

Blockchain-based Bidirectional Updates on Fine-grained Medical Data

Chunmiao Li^{1,3}, Yang Cao², Zhenjiang Hu^{1,3,4}, Masatoshi Yoshikawa²

¹ National Institute of Informatics, Japan ² Kyoto University, Japan

³ SOKENDAI (The Graduate University for Advanced Studies), Japan ⁴ University of Tokyo, Japan

Abstract—Electronic Medical data sharing between stakeholders, such as patients, doctors, and researchers, can promote more effective medical treatment collaboratively. These sensitive and private data could only be accessed by authorized users. Given a total medical data, users may care parts of them and other unrelated information might interfere with the user-interested data search and increase the risk of exposure. Besides accessing these data, users may want to update them and propagate to other sharing peers so that all peers keep identical data after each update. To satisfy these requirements, in this paper we propose a medical data sharing architecture that addresses the permission control using smart contracts on blockchain and splits data into fine-grained pieces shared with different peers then synchronize complete data and these pieces with bidirectional transformations. Medical data reside on each user's local database and permission-related data are stored on smart contracts. Only all peers have gained the newest shared data after updates can they start to do next operations on it, which are enforced by smart contracts. Blockchain-based immutable shared ledger enables users to trace data updates history. This paper can provide a new perspective to view complete medical data as different slices to be shared with various peers but consistency after updates between them are still promised, which can protect privacy and improve data search efficiency.

Index Terms—medical data, blockchain, update, bidirectional transformations

I. INTRODUCTION

Now a lot of medical data are digitalized so as to be stored and accessed conveniently. A medical record is produced after a patient goes to see a doctor and often resides on the hospital's database. Medical records usually contain highly sensitive information about patient privacy. HIPPA Privacy Rule [1] in the U.S. regulates the use and disclosure of personally identifiable health information to protect patients' privacy. However, it is hard to make sure that all medical institutes would follow these rules and they may expose patient privacy for profit. Moreover, patients might visit many hospitals and leave their records scattered [2] in different places, so that it is hard for them to manage records efficiently. Patients should better be provided a platform to manage and review their historical medical data in case of exposure or being tampered. What's more, better communication between patients and doctors can contribute to patient adherence [3] and improve health [4]. In addition to provide data to doctors, patients tend to share their medical data with health experts to help understand some complex statistics. Many kinds of research have been done to study how medical specialists with different expertise

collaborate with each other [5]. Researchers can identify public health risks and then develop a better treatment by analyzing existing medical data [6]. Sharing medical data under some constraints could benefit all relating stakeholders such as patients, researchers, and doctors.

To protect patients' data from being exposed or tampered, shared medical data could reside in encrypted formats on a trusted cloud storage server and can only be accessed by authorized users. However, in that case, centralized access control might lead to a single point of failure and become the bottleneck of the sharing system. Some decentralized medical data sharing systems [7]–[9] have been proposed to manage authentication based on blockchain [10] technology, which achieves consensus among distributed nodes. The access control logic of medical data is encoded into smart contracts [7] or Chaincode [11] so that anyone who wants to access medical data should firstly be authorized permission from the blockchain side.

Still, we identified a limitation on current works. Users might be overwhelmed by a complete medical record since they tend to have a unique focus on it. For example, given a medical record, researchers are interested in the mechanism of medicine action, whereas patients care more about the medicine dosage standard. In addition, additional but unnecessary information might influence or even mislead users' judgment. Imagine that doctors may add some symptom description on records which might put patients in confusion and fear [12]. In addition, treatment steps are exclusive to a hospital and can not be directly accessed to other users.

To fill this gap, we propose an idea that a complete medical data (i.e., data source) might be split into lots of smaller pieces (i.e., data views). A user can share different fine-grained data pieces¹ with different users based on predefined protocols. Imagine a doctor can share dosage usage with a patient and medicine mechanism with a researcher respectively. In this way, only users related data are exposed to them, which can avoid additional data interference and protect private data from being leaked.

However, if we adopt this idea, we have to dispose of the synchronization between source and multiple views when updates on shared data occur. Consider this scenario: a researcher

¹Our work assume that the initialization of shared data has been finished. We only consider management on existing shared data.

updates the medicine mechanism on shared data with a doctor. Note here the shared data is actually a view from the complete medical data (source) on the doctor side. Thus he needs to synchronize this change on view to create a new source.

