IEEE Access
The journal for rapid open access publishing

IEEE Xplore Digital Library
http://ieeexplore.ieee.org/Xplore

A Review of Secure and Privacy-Preserving Medical Data Sharing

HAO JIN [1], YAN LUO [1], PEILONG LI[2] AND JOMOL MATHEW[3]

***Abstract*—In the digital healthcare era, it is of the utmost importance to harness medical information scattered across healthcare institutions to support in-depth data analysis and achieve personalized healthcare. However, the cyberinfrastructure boundaries of healthcare organizations and privacy leakage threats place obstacles on the sharing of medical records. Blockchain, as a public ledger characterized by its transparency, tamper-evidence, trustlessness and decentralization, can help build a secure medical data exchange network. This paper surveys the state-of-the-art schemes on secure and privacy-preserving medical data sharing of the past decade with a focus on blockchain-based approaches. We classify them into permissionless blockchain-based approaches and permissioned blockchain-based approaches, and analyze their advantages and disadvantages. We also discuss potential research topics on blockchain-based medical data sharing.**

***Index Terms*—access control, blockchain, encryption, medical data, privacy, security**

I. Introduction

Data is an asset with value, and particularly so today when cloud computing, big data, and Internet of things are embracing each other. This unprecedented era of technological confluence poses great challenges for data security and privacy. As an example, in 2013, Yahoo experienced a data breach that put the information of over 3 billion users at risk, which is almost half of the entire human population. And this incident is just one example of countless data breach events[1].

Electronic Medical Record (EMR) data, especially Protected Health Information (PHI), suffers from an even greater risk. According to a recent investigation [2], there has been an upward trend in the number of medical records exposed each year. Healthcare data breaches are now happening at a rate of more than one per day. To strengthen medical data governance, privacy protection regulations, such as the Health Insurance Portability and Accountability Act (HIPAA)[3] in the United States or the General Data Protection Regulation (GDPR)[4] in Europe, require data to be stored and shared in a secure and privacy-preserving way and may inflict severe penalties for events of healthcare data breach.

Consequently, to enhance security safeguards and avoid privacy leakage, most healthcare providers and hospitals choose to build their healthcare systems in a closed domain with a defensive perimeter, such as a private network equipped with firewalls and intrusion detection systems. This has created the medical data silos of today that are scattered throughout various healthcare institutions, preventing collaborative healthcare treatment and medical research. On the other hand, the era of cloud computing and big data necessitates that medical data be shared among various users and institutions to allow analysis, so that better healthcare service and new treatment plans can be provided.

In summary, the secure and privacy-preserving sharing of clinical information mainly faces the following obstacles:

1. Massive data increasing at a rapid speed. Medical data such as X-ray images, computed tomography, and genetic data are large in size, and their volumes are increasing at a rate of 20-40 percent every year. In 2015, an average healthcare provider in United States needed to manage 665 terabytes of patient information, 80 percent of which was unstructured medical images. Even worse, it is estimated that big data in healthcare will reach 25,000 petabytes in 2020[5].

The challenges include not only how to store such a massive amount of data with existing IT infrastructure, but also how to ensure its confidentiality and integrity while maintaining high availability among clinicians, medical researchers, and collaborators.

2. Cross-institutional data interoperability. Most existing healthcare systems are built on an enclosed domain with a network defense perimeter to prevent outside attacks and threats. This is a huge hurdle to cross-institutional data sharing due to two reasons:

1) domain cyberinfrastructure and its perimeter impede data access from outside;
2) an independent domain usually has its own data management policy, making it difficult to guarantee the compatibility of any two domains.

The direct consequence of this network defense perimeter is the lack of data interoperability on medical information, which further poses a barrier to medical analytics that require a large amount of clinical information. Moreover, it also creates an inconvenience for patients seeking better treatment plans when their medical records are scattered across multiple hospitals. Here, a healthcare domain refers to an enclosed hospital ecosystem that is built on a private network, where all external access to internal databases and devices are through authenticated connections such as VPN. It is a widely adopted architecture for today's healthcare data management.

Hence, a more holistic and integrated healthcare infrastructure is needed to facilitate the secure sharing and interoperation on medical data among various healthcare domains, and to enable collaborative healthcare service and research.

3. Security and privacy. Security should provide protection for medical data in transit and at rest, with traditional security goals on data confidentiality, integrity, and availability being fulfilled. Currently, the Transport Layer Security (TLS) protocol can be used to guarantee the security of data in transit. For data at rest, cryptography primitives such as data

Digital Object Identifier: 10.1109/ACCESS.2019.2916503

TABLE I: HIPAA 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 disposal
	Encryption: a mechanism to encrypt ePHI whenever deemed appropriate.

encryption, digital signature, and access control mechanisms can ensure secure access in a single domain. However, how to enforce cross-domain access control and secure sharing of medical data in a state-wide or even national scale remains a challenging task.

Privacy is a closely-related concept to security but has its own concentrations, i.e., it assures that personal information is collected, used, and protected legally. For example, 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.

