Blockchain for healthcare records: a data perspective

Mian Zhang[1] **and Yuhong Ji**[2]

[1] **Department of Computer Science and Information Technology, College of Art and Science, Roosevelt University**
[2] **College of Education and Human Development, Texas A&M University**

Corresponding author:
Mian Zhang[1]

Email address: mzhang1@mail.roosevelt.edu

ABSTRACT

A problem facing healthcare record systems throughout the world is how to share the medical data with more stakeholders for various purposes without sacrificing data privacy and integrity. Blockchain, operating in a state of consensus, is the underpinning technology that maintains the Bitcoin transaction ledger. Blockchain as a promising technology to manage the transactions has been gaining popularity in the domain of healthcare. Blockchain technology has the potential of securely, privately, and comprehensively manage patient health records. In this work, we discuss the latest status of blockchain technology and how it could solve the current issues in healthcare systems. We evaluate the blockchain technology from the multiple perspectives around healthcare data, including privacy, security, control, and storage. We review the current projects and researches of blockchain in the domain of healthcare records and provide the insight into the design and construction of next generations of blockchain-based healthcare systems.

INTRODUCTION

Bitcoin was originally proposed as a peer-to-peer electronic cash system without relying on the trust of any third parties [1]. Bitcoin is a crypto currency that has been recently adopted due to its tamper-proof and decentralized properties. One of the major contributions by Bitcoin system is that it presents a new way to develop open networks. Anything transactions on the blockchain is a function of the network as a whole. Crypto networks represent a fundamental change in the way our society interacts with each other. The intrinsic characteristic of decentralization in blockchain allows information to propagate in a peer-to-peer manner. The decentralized principle is consistent in all aspects of a blockchain system including data access, storage, and dissemination.

Blockchain, as the underlining technology for bitcoin and many other cryptocurrencies, can be considered as a number of networking-connected computers create a permanent, public record of every single transaction that has ever occurred using cryptographic techniques. Blockchain is a promising technology to manage the transactions, contracts, and agreements which is essential to the financial and legal systems in this world [2]. Blockchain is defined as a public, permanent, append-only, distributed ledger [3]. Blockchain has a variety of advantages due to its decentralized property. The development of Blockchain technology today is considered as an equivalent of the TCP/IP during the 90s. It is more secure to store information in face of integrity-based attacks due to its tamper-proof characteristic. Blockchain provides the possibility for trust to be hard-coded into the process of collaborative research and development in an unprecedented manner.

There are currently two types of blockchain systems: public and private. They are also called permissioned and permissionless blockchain. An essential mechanism that the fuels a blockchain system is that consensus algorithm running across the network that validates each transaction. As compared to

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26942v1 | CC BY 4.0 Open Access | rec: 17 May 2018, publ: 17 May 2018

the computing-intensive proof-of-work and proof-of-stake validation in a public blockchain, there is no mathematical guarantee of irreversibility of transaction in a private blockchain, since the consensus relies on the benevolent actors in the blockchain. In a private blockchain, blocks can be validated using other consensus algorithm such as Juno [4], which makes it less secure in term of data integrity. There are two major issues of public blockchain that hinders it as a large-scale business solutions: scalability and privacy. The transaction throughput in a public blockchain is several orders of magnitude slower than the requirements for the real business product solutions.

The healthcare records are fragmented, distributed over different healthcare institutions in a chaotic way. In 2015, 140 million patient records were breached according to Protenus Breach Barometer report. Furthermore, medical error is the third leading cause of death in the US [5]. Medical errors are estimated to be the third leading cause of death for Americans. A patient's medical history is fragmented pieces dispersed and locked within multiple providers and organizations. The current state of healthcare records is disjointed and fragmented due to a lack of common protocols, architectures, and standards. With all above issues considered, blockchain could potentially be one of the most promising solutions to them. With a blockchain-based protocol, every modification regarding a patient's record would be verified, linked with other transactions and added as a block to a chain of blocks. To have patients really own their data, it requires a sophisticated access-control model. To better understand how blockchain could potentially help patients to better control and access EHR (Electronic Healthcare Records), we need to answer the following questions:

1. What is the life cycle of the EHR?
2. Who really owns the EHR?
3. How to manage and store gigantic size of personal data?
4. How to make data compatible considering data model could potentially change over time?