To solve this problem, we apply bidirectional transformations [13] to synchronize them after updates on either one side. For example, we can invoke *put* direction of a BX program to reflect modifications on shared data in complete data and *get* to produce shared data from complete data. Because different views produced from the same source might have overlapped data. The doctor still has to judge whether he needs to modify the shared data with patients by reproducing a new view.

Moreover, we encode permission information of shared data into smart contracts. Shared data management should be conducted after the peer has been authorized.

In this paper, our contributions are as follow.

- 1) We proposed to partition a complete medical record to multiple more fine-grained data pieces shared with different peers and apply bidirectional transformations to synchronize a record and these multiple pieces.
- 2) We designed a decentralized medical data sharing architecture where data reside on user's local database and metadata are stored on smart contracts of blockchain to control permission for managing data.
- 3) We sketched procedures for data management (i.e., Create, Read, Update, Delete) operations on shared data.

The remainder is organized like this. Section II gives some preliminaries about blockchain and bidirectional transformations. Section III sketches our system design and provides an implementation architecture. Section IV discusses identified threats and proposes countermeasures. Section V compares our work with existing ones to clarify our improvement over them. Section VI concludes and directs our future work.

II. PRELIMINARIES

A. Blockchain

Proposed with Bitcoin [10] in 2008, blockchain technology has been widely used in many fields. Blockchain provides a solution for data storage, data transfer and consensus protocol in a distributed and decentralized environment. Generally speaking, blockchain is a shared ledger and replicated by all nodes on a distributed network, which records the historical valid transactions in chronologically chained blocks. The nodes who generate new blocks by solving a computational puzzle (the proof-of-work problem) are called miners.

Not only can support the platform of cryptocurrency, but blockchain can also be applied to other scenes. Ethereum [14] extend blockchain with additions such as a built-in Turing-complete programming language so that one can use this scripts (i.e., Ethereum Virtual Machine (EVM) bytecodes) to write programs (i.e., smart contracts²) on the blockchain. We can just write Solidity³ programs and then compile it to EVM code. Besides the user accounts controlled by private keys

like in Bitcoin, the accounts for smart contracts are allowed in Ethereum. Anyone can build decentralized applications which consist of a collection of smart contracts. Once a transaction involving smart contract creation gets confirmed, an address is generated for the contract and later anyone can send transactions to this address to execute the programs on it. A smart contract transaction is enforced when a miner includes it in a new produced block. Other nodes will validate it and re-run contracts if it is valid.

B. Bidirectional transformations

Maintaining consistency between different data representations having overlapping contents is important [15]. For example, in databases, a view table can be produced by querying a base source table; this view table can be modified, in which case we will want to “restore consistency”, i.e., we need to change the source such that the modified view coincides with the result of the query on the changed source — this is the well-known view update problem [16]. To achieve this, one may consider providing two separate programs to represent the two directions to propagate updates from one side to the other. But it is hard to prove that the source and view can still be kept consistent after updates. Bidirectional transformations (BXs) were proposed [17] to solve this.

BX programs⁴ can be invoked in two ways as forward and backward transformations. A forward transformation (denoted as *get*) extracts some information from the source to build an abstract view, and the backward transformation (denoted as *put*⁵) embeds information of the view back into the source and produces an updated source. This pair of transformations should satisfy the *round-tripping* laws (also referred to as *well-behavedness*) called *PutGet* and *GetPut*.

$$\begin{aligned} get(put(\mathbf{source}, \mathbf{view})) &= \mathbf{view} && (PutGet) \\ put(\mathbf{source}, get(\mathbf{source})) &= \mathbf{source} && (GetPut) \end{aligned}$$

Intuitively, *GetPut* states that no update should be performed on the source when there is no change on the view, while *PutGet* hints that *put* should take all updates on the view into account so that the view can be regenerated from the updated source by *get*. The most distinguished point of BX is that a view can contain only a part of a source. With respect to some consistency between a source and a view, BX programs can synchronize the source and view, and their well-behavedness guarantees that the source and view are kept consistent after updates on either side. There are some languages for constructing well-behaved BX programs, such as Boomerang [19], BiGUL [20] and HOBiT [21].

III. SYSTEM ARCHITECTURE

In this section, we first present the way to split a complete medical record into multiple pieces in Section III-A. Then

⁴The BXs we refer to in this paper are asymmetric lenses [18], one of the synchronization models studied by the BX community.