Generally, the difficulty primarily lies in that the security and privacy of healthcare information should be protected not only from external attackers, but also from unauthorized access from within the network or system[6]. Therefore, new methods, architectures, or computing paradigms may be needed to address security and privacy problems in medical data sharing area.

In this paper, we surveyed the state-of-the-art approaches in secure medical data sharing and management. The remainder of the paper is organized as follows. Section II briefly introduces the HIPAA and blockchain background that is necessary to understand the schemes surveyed in following sections. Section III describes the schemes on healthcare data sharing based on cloud computing, cryptography, and blockchain technology. Finally, we point out several potential future directions for blockchain-based approaches.

II. Background

A. Regulatory Compliance Requirements on Security and Privacy

HIPAA and HITECH Act [3] extend 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 I lists a collection of technical safeguard standards along with implementation specifications, where 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, new requirements such as access control with identity tracking and emergency access, and activity auditing 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 incorporated to meet these security and privacy targets.

B. Blockchain and Smart Contract

Since the emergence of Bitcoin[7] in 2009, blockchain technology has garnered a wide reputation in decentralized computing. In essence, blockchain can be viewed as a decentralized, immutable, public ledger where transactions are stored in chained blocks without the existence of a trusted central authority. Many cryptographic primitives (e.g., Merkle hash tree, chained hash, and digital signatures) are adopted in blockchain to guarantee its security.

1. Permissionless and permissioned blockchains.

Generally, blockchain can be categorized into two types: permissionless and permissioned. The difference mainly lies in the consensus protocol executed behind the peer-to-peer network.

Permissionless or public blockchains allow every user to participate in the network by creating and verifying transactions and adding blocks to the ledger. Bitcoin is the most famous example of permissionless blockchains, which applies a Proof of Work (PoW) algorithm to ensure network consensus [8]: a mathematical puzzle needs to be solved for each new mined block. Ethereum, the successor of Bitcoin, uses a combination of Proof of Work and Proof of Stake [9]. Both strategies require participating nodes to add blocks at a certain cost, either at the expense of computation or capital.

In contrast, permissioned or consortium blockchains act more like a closed ecosystem: they maintain an access control layer to allow certain actions to be performed only by certain kinds of nodes. That means nodes in the network are not equal to each other. In essence, they sacrifice some degree of decentralization to regain some centralization so that better control can be enforced to achieve their goals. Hyperledger[10] is an increasingly popular, collaborative permissioned blockchain that aims at advancing cross-industry blockchain technologies. With Fabric being its most influential project, Hyperledger adopts BFT-SMART state machine replication algorithm[11, 12], a variant of the practical byzantine fault tolerance (PBFT)[13, 14] consensus algorithm, as its consensus protocol. Hyperledger provides the opportunity to broaden the scope of blockchain technology beyond cryptocurrency transactions to other fields including healthcare data management.

2. Smart Contract.

The script language embedded in Bitcoin is not Turing-complete, which is implemented with stack-based operations. Hence it is difficult to extend Bitcoin to support various applications. It was not until 2015 when Ethereum[15] pioneered to instantiate the "Smart Contract" concept by designing a rich programming language and enabling it with Turing-completeness. It has become a trend to build various decentralized applications upon blockchain and smart contracts.

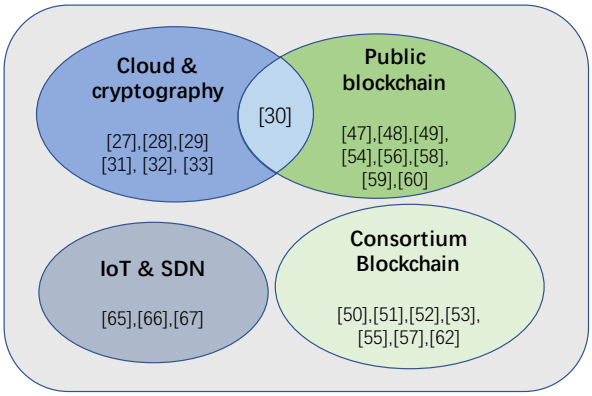

Fig. 1: Classification of the state-of-the-art schemes

Smart contracts are small-size user-defined computer programs that specify rules governing transactions, which run atop blockchain and are enforced by a network of peers. Smart contracts automatically execute whenever certain predefined conditions are met. Currently, Solidity[16] under Ethereum platform and Chaincode[17] under Hyperledger platform are the two most widely used programming languages for smart contracts.

III. Survey on Medical Data Management

A. Cloud and Cryptography in Healthcare

Since the emergence of cloud computing, secure data sharing in a distributed setting has long been a 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. Such a lack of trust stems from the lack of transparency and the loss of data control[18, 19] by users in cloud environments:

- outsourcing data to the cloud is essentially the handover of physical control from one trust domain (local storage) to another trust domain (cloud storage).
- user's data are stored across many physical locations and web sites. Users are not aware of where their data actually are and whether the security mechanisms of these sites meet their requirements.

Medical information management based on cloud computing faces the same problems. Moreover, due to the security and privacy regulations of HIPAA, cross-institutional medical data sharing becomes even more complicated and challenging.

