. 2017. BBDS: Blockchain-based data sharing for electronic medical records in cloud environments. *Information* 8, 2 (2017), 44.
[18] Xiao Yue, Huiju Wang, Dawei Jin, Mingqiang Li, and Wei Jiang. 2016. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. *Journal of medical systems* 40, 10 (2016), 218.
[19] Guy Zyskind, Oz Nathan, et al. 2015. Decentralizing privacy: Using blockchain to protect personal data. In *Security and Privacy Workshops (SPW), 2015 IEEE*. IEEE, 180–184.