⁵*put* is not a simple inverse of *get*. Instead, it accepts the view and the original source as input and produces an updated source as output.

²Hyperledger and others still provide platforms to write smart contracts.

³<https://solidity.readthedocs.io/en/v0.5.2/>

Section III-B shows the permission encoding on smart contracts. Our system can not only allow updates on shared data but other operations such as creating data, which are described in Section III-C and explained by a concrete case in Section III-D. Lastly, we give our system architecture on Section III-E.

To simplify our expression next, we adopt nodes and users to denote devices connecting to blockchain network and stakeholders (doctors, patients, etc.) in medical scenarios respectively. Users sharing data are called sharing peers.

A. Fine-grained shared data

Shared data⁶ can exist in different peers, as shown in Fig. 1. Suppose there are three users: patient Alice, researcher Charlie, and doctor Bob. Each user has their own complete base table which is named as D1, D2, D3 respectively and different shared data with other users. For example, Bob and Charlie share some data which are stored in D32 on Bob side and D23 on Charlie side respectively. D32 and D23 should contain the same contents, which means either one is updated and the other one need be modified to become identical with it again. Similarly, Alice and Bob share the same contents that stored in D13 and D31 separately in their sides. Notably, the formats and contents of shared data are predefined by sharing peers.

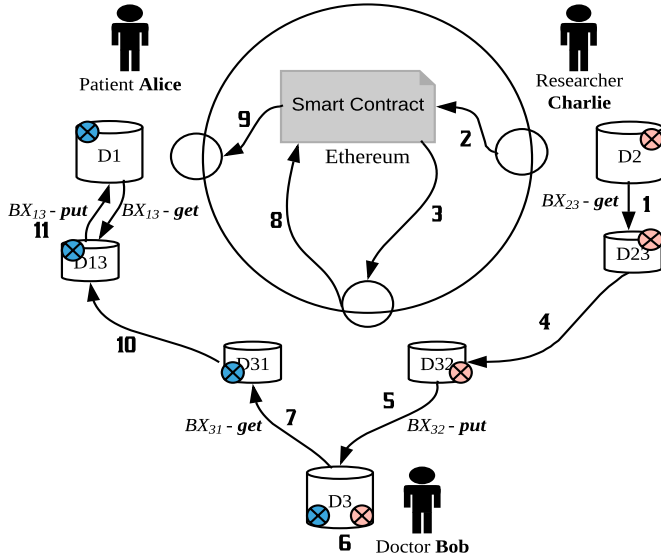


Fig. 1. A workflow for updating data fields of shared data

Figure 2 presents a concrete example with statistics to illustrate this structure. Each user stores a complete medical record and different shared data pieces on their local database. As we said before, BX programs are used to synchronize the complete data and shared data. Shared data can be seen as views which can be produced from complete records named as sources. For example, Table D13 is shared by Alice and Bob and can be produced from D1 by $BX_{13}\text{-get}$ (i.e., applying the

⁶In our prototype, medical data are entered directly by users. Later we may consider using the data from wearable devices.

Patient ID	Medication Name	Clinical Data	Address	Dosage
188	Ibuprofen	CliD1	Sapporo	one tablet every 4h

D1 (Patient **Alice**)

Patient ID	Medication Name	Clinical Data	Mechanism of Action	Dosage
188	Ibuprofen	CliD1	MeA1	one tablet every 4h
189	Wellbutrin	CliD2	MeA2	100 mg twice daily

D3 (Doctor **Bob**)

Patient ID	Medication Name	Clinical Data	Dosage
188	Ibuprofen	CliD1	one tablet every 4h

D13 (also D31)

Medication Name	Mechanism of Action
Ibuprofen	MeA1
Wellbutrin	MeA2

D23 (also D32)

Medication Name	Mechanism of Action	Mode of Action
Ibuprofen	MeA1	MoA1
Wellbutrin	MeA2	MoA2

D2 (Researcher **Charlie**)

Fig. 2. Data distribution

get direction of the BX program between D1 and D13). If D13 is modified, then D1 need to be updated from original D1 and D13 by using $BX_{13}\text{-put}$ (i.e., invoking put direction of the BX program between D1 and D13) to ensure that the modified D13 can be regenerated from the updated D1.