1. Problems of cloud-based healthcare management.

Existing IT infrastructure deployments within a medical organization are usually based on private cloud architectures, which bring limitations on scalability and data sharing[20]. Private cloud [21] refers to a cloud computing model where IT services are provided over private infrastructure for the dedicated use of a single organization.

Because building highly scalable private clouds requires a large investment on computing and storage devices, and the rapidly changing volume of clinical data makes it difficult to accurately estimate required cloud capacity in the future[6], private cloud-based approaches are inconvenient for collaborators who reside outside of the domain perimeter to access data stored in the domain. These limitations prevent the further sharing of medical information demanded by big data analytics.

On the other hand, public clouds support scalability and data sharing well. However, the multi-tenancy characteristic of public cloud services decides virtual machines are shared among various applications that expose the data to different types of attacks. Worse still, it is difficult to detect or monitor such attacks in a shared VM environment.

Whether private or public clouds are adopted for healthcare management to guarantee data security and privacy, a basic requirement is data encryption. Unfortunately, a dilemma comes from the key management problem. Letting cloud users manage encryption keys certainly will enhance data security and provide better control; however, it would be a troublesome burden for users to distribute corresponding keys to authorized users, which limits its scalability of sharing data among a large number of institutions. This is the primary limitation of traditional key distribution center (KDC)-based solutions. Allowing cloud providers to control the keys could potentially increase the risks of data leakage because cloud administrators have the chance to "touch" the keys and further to decrypt data. This is the dilemma faced by HIPAA-compliant clouds[22] such as Amazon, Google, and Microsoft that provide externally-hosted clouds for medical information management.

2. Cryptography for medical data sharing.

To address the aforementioned problems, one possible solution is to enable owner-dominated security mechanisms for medical data outsourced to clouds. Such mechanisms[23–26] are designed to protect the security of remotely stored data in cloud computing, which demonstrate that providing owners with data access control is more important than letting the cloud take the full control over their data. Since users no longer physically possess their data, they want to at least be able to decide who can visit their data. This can return some control back to data owners, therefore promoting users' confidence in data security.

Another trend revealed from these cloud-based data sharing schemes is that traditional security means adopted in a single administrative domain are insufficient for medical data sharing across multiple healthcare domains. Hence, more advanced cryptographic primitives with rich access control semantics and strict confidentiality enforcement are required. Currently, there are some research projects that focus on adopting advanced cryptography to secure medical data sharing based on cloud storage platforms.

Li et al.[27] used attribute-based encryption (ABE) for secure sharing of personal health records stored in semi-trusted cloud servers. Their design divides security domains into public domains (physicians and medical researchers) and personal domains (family members and friends), where two types of ABE schemes (e.g., a revocable key-policy ABE scheme and a multi-authority ABE scheme) are adopted to address data sharing in public and private domains, respectively. Despite patients' full control of their medical information, the scheme poses too much burden on patients, since the

patient side applications are required to generate and distribute corresponding keys to authorized users.

J. Liu et al.[28] proposed using ciphertext-policy attrubute-based signcryption to provide fine-grained access control and secure sharing of Personal Health Records (PHRs). B. Fabian el al.[29] put forward a collaborative architecture for inter-organizational sharing of medical data in semi-trusted cloud computing environments, which adopts attribute-based encryption for selective authorization of data access and a secret sharing technique to securely distribute data across multiple clouds.

R. Guo el al.[30] combined blockchain technology with a multi-authority attribute-based signature scheme to secure the storage and access of electronic health records. An ABE signature reveals only that the verified message must be signed by a signer whose attributes satisfy certain predicates, which prevents identity leakage when a user signs a message. 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.

Narayan et al. [31] presented a patient-centric EHR system to let patients selectively share portions of their health data stored in cloud. They adopted a broadcast attribute-based encryption (bABE) to enforce access control to medical files. Meanwhile, they provided public-key encryption with keyword search (PKES) on encrypted data. However, their design lacked algorithmic details about adopted bABE and PKES schemes.

Barua et al. [32] proposed an efficient and secure patient-centric access control (ESPAC) scheme on the basis of the ciphertext-policy attribute-based encryption to allow patient-centric access control. Identity-based encryption was adopted to secure end-to-end communications where identity privacy, message integrity, and non-repudiation are ensured.

Chen et al. [33] gave a cloud-based privacy-aware role-based access control (CPRBAC) model for data controllability and traceability, and authorized access to healthcare cloud resources. They also designed an active auditing scheme to monitor and report illegal operations. However, their work does not contain any cryptographic primitive to ensure data confidentiality and integrity.

Discussion. It should be noted that data interoperability remains a significant issue in cloud environments due to the incompatibility of various cloud services. Let us consider medical data sharing on a statewide or national scale. It involves many cloud providers, and each provider has its own data security and privacy safeguards. To what extents will these mechanisms of various providers be compatible with each other? Unfortunately, the answer is unclear. We will discuss it further in following sections.

B. Anonymization-based Privacy Preservation in Healthcare

1. Data Anonymization Models

Privacy-preserving data publishing has gained much attention recently, especially when data mining and analytics is becoming a mainstream technological trend in the big data era. Researchers have designed various data anonymization algorithms such as generalization, suppression, and diversity slicing to protect individuals' privacy in transactional data publishing.