In the next section, we evaluate the blockchain technology in multiple dimensions and how blockchain could compensate other solutions in healthcare records systems. We address the problem primarily from the perspective of data.

BLOCKCHAIN FOR HEALTHCARE RECORDS

A blockchain system can be considered as a virtually incorruptible cryptographic database where critical medical information could be recorded. The system is maintained by a network of computers, that is accessible to anyone running the software. Blockchain operates as a pseudo-anonymous system that has still privacy issue since all transactions are exposed to the public, even though it is tamper-proof in the sense of data-integrity. The access control of heterogeneous patients' healthcare records across multiple health institutions and devices needed to be carefully designed. Blockchain itself is not designed as the large-scale storage system. In the context healthcare, a decentralized storage solution would greatly complement the weakness of blockchain in the perspective.

Date Privacy and Security

The blockchain network as a decentralized system is more resilient in that there is no single-point attack or failure compare to centralized systems [6]. However, since all the bitcoin transactions are public and everybody has access, there already exist analytics tools that identify the participants in the network based on the transaction history [7]. With popularity analytics, similarity or closeness among topics within large volume of data can be detected.. Groups of items or topics can be system generated using closeness relationship formulation [7]. As information flows among different nodes in bitcoin network, Bitcoin transaction is slow due to the fact that information needs to be propagated across the network to synchronize the ledger replicas. The slow dissemination of information exposes a potential security hole for the malicious attacks. Some measures have been implemented to mitigate the number of the blockchain forks in the network by 50%. However, a long-term solution is still

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26942v1 | CC BY 4.0 Open Access | rec: 17 May 2018, publ: 17 May 2018

needed [8]. Like any other networks, Bitcoin network is no exception when it comes to malicious attacks. One of the notable form of attack against Bitcoin network topology is eclipsing attack by using information propagation knowledge [8]. Bitcoin peer-to-peer network topology can be inferred and utilized by malicious attackers to perform precise attacks such as eclipsing attack. By observing the flooding process of the information flow, a flooding network's topology can be inferred. A network topology inference method has been proposed along with a proof of concept in real network [9]. The critical players of bitcoin transactions can be identified use various network centrality metrics [10–19].

Blockchain might replace conventional methods of keeping track of valuable information such as contracts, intellectual-property rights, and corporate accountings. Personal healthcare records need to be protected with the highest standard. With the increasing number of data breach incidents in the past several years, the awareness of the general public about the personal data privacy will continue to increasing. The necessity for data privacy will grow stronger with an increasing number of services and device collecting our personal data associated with our personal identity. There are techniques that obfuscate the linking of pseudonymous address and the real person such as CoinJoin [20].

Data Access and Control

In terms of patients' full control of their health record history, there are three major aspects of privacy need to be considered: a. data ownership; b. fine-grained access control; c. data transparency, integrity, and auditability [21]. The decrypted patients' data inevitably resides in the computer of whoever requests the information whenever the patient grants an one-time access. Potentially, the health records of patients would not be deleted unless whoever has accessed the data makes a specific effort to do that. To mitigate the risk of patients' data residing in the machine forever, a specific data viewing client can be implemented so that the data can be set to expire automatically. The next-generation system for healthcare records should not only allow the patients own the data and have full access to all their data, but also they should be able to control who can access the data. There are a lot of players in the healthcare business: insurance provider, medicine provider, health institutions, etc. A fine-grained access-control of healthcare recording is one of the most challenging problems.