In our design, shared data between any two peers are not exposed to the third party, which can keep data privacy between them to some degree. For example, any operations on D23 or D32 can only be known by Charlie and Bob and Alice have no information about this. However, after the change on D32 are reflected in D3, since D3 has been modified, D31 might need to be regenerated by using $BX_{31}\text{-get}$ on D3.

B. Permission on data management

Figure 3 presents a metadata collection table which dictates the update permission on each attribute of the shared data. These kinds of tables reside in smart contracts on the blockchain. Each metadata entry corresponds to a shared table. For example, the entry for D13 or D31 declares that it is shared by Alice and Bob and Bob can update all attributes value but Alice can only change the clinical data. The “Latest Update Time” shows when the metadata was modified most recently. The value on “Authority to Change Permission” delegates Bob to change other peers’ (here just Alice) authority. For instance, for D13 and D31, Bob can change the permission to update “Dosage” to “Bob, Alice” so that Patient Alice can also update the “Dosage” later.

- Blockchain: keep the manage permission of shared data on smart contracts and notify sharing peers the change on them.
- Database manager: disposes of the synchronization between shared data and local data according to consistency logic relations. These synchronizations are implemented by executing BX programs.
- Database: each user has a complete database and many data pieces shared with other users. The latter (seen as a view) can always be reproduced from the former (seen as a source).

Next, we discuss more details about this architecture.

Firstly, Raw medical data always stay in each peer's local database and data transfer only exist between sharing peers, which avoid data being leaked to the third party so as to keep shared data security. The data can not only be provided by doctors. Instead, each node can be a shared data provider. As referred in [22], many clinics encourage patients to collect data by themselves that are supposed to be gathered by doctors and expect to increase clinic efficiency and promote patient awareness.

Moreover, blockchain's consensus protocol scheme keep the shared data between sharing peers are same after updates since each peer will receive the notification from contracts and request new shared data from other sharing peers. Additionally, any modification on shared data can be recorded on the blockchain. Blockchain properties such as immutability, auditability, and transparency enable nodes to check and review update history on shared data. Still, simultaneously updates to the same shared data by multiple peers are forbidden. Smart contracts dispose of the updates according to received requests in chronological order. If a transaction for updates on shared data has been included in a block, then other requests on this shared data will not be accepted, i.e., one block can contain one transaction at most on some shared data at one time. This can promise that only when all sharing peers have had the newest shared data can they execute further operations.

Lastly, this system architecture can also be applied to other data sharing scenarios.

IV. THREATS AND COUNTERMEASURES

In this section, we identify some threats to our system and propose relating countermeasures.

1) *Throughput*: We employ smart contracts to control access to shared data. As we all know that the block creation time is approximately 12 seconds on Ethereum. We argue that this time interval is acceptable since nodes may choose to collect a lot of updates then send requests to contracts. Usually, it is not so urgent for a patient or doctor get the immediate updated shared data.

2) *Correctness of smart contracts*: Smart contracts might be inconsistent with specifications. We may apply some theorem prover such as Coq [23] to prove or verify the correctness of smart contracts to prevent these attacks.

3) *Public blockchain*: Once deployed to the public Ethereum blockchain, transactions relating to our systems might not be chosen into a block by miners. So a private blockchain might be a better choice for our system.

4) *Incentive*: Like in [24], we don't include any incentive for mining beyond the use of our system. We presume that all nodes on the blockchain already have incentives to keep medical data from being tampered and illegal access or updates.

V. RELATED WORK

In this section, we review existing blockchain-based research on medical data sharing field and state the advantages of our system compared with them.

Zyskind et al. suggested using blockchain for access control in [25] where encrypted data reside on the third party storage. But data might be exposed by this "trusted" the third party so that data privacy is violated.

The idea of introducing Blockchain technology to healthcare was presented firstly in [26] where they use blockchain for data storage to guarantee medical data cannot be modified by anyone. Also, they designed a Healthcare Data Gateway (HDG) to control access to the shared data. However, the medical data size can become huge so that become a burden for blockchain nodes' storage since each node has the same copy of blockchain. Usually, the size of metadata is smaller than data. (It also depends on the structure of metadata and data.) We store metadata on smart contracts so as to reduce the storage pressure for each blockchain node.