Generally, there are three different privacy-preserving models (i.e., k-anonymity, l-diversity, and t-closeness, in the order of increasing complexity). Among them, k-anonymity was proposed by Sweeney and Samarati in 2001. The philosophy behind it is to allow each combination of quasi identifiers (non-identifiable attributes that could jointly identify an individual, e.g., birth date and zip code) be indistinctly matched to at least k individuals, which means a specific person's information cannot be distinguished from other $k-1$ persons' information in a dataset.

A stronger privacy protection model is l-diversity, which requires that each sensitive attribute include at least l well-represented values in the published dataset besides keeping k-anonymity property. t-closeness is a further refinement of l-diversity model that preserves privacy by reducing the granularity of data representation, which treats values of an attribute distinctly by taking into account the distribution of values of the attribute. It is a trade off that leads to some loss of effectiveness of data mining in order to gain some privacy.

Based on these three models, various algorithms [34–36] focusing on improving these anonymization models have been proposed. Comprehensive surveys of this area can be found in [37, 38].

Differential privacy [39] is another technique to provide data anonymization by adding noise to a dataset so that an attacker cannot determine whether a particular data portion is included. Soria el al. [40, 41] proposed using microaggregation-based k-anonymity to reduce the noise to be added to generate differential private datasets.

2. Data Anonymization in Healthcare

HIDE [42] is an integrated health data de-identification system for both structured and unstructured data. Basically, it deploys a conditional random fields-based technique to extract identifiable attributes from unstructured data and a k-anonymity-based technique to de-identify data while maintaining maximum data utility.

Emam et al. [43] proposed an optimal lattice anonymization (OLA) algorithm based on k-anonymity. It produces a globally optimal de-identification solution suitable for health datasets. Their evaluation on six datasets shows that OLA results in less information loss and faster performance compared to existing de-identification algorithms.

Belsis et al. [44] presented a clustering-based anonymity scheme for sensor data collection and aggregation in wireless medical monitoring environments. Their design is based on k-anonymity since it protects user privacy by making an entity indistinguishable from other $k-1$ similar entities.

Loukides et al. [45] proposed an approach to allow data owners to share personal health data without disclosing identities, incurring excessive information loss, or harming data usefulness. The approach transforms data via disassociation, which is an operation that splits health records into carefully constructed subrecords to hide combinations of diagnosis codes.

TABLE II: Metrics of Surveyed Schemes

Research Work	Metrics								
	identity management	access control	data authenticity	data encryption	encrypted keyword search	blockchain type	smart cotract	data storage	data inter-operability
MedRec[47]	✓	coarse grained	✓	–	–	public	✓	off-chain	inter-domain
MedRec+[48]	✓	coarse grained	✓	✓	–	public	✓	off-chain	inter-domain
HDG[49]	✓	coarse grained	–	–	–	*	–	*	intra-domain
BBDS[50]	✓	coarse grained	✓	–	–	consortium	–	cloud	inter-domain
MeDShare[51]	✓	coarse grained	✓	–	–	consortium	✓	cloud	inter-domain
MedBlock[52]	✓	coarse grained	✓	–	–	*	–	off-chain	inter-domain
Peterson et al.[53]	✓	coarse grained	✓	–	–	*	–	*	inter-domain
PSN healthcare [54]	✓	coarse grained	–	–	–	public	–	–	*
Liang et al.[55]	✓	coarse grained	✓	×	×	consortium	×	off-chain	inter-domain
Patientory [56]	✓	coarse grained	✓	✓	✓	consortium	✓	off-chain	inter-domain
Guo et al.[30]	✓	coarse grained	✓	–	–	*	–	on-chain	inter-domain
BSPP [57]	✓	coarse grained	✓	✓	✓	hybrid	–	on-chain	inter-domain

– denote the metric item not implemented in the paper.
* denote the metric item not mentioned in the paper.
inter-domain means cross-institutional data sharing.
intra-domain means data sharing in a healthcare domain.
coarse-grained means no classification of medical records according to data sensitivities.

Discussion. Data anonymization is an ongoing research area. However, it needs to strike a balance between anonymity and data utility. Currently, none of k-anonymity, l-diversity, and t-closeness can completely ensure that no privacy leakage occurs while maintaining a reasonable level of data utility [37, 38]. Specifically, k-anonymity and l-diversity do not protect anonymity from every attack (e.g., homogeneity attack, background attack, similarity attack, and skewness attack etc.)[46]. In contrast, t-closeness offers complete privacy but severely impairs the correlations between key attributes and confidential attributes. Hence, it would be better to integrate data anonymization with other techniques to achieve a good trade-off between privacy preservation and data utility.

C. Blockchain in Healthcare

Recently, with the adoption of blockchain technology becoming a widespread trend in distributed computing, many researchers now consider using blockchain to secure medical data sharing and management. Table II surveys the state-of-the-art medical data sharing schemes[1] based on blockchain technology. We compare security metrics (identification, access control, data authenticity, data encryption) to architecture metrics (blockchain type, data storage method) and functionality metrics (smart contract, interoperability). Moreover, we classify these schemes into two types: permissioned blockchain-based approaches and permissionless blockchain-based approaches.