Data Storage

One of the biggest challenge of storage personal healthcare record in blockchain is the sheer size of all combined healthcare records. The size of personal health data way larger than the amount of data stored in Bitcoin or most of the public Blockchains [22]. The size of the entire blockchain is close to 160 gigabytes by May 2018, considering that only a small portion of the blockchain is used to store non-metadata. Bitcoin is not designed to store data. It only allows write small of data information in each block. Some blockchain solutions that try to store healthcare information in the chain fail to recognize this issue. One approach that could mitigate the size of single blockchain is by creating a different chain for each disease type [23], similar to the concept of database sharding. However, this does not really solve the problem and also introduce complexity of fetching data from multiple blockchains.

A feasible solution would be storing the large size of medical data in a separate off-chain storage system [24] and store the hash reference to the storage in the chain [25]. To have a really decentralized system, it would be ideal for the storage system to operate in a decentralized manner. The decentralized storages solution include IPFS (The InterPlanetary File System) which is a content addressed, versioned, peer-to-peer file system [26]. An important category of medical data is the medical image. A cross-domain image sharing framework has been proposed [27].

CONCLUSIONS AND FUTURE WORK

Blockchain has its own issues and challenges. However, it is a promising technology to provide an open and secure access to healthcare data. Blockchain is not mature enough for large-scale business deployment yet in terms of general public education and technology. Many barriers need to be overcome before blockchain can be fully adopted as a mainstream technology for real-world business solutions. Aside from the issues inside of the blockchain technology, a lack of standards, the need for off-chain

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26942v1 | CC BY 4.0 Open Access | rec: 17 May 2018, publ: 17 May 2018

development of open system interfaces, and having intermediaries who currently profit from controlling the data to interface with blockchains are all impediments to the adoption of blockchain system..

Most people realize and understand the hard problems in current healthcare industry. However, blockchain is no elixir. To power blockchain applications, we need to continuously improve low-level blockchain protocols that are equivalent to the backbone of current internet services such as TCP/IP and HTTP. We also need to improve trust and decision-making process in blockchain so that the system is more scalable with higher transaction throughput without sacrificing the security. One way that has been used is that blockchain network could assign more weight into trusted nodes to expedite the computation of the block.

Healthcare data has its value, particularly with a tremendous size of data pool. Blockchain token can be used to incentivize the data donors. It is important to have a incentive mechanism as a built-in feature of the system to power the functioning of the chain. Build a application that allow patients to donate their data in an anonymous way. The pricing model for various categories of data needs to be figured out.

Another interesting topic is the communications across multiple blockchain systems. It is foreseeable that there will be multiple blockchain system even for healthcare system that might interact with each other.

REFERENCES

[1] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system.

[2] Marco Iansiti and Karim R Lakhani. The truth about blockchain. *Harvard Business Review*, 95(1):118–127, 2017.

[3] Steven Norton. Cio explainer: What is blockchain? *The Wall Street Journal*, 2, 2016.

[4] Christian Cachin and Marko Vukolić. Blockchains consensus protocols in the wild. *arXiv preprint arXiv:1707.01873*, 2017.

[5] Martin A Makary and Michael Daniel. Medical error-the third leading cause of death in the us. *BMJ: British Medical Journal (Online)*, 353, 2016.

[6] Mian Zhang and Yujong Ji. A review of crypto networks. *PeerJ PrePrints*, 2018.

[7] Paul Tak Shing Liu. Medical record system using blockchain, big data and tokenization. In *International Conference on Information and Communications Security*, pages 254–261. Springer, 2016.

[8] Christian Decker and Roger Wattenhofer. Information propagation in the bitcoin network. In *Peer-to-Peer Computing (P2P), 2013 IEEE Thirteenth International Conference on*, pages 1–10. IEEE, 2013.

[9] Till Neudecker, Philipp Andelfinger, and Hannes Hartenstein. Timing analysis for inferring the topology of the bitcoin peer-to-peer network. In *Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), 2016 Intl IEEE Conferences*, pages 358–367. IEEE, 2016.