MedRec [7] choose to store raw medical data on providers' database and patients can download the data from it after authorized by smart contract on the blockchain. They aimed to enable patients to engage in their healthcare. Whereas in our system, all parties, such as doctors, patients, and researchers can benefit from sharing data with others. MedRec recognized that not all provider data such as physician intellectual property can be exposed to patients [27], [28] so that they don't claim to manage contents automatically from physician's output. In our work, instead, we allow each node to share a piece of medical data not total but still keep consistency between them after the updates to the shared ones. Additionally, any modifications on data shared by two nodes will not be disclosed to the third party which keeps the consistency only exists in sharing peers. Moreover, since all shared data with others can be a part of each nodes' local total databases, we can decide whether one shared data have some influence on the other shared pieces and then propagate this change to the third party.

Dubovitskaya et al. gave architecture to manage and share medical data for cancer patient care [11]. They stored encrypted categorized shared data on the cloud and relating metadata in blockchain and implemented the prototype on Hyperledger [29]. The access control policy is defined in the chaincode Logic by patients. Whereas we think that each data provider, not just patients can use smart contracts to encode the control policy when they deploy them to Ethereum.

Notably, previous three works and others [8], [9], [24], [30], [31] mostly targeted to the access problem on shared data but did not pay much attention to updates on them. Additionally, they presumed that different parties can share the same data. Unlike them, we aim to solve the updates issues on the shared data and allow one party can split total data into multiple pieces (i.e, views) which are shared with different parties but still keep consistency between source and views.

VI. CONCLUSION

Medical data sharing are necessary and important, which allow stakeholders on medical scenarios to contribute their knowledge to better the medical treatment. Users may have different interests in the same complete medical record. Some peers might update some values of fields in the existing data. These updates need to be propagated to sharing peers. Our architecture divides a record into pieces that shared with different users separately, which can protect data privacy by limiting essential data between two peers and reduce the unrelated data interference. Any updates on data pieces can be synchronized to complete records by bidirectional transformations. Moreover, based on smart contracts on the blockchain, we can promise that only authorized users can update the existing shared data and only when all peers have updated to the newest data contents they can continue the operations on shared data.

We are still developing the prototype to implement our idea. In the future, we will use real patient data to do the experiment but use some de-identification technology to protect patient data from being exposed.