[1]We only include in the table schemes with a complete framework or system addressing secure medical data sharing. Schemes focusing on a single security functionality are not included in the table, they are discussed in the paper instead.

1. Approaches based on Permissionless Blockchain

Zyskind et al. [58] proposed using blockchain to provide secure and privacy-preserving data sharing among mobile users and service providers. Their design proposes two types of transactions, i.e., transaction T_{data} is used for data storage and retrieval, and transaction T_{access} is used for access control.

MedRec [47] is a decentralized EMR management system based on blockchain technology that provides a functional prototype implementation. MedRec has 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. Specifically, registrar contracts maps node identity strings to their Ethereum addresses. A patient-provider relationship (PPR) contract defines the stewardship and ownership of a patient's clinical data, where access permissions and query strings indicating data positions are also included. A summary contract holds a list of PPR references to denote its engagements with other patient nodes or hospital nodes. In implementation, four software components (e.g., backend library, Ethereum client, database gatekeeper, and EMR manager) are deployed on a system node to implement the business logic of medical data sharing and management.

Based on the work of MedRec, H. Yang et al. [48] proposed using signcryption and attribute-based authentication to enable the secure sharing of healthcare data. EHRs are encrypted with a symmetric key, which is further encrypted with an attribute key set. The concatenation of both ciphertexts(encrypted EHRs and encrypted key) is signed with a private key. For data accessing, a user verifies the signature and performs key decryption and EHRs decryption to get the plaintext EHRs.

Yue X. et al. [49] presented a healthcare data gateway, which is a blockchain-based architecture equipped with a purpose-centric access control policy to let patients own, control, and share their medical information without violating privacy. But their scheme lacks the details of how a service is prevented from knowing the data content when a computation runs on the raw medical data.

Zhang et al. [54] proposed a pervasive social network (PSN)-based healthcare environment, which consists of a wireless body area network and a PSN area. Their design incorporates an authenticated association protocol to initiate a secure link between medical sensors. Afterwards a coordinator node in the PSN area can broadcast a transaction and add it to new blocks. However, the authors did not provide details about their consensus protocol and smart contracts. Zhao et al. [59] proposed using fuzzy vault technology to design a lightweight backup and recovery scheme to manage keys, which are used to encrypt health signals collected from body sensor networks (BSN) and stored on a health blockchain. But their work lack details of how their health blockchain works.

Modelchain [60] was designed to adapt blockchain for privacy-preserving machine learning to accelerate medical research and facilitate quality improvements. In the design, a proof-of-information algorithm on top of PoW consensus protocol determines the order of online machine learning to increase efficiency and accuracy.

These schemes are proposed to adopt a permissionless or public blockchain to secure medical data sharing and various applications (e.g., healthcare sensors, machine learning). However, public blockchain is usually crypto-currency driven (bitcoins in Bitcoin or ether in Ethereum), which means a certain amount of cryptocurrency[2] has to be paid for transaction inclusion and block mining. According to Ethereum yellow paper[15], storing a kilobyte cost 640 thousand gas, which amounts to $2.3 even at a relatively low gas price of 20G wei (1 Ether = 10^9 Gwei) and with Ether recently valued at $168 (May 8, 2019).

Storing data on a public blockchain can be very expensive. It is financially impractical to store detailed clinical information of millions of patients on chain. Instead, only a very tiny subset of critical metadata can be stored on the blockchain. Data-related behavior in a public blockchain, like access request, access policy validation, and message transferring, can all be costly since they require transactions that describe them to be generated and included in blocks.

2. Approaches based on Permissioned Blockchain

K. Peterson el. al[53] proposed a blockchain-based approach for cross-institutional health information sharing. They designed new transaction and block structures to enable secure access of fast healthcare interoperability resources (FHIR) that were stored off-chain. Moreover, they designed a new consensus algorithm that avoids the expensive computational resources consumed by the PoW consensus in Bitcoin. In their design, a block would undergo a transaction distribution phase, a block verification request phase, a signed block return phase, and a new blockchain distribution phase before being added to the blockchain. A proof-of-interoperability concept was proposed in their consensus mechanism to ensure transaction data be in conformance with FHIR structural and semantic constraints. They also designed a random miner election algorithm where each node in the network has an equal probability to become a miner in the future. However, the paper does not mention how the medical data are organized, stored, and accessed. The privacy-preserving keyword searches adopted in their framework lack algorithmic details.

Q. Xia el al. proposed BBDS[50], a high-level blockchain-based framework that permits data users and owners to access medical records from a shared repository after successful verification of their identities and keys. An identity-based authentication and key agreement protocol in [61] is used to achieve user membership authentication. However, their secure sharing of sensitive medical information is limited to invited and verified users. The authors also proposed MedShare[51], a similar 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[52], a hybrid blockchain-based architecture to secure electronic medical records (EMR), where nodes are divided into endorsers, orderers and committers. Its consensus protocol is a variant of the PBFT[14] consensus protocol. However, the authors did not explicitly explain the access control policy to allow third-party researchers to access medical data. Moreover, their proposal of using asymmetric encryption algorithms to encrypt medical information does not seem to be a good option considering the encryption/decryption performance of asymmetric encryption. S. Wang et al.[62] presented a parallel healthcare system (PHS) where descriptive intelligence, predictive intelligence, and prescriptive intelligence in healthcare are achieved on the basis of artificial systems, computational experiments, and parallel executions. In their framework, a consortium blockchain containing patients, hospitals, health bureaus and communities, and medical researchers, is deployed. Smart contracts are implemented to enable medical records sharing, review, and auditing.

X. Liang et al.[55] proposed a user-centric framework on a permissioned blockchain for personal health data sharing, where the Hyperledger Fabric membership service and channel formation scheme are used to ensure privacy protection and identity management. They implement a mobile app to collect health data from wearable devices and synchronize data to the cloud for storage and sharing with healthcare providers. A. Zhang et al. [57] designed a hybrid blockchain-based secure and privacy-preserving (BSPP) PHI sharing scheme, where a private blockchain is used to store PHI for each hospital and a consortium blockchain is used to keep secure indices of the PHI. In their design, a public encryption-based keyword search scheme[63] is adopted to secure the search of PHI and to ensure identity privacy.

Patientory [56] is a healthcare peer-to-peer EMR storage network that leverages blockchain and smart contracts to provide HIPAA compiant health information exchange. The authors also proposed a software framework to address the

[2]Actually, a user has to use real money to buy cryptocurrenty, e.g., $5892 for one BTC(bitcoin currency) in Bitcoin (May 8, 2019). So here we regard cryptocurrency expenses as real monetary expenses.

authentication, authorization, access control, and data encryption in system implementation, as well as interoperability enhancement and token management.

The ChainAnchor [64] system provides anonymous identity verification for entities performing transactions in a permissioned blockchain. The system employs Enhanced Privacy ID (EPID) zero-knowledge proof scheme to prove participants' anonymity and membership.

The aforementioned schemes choose consortium or permissioned blockchain to secure the storage of medical information. This is different from approaches based on public blockchains, such as Bitcoin and Ethereum, which are totally decentralized. Instead, consortium blockchain requires certain permission to access the blockchain. This means that participants are selected in advance and only those authorized nodes can be allowed to access information stored on the blockchain. Such a setting is similar to the medical data sharing scenario, where only healthcare stakeholders (patients, healthcare providers, and authorized medical researchers) can be allowed to access that information based on their authorized permissions.

However, in spite of its high throughput, permissioned blockchain is far from a perfect solution for secure medical data sharing. The most notable disadvantage is the necessity of a central authority, which is usually comprised of a group of companies with a shared interest that will run the blockchain network and oversee the whole system. Therefore, the data immutability in public blockchain is discounted in consortium blockchain, which opens up the possibility of blockchain rollback by an attacker or a certain authority member.

D. Software-Defined Infrastructures for Healthcare

While cloud platforms provide flexible and cost-effective computing resources on demand, Software-Defined Infrastructures (SDIs) provisioned at the network edge support applications with significant performance requirements, especially in terms of throughput and latency. SDI technologies are fundamental to many home-based medical applications [65, 66] due to their programmability of networks via software-defined networking (SDN), and the feasibility of resource management in the cloud via OpenStack.

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 has gained wide attention in network-based data management systems. A typical example is home-based medical applications, where abundant programmable resources are installed at a given patients premises, such as desktops, embedded controllers, and smart routers, by which apps are allowed to interface with various home sensors that capture a patient's real-time activities. All of the heterogeneous resources at every part of the network, including the end point, the edge, and the core, enable the deployment of high-performance medical services.

P. Li et al. proposed CareNet [65, 66], 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. [67] proposed a smart health monitoring method on the basis of software-defined networking, where a centralized smart controller is designed to manage all physical devices and provide interfaces to data collection, transmission and processing.

IV. A Blockchain Future for Medical Data Sharing

A. Blockchain for medical data sharing

In the previous section, we surveyed the state-of-the-art approaches on secure medical data sharing with a focus on blockchain technology adoption. Regardless of whether the adopted blockchain is permissioned or permissionless, these schemes [30, 47–53, 55, 57] shed a light on the blockchain application in medical data sharing and management. However, blockchain itself is not a panacea to solve all security and privacy problems in medical data sharing. In truth, we should be more aware of the limitations of blockchain technology than of its advantages so that we can compensate for its disadvantages by integrating with other techniques (e.g., cryptographic primitives) to address the security problems of medical information management.

Secure sharing of medical data involves patients, healthcare providers, and third-party medical researchers. Due to the privacy and security regulations of HIPAA, medical data management should provide secure storage of raw medical information (confidentiality, integrity), privacy-preserving data provision (data authenticity, user authentication, access control), auditability, traceability, and data interoperability. Besides, when blockchain is adopted for healthcare data sharing, the following key features may need further investigation.

1. On-chain or off-chain storage of medical data?

Blockchain was originally designed to record small size trading transactions, so its data capacity is usually limited. For instance, the block size in Bitcoin is limited to one megabyte, which is insufficient to store medical data such as X-ray images. Furthermore, there remain other aspects with regard to the data cycle that need to be seriously considered.

- On-chain stored data cannot be altered or deleted because blockchain is a continually growing public ledger. However, some regulations such as GDPR in Europe have strengthened patient rights to erase their personal health information since a patient owns his or her medical records.
- 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[6].

Blockchain itself is a secured and transparent public ledger that can guarantee the integrity of on-chain stored data (transactions and blocks). That means blockchain can be leveraged to secure the storage of medical information if we choose on-chain data storage. However, this naive approach will lead to poor throughput and performance since on-chain transactions and blocks need to be downloaded locally by every peer node,

TABLE III: Characteristics of permissioned and public blockchain

Characteristic	Permissioned	Public
Consensus	PoS, PBFT	PoW
Currency	optional	required (incentivization)
Access	by invitation	anybody
Data immutability	medium	high
infrastructure	decentralized	distributed
Throughput	high	low
Privacy	private from non-participants	pseudonymity

which leads to a great bandwidth waste. This explains why most of the state-of-the-art approaches[47, 48, 50–52, 55] on medical data sharing chose to store medical information off-chain while data query strings and hash values are stored on-chain for authenticity and integrity verification. In such an architecture, medical data can be secured, modified and deleted as necessary.

2. Data encryption or not?

From the above analysis, it can be seen that on-chain storage of medical information is not a good choice due to the limited block size in current blockchains and the bandwidth waste to achieve network consensus. Off-chain storage of medical information seems to be a feasible alternative. However, in this case, we should be aware of one fact: blockchain can only guarantee the security of on-chain stored data. Hence, for those off-chain stored data, we still need to design data storage and access mechanisms with appropriate cryptographic primitives to fulfill its security and privacy goals.

Before going ahead, a basic question should be answered: should off-chain stored medical data be encrypted? According to a 2014 study[68], over 50% of security breaches occur in the medical industry, with up to 90% of healthcare organizations having exposed their data or had it stolen. It is obvious that storing plain-text medical records in a medical database undoubtedly will increase leakage risks, which is primarily due to following reasons:

1) once a healthcare system is compromised, then all medical information could be leaked;
2) despite the strict access control policy deployed in a healthcare system, an internal IT technical staff member still can easily "touch" the data, which makes data confidentiality difficult to guarantee.

In this context, we believe that encryption of medical data and secured key storage are two necessary steps to enhance the security and privacy of medical information. Data encryption can be the last line of defense when a healthcare system is compromised because an attacker can learn nothing about the encrypted data if one cannot obtain the corresponding encryption key.

3. Permissioned or permissionless blockchain?

As we have introduced in Section II, permissioned and permissionless blockchain primarily differ in their adopted consensus protocols, which in turn have a great impact on throughput, block mining time, access policies, and privacy. Table III shows the main difference between the two types.

Currently, the most concerned performance metric is throughput. For example, Hyperledger Fabric can process up to 10000 transactions per second (TPS), which is much faster than Ethereum's 20 TPS and Bitcoin's 7 TPS. The last two are insufficient to address the data access events that happen in a real-world healthcare management system. Fortunately, with the evolution of new consensus protols and technologies, the blockchain throughput undoubtedly will increase. For example, the Casper version of Ethereum (Ethereum 2.0) that adopts the Proof of Stake(PoS) consensus and a sharding technique can attain a 8-million TPS throughput[69].

Another challenge of adopting permissionless blockchain for medical data sharing would be cryptocurrency, which is the incentive that makes the behind consensus protocol work. In medical data management, data access happens very frequently. That means a great amount of money (cryptocurrency) is needed to run the network for healthcare data management. A possible option is to issue an altercoin in the system to pay contributors (miners). When a contributor has accumulated a certain amount of altercoin, whose level of trustworthiness will be promoted and, as a result, the contributor can get better service in the system.

B. Cryptography for medical data sharing

Considering that current blockchain cannot accommodate medical information due to its limited block size, storing medical data off-chain seems to be the only feasible solution. Securing the storage of these off-chain stored data becomes a challenge. This section briefly introduces some mainstream cryptographic primitives used for access control, key and privilege management.

1. Broadcast encryption

Broadcast encryption was first introduced in [70] and improved in [71, 72], which let an owner encrypt a small piece of data to a subset of users. Only users in the subset can decrypt the broadcast message to recover the data. In cryptographic cloud storage [24–26], instead of directly encrypting data content, keys are encrypted by broadcast encryption schemes to enforce access control where authorized users can recover the key by decrypting the broadcast message, whereas unauthorized or revoked users cannot find sufficient information to decrypt the message.

2. Identity-based encryption

The concept of identity-based encryption (IBE) was first proposed by Shamir[73] in 1984, who suggested that a public key can be an arbitrary string, and then improved by Boneh and Franklin[74] using Weil pairing over elliptic curves. In IBE, a trusted third party called the Private Key Generator (PKG), generates a master public-private key pair for each identity string. In practice, given a master public key, any party can compute a public key corresponding to the identity by combining the master public key with the identity string. To obtain a corresponding private key, the authorized party with identity ID needs to contact the PKG, which uses the master private key to generate the private key for identity ID.

IBE eliminates the need for a public key distribution infrastructure. It allows any pair of users to communicate

securely without exchanging private or public keys, which is ideal for data sharing among a closed group (e.g., within an organization).

3. Attribute-based encryption

In many applications, there is the need to share data according to a specific policy without prior knowledge of who will be the data receiver. Suppose a patient wants to share his medical records only with a user who has the attribute of "PHYSICIAN" issued by a medical organization and the attribute "RESEARCHER" issued by a clinical research institute. With attribute-based encryption[75], the patient can define an access policy ("PHYSICIAN" AND "RESEARCHER") and encrypt his medical records with this policy, so that only users with attributes matching this policy can decrypt the records.

Attribute-based encryption is a promising cryptographic technique for access control of encrypted data. Generally, it can be divided into two categories: (a) key-policy attribute-based encryption (KP-ABE)[76] where keys are associated with access policies and ciphertext is associated with an attributes set; and (b) ciphertext-policy attribute-based encryption (CP-ABE) [76] where keys are associated with an attributes set and ciphertext is associated with access policies. In both schemes, a central authority is required to issue and validate private keys, rendering them unsuitable for a distributed environment where data sharing takes place across different administrative domains.

To address the single authority problem in existing ABE, Multi-Authority Attribute-Based Encryption (MA-ABE)[77–79] schemes are proposed, where no central authority is needed and collusion resistance is guaranteed.

4. Proxy re-encryption

Proxy re-encryption (PRE), proposed by Blaze[80] et al. in 1998 and improved by G. Ateniese[81, 82] et al. in 2006, is a cryptosystem that allows a third party (proxy) to alter a ciphertext encrypted by one party so that it can be decrypted by another authorized party. The basic idea behind it is that two parties publish a proxy key that allows a semi-trusted intermediate proxy to convert ciphertext, which avoids data decryption and re-encryption at the sender side. Thus, it is suitable for data sharing across multiple domains where data owners can leave the task of data re-encryption to a proxy (e.g., cloud) after user revocations.

5. Search on encrypted data

Searchable symmetric encryption (SSE)[83] can enforce keyword search on outsourced encrypted data, which avoids the decryption process and thereby enhances query efficiency without the risk of data leakage. Otherwise, data owners either have to send service providers the keys for data decryption before executing a query, or download encrypted data locally and decrypt it to perform a query. Both approaches are unacceptable due to security or efficiency reasons. The idea behind SSE is to deploy a masked index table as metadata[84, 85] that facilitates searches on encrypted data. The data owner needs to create an index table based on pre-processed message-keyword pairs. To perform a search, a search token is provided by the user with which the server searches through the index. If a match is found, then the matching encrypted data is returned to the user.

Discussion. As we have pointed out, that relying on blockchain technology to secure off-chain stored medical data is infeasible. Hence, a secure healthcare system still needs to employ appropriate cryptographic primitives to achieve confidentiality, integrity, access control, and privacy protection. Specifically, for encrypted data, advanced cryptographic primitives (e.g., IBE, ABE, PRE) is becoming widely deployed to enforce strict and flexible access control of encryption keys. Hence, in the near future, it can be expected that cryptography will play a more important role in blockchain-based data sharing.

C. Future Research Work

According to the analysis in SectionIV-A, it is clear that medical data should be stored off-chain in encrypted form due to network throughput and security reasons. Yet the third question—whether to adopt permissioned or permissionless blockchain—remains open with no apparent solution. Despite the debate over permissioned and permissionless blockchains throughout academia and industry, there is no strong evidence showing that one type can completely substitute the other type. One possible method is that researchers can leverage the advantages of both types by constructing a hybrid blockchain architecture as in [57]. However, this may cause great complexities in the management of consensus executions, including block mining, data access control, and data provision.

Therefore, future research on designing blockchain-based approaches for secure medical data sharing can focus on following areas.

1) **Cryptography-based access and privacy control.** To ensure the security and privacy required by HIPAA regulations, cryptography needs to be embedded in the design to enforce strict access control and privacy preservation. The state-of-the-art schemes [30, 47–52, 55, 57] in medical data areas rely more or less on the adoption of certain cryptographic primitives to implement authentication, access control, key management, and privacy protection for medical information.
2) **Smart contract-driven business logic.** Smart contracts, as a series of self-executing contractual states without third parties, are the core element to implementing the business logic of blockchain-based medical data sharing. By designing smart contracts specific to certain requirements, the creation of medical records, authorization and revocation of access permissions, and auditing and provenance of access behavior can be implemented on the blockchain.

Figure 2 depicts a general architecture for blockchain-based healthcare data management, where three layers (i.e., health domain layer, blockchain layer, and user layer) are included. A healthcare system residing in an enclosed network domain is regarded as a health domain, which usually has one or more databases to store patients' medical records and clinical trials. The blockchain layer is used to connect scattered health domains, where smart contracts are responsible for the implementation and execution of the business logic of cross-institutional data sharing. The user layer consists of patients,