[10] Dongsheng Zhang. Resilience enhancement of container-based cloud load balancing service. Technical report, PeerJ Preprints, 2018.

[11] Dongsheng Zhang. *Resilience Evaluation and Enhancement in Mobile Ad Hoc Networks*. PhD thesis, University of Kansas, 2015.

[12] Dongsheng Zhang and James P.G. Sterbenz. Modelling critical node attacks in MANETs. In *Self-Organizing Systems*, volume 8221 of *Lecture Notes in Computer Science*, pages 127–138. Springer Berlin Heidelberg, 2014.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26942v1 | CC BY 4.0 Open Access | rec: 17 May 2018, publ: 17 May 2018

[13] Dongsheng Zhang and James P. G. Sterbenz. Analysis of Critical Node Attacks in Mobile Ad Hoc Networks. In *Proceedings of the 6th IEEE/IFIP International Workshop on Reliable Networks Design and Modeling (RNDM)*, pages 171–178, Barcelona, Spain, November 2014.

[14] Dongsheng Zhang, Santosh Ajith Gogi, Dan S. Broyles, Egemen K. Çetinkaya, and James P.G. Sterbenz. Modelling Wireless Challenges. In *Proceedings of the 18th ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*, pages 423–425, Istanbul, August 2012. Extended Abstract.

[15] Dongsheng Zhang, Santosh Ajith Gogi, Dan S. Broyles, Egemen K. Çetinkaya, and James P.G. Sterbenz. Modelling Attacks and Challenges to Wireless Networks. In *Proceedings of the 4th IEEE/IFIP International Workshop on Reliable Networks Design and Modeling (RNDM)*, pages 806–812, St. Petersburg, October 2012.

[16] Dongsheng Zhang and James P. G. Sterbenz. Measuring the Resilience of Mobile Ad Hoc Networks with Human Walk Patterns. In *Proceedings of the 7th IEEE/IFIP International Workshop on Reliable Networks Design and Modeling (RNDM)*, Munich, Germany, October 2015.

[17] Dongsheng Zhang and James PG Sterbenz. Robustness Analysis and Enhancement of MANETs using Human Mobility Traces. *Journal of network and systems management*, 24(3):653–680, 2016.

[18] Dongsheng Zhang and James P. G. Sterbenz. Robustness analysis of mobile ad hoc networks using human mobility traces. In *Proceedings of the 11th International Conference on Design of Reliable Communication Networks (DRCN)*, Kansas City, USA, March 2015.

[19] Dongsheng Zhang and James P. G. Sterbenz. Measuring the resilience of mobile ad hoc networks with human walk patterns. In *2015 7th International Workshop on Reliable Networks Design and Modeling (RNDM)*, pages 161–168, Oct 2015.

[20] Greg Maxwell. Coinjoin: Bitcoin privacy for the real world. In *Post on Bitcoin forum*, 2013.

[21] Guy Zyskind, Oz Nathan, et al. Decentralizing privacy: Using blockchain to protect personal data. In *Security and Privacy Workshops (SPW), 2015 IEEE*, pages 180–184. IEEE, 2015.

[22] John D Halamka and Ariel Ekblaw. The potential for blockchain to transform electronic health records. *Harvard Business Review*, 3, 2017.

[23] MediChain The Medical Data Blockchain. https://medichain.online/wp-content/uploads/2018/03/medichain-ico-whitepaper.pdf.

[24] Eberhard Scheuer. Health information traceability foundation, 2018.

[25] A Park, J Mullin, A Mah, M Gallo, L Cyca, D Bacinello, and C Manning. The blockchain for personalized medicine. 2017.

[26] Juan Benet. Ipfs-content addressed, versioned, p2p file system. *arXiv preprint arXiv:1407.3561*, 2014.

[27] Vishal Patel. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. *Health informatics journal*, page 1460458218769699, 2018.