REFERENCES

- [1] H. Centers for Medicare & Medicaid Services *et al.*, "Hipa administrative simplification: standard unique health identifier for health care providers. final rule." *Federal register*, vol. 69, no. 15, p. 3433, 2004.
- [2] J. Zhang, N. Xue, and X. Huang, "A secure system for pervasive social network-based healthcare," *IEEE Access*, vol. 4, pp. 9239–9250, 2016.
- [3] K. B. H. Zolnierok and M. R. DiMatteo, "Physician communication and patient adherence to treatment: a meta-analysis," *Medical care*, vol. 47, no. 8, p. 826, 2009.
- [4] R. L. Street Jr, G. Makoul, N. K. Arora, and R. M. Epstein, "How does communication heal? pathways linking clinician–patient communication to health outcomes," *Patient education and counseling*, vol. 74, no. 3, pp. 295–301, 2009.
- [5] G. Fitzpatrick and G. Ellingsen, "A review of 25 years of cscw research in healthcare: contributions, challenges and future agendas," *Computer Supported Cooperative Work (CSCW)*, vol. 22, no. 4–6, pp. 609–665, 2013.
- [6] O. of the National Coordinator for Health Information Technology, "Report to congress: Report on health information blocking," 2015.
- [7] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "Medrec: Using blockchain for medical data access and permission management," in *Open and Big Data (OBD), International Conference on*. IEEE, 2016, pp. 25–30.
- [8] K. Fan, S. Wang, Y. Ren, H. Li, and Y. Yang, "Medblock: Efficient and secure medical data sharing via blockchain," *Journal of medical systems*, vol. 42, no. 8, p. 136, 2018.
- [9] Q. Xia, E. B. Sifah, A. Smahi, S. Amofa, and X. Zhang, "Bbds: Blockchain-based data sharing for electronic medical records in cloud environments," *Information*, vol. 8, no. 2, p. 44, 2017.
- [10] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.
- [11] A. Dubovitskaya, Z. Xu, S. Ryu, M. Schumacher, and F. Wang, "Secure and trustable electronic medical records sharing using blockchain," in *AMIA Annual Symposium Proceedings*, vol. 2017. American Medical Informatics Association, 2017, p. 650.
- [12] T. Delbanco, J. Walker, J. D. Darer, J. G. Elmore, H. J. Feldman, S. G. Leveille, J. D. Ralston, S. E. Ross, E. Vodicka, and V. D. Weber, "Open notes: doctors and patients signing on," *Annals of internal medicine*, vol. 153, no. 2, pp. 121–125, 2010.
- [13] Z. Hu, H. Pacheco, and S. Fischer, "Validity checking of putback transformations in bidirectional programming," in *FM*, 2014, pp. 1–15.
- [14] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger,"
- [15] F. Abou-Saleh, J. Cheney, J. Gibbons, J. McKinna, and P. Stevens, "Introduction to bidirectional transformations," in *Bidirectional Transformations*. Springer, 2018, pp. 1–28.
- [16] F. Bancillon and N. Spyrtatos, "Update semantics of relational views," *ACM Transactions on Database Systems (TODS)*, vol. 6, no. 4, pp. 557–575, 1981.
- [17] K. Czarnecki, J. N. Foster, Z. Hu, R. Lämmel, A. Schürr, and J. F. Terwilliger, "Bidirectional transformations: A cross-discipline perspective," in *International Conference on Theory and Practice of Model Transformations*. Springer, 2009, pp. 260–283.
- [18] J. N. Foster, M. B. Greenwald, J. T. Moore, B. C. Pierce, and A. Schmitt, "Combinators for bidirectional tree transformations: A linguistic approach to the view-update problem," *ACM Transactions on Programming Languages and Systems (TOPLAS)*, vol. 29, no. 3, p. 17, 2007.
- [19] A. Bohannon, J. N. Foster, B. C. Pierce, A. Pilkiewicz, and A. Schmitt, "Boomerang: resourceful lenses for string data," in *ACM SIGPLAN Notices*, vol. 43, no. 1. ACM, 2008, pp. 407–419.
- [20] H.-S. Ko, T. Zan, and Z. Hu, "BiGUL: A formally verified core language for putback-based bidirectional programming," in *Proceedings of the 2016 ACM SIGPLAN Workshop on Partial Evaluation and Program Manipulation*, ser. PEPM '16. New York, NY, USA: ACM, 2016, pp. 61–72. [Online]. Available: <http://doi.acm.org/10.1145/2847538.2847544>
- [21] K. Matsuda and M. Wang, "Hobit: Programming lenses without using lens combinators," in *European Symposium on Programming*. Springer, 2018, pp. 31–59.
- [22] C.-F. Chung, "Using personal informatics data in collaboration among people with different expertise," Ph.D. dissertation, 2018.
- [23] G. Huet, G. Kahn, and C. Paulin-Mohring, "The coq proof assistant a tutorial," *Rapport Technique*, vol. 178, 2004.
- [24] G. G. Dagher, J. Mohler, M. Milojkovic, and P. B. Marella, "Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology," *Sustainable Cities and Society*, vol. 39, pp. 283–297, 2018.
- [25] G. Zyskind, O. Nathan *et al.*, "Decentralizing privacy: Using blockchain to protect personal data," in *Security and Privacy Workshops (SPW), 2015 IEEE*. IEEE, 2015, pp. 180–184.
- [26] X. Yue, H. Wang, D. Jin, M. Li, and W. Jiang, "Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control," *Journal of medical systems*, vol. 40, no. 10, p. 218, 2016.
- [27] U. D. of Health, H. Services *et al.*, "Individuals' right under hipaa to access their health information 45 cfr 164.524," 2017.
- [28] C. Grossman, W. A. Goolsby, L. Olsen, and J. M. McGinnis, "Clinical data as the basic staple of health learning: creating and protecting a public good," *Washington, DC: Institute of Medicine*, 2011.
- [29] Hyperledger, "Hyperledger," 2017.
- [30] J. Liu, X. Li, L. Ye, H. Zhang, X. Du, and M. Guizani, "Bpds: A blockchain based privacy-preserving data sharing for electronic medical records," *arXiv preprint arXiv:1811.03223*, 2018.
- [31] S. Amofa, E. B. Sifah, O.-B. Kwame, S. Abia, Q. Xia, J. C. Gee, and J. Gao, "A blockchain-based architecture framework for secure sharing of personal health data," in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, 2018, pp. 1–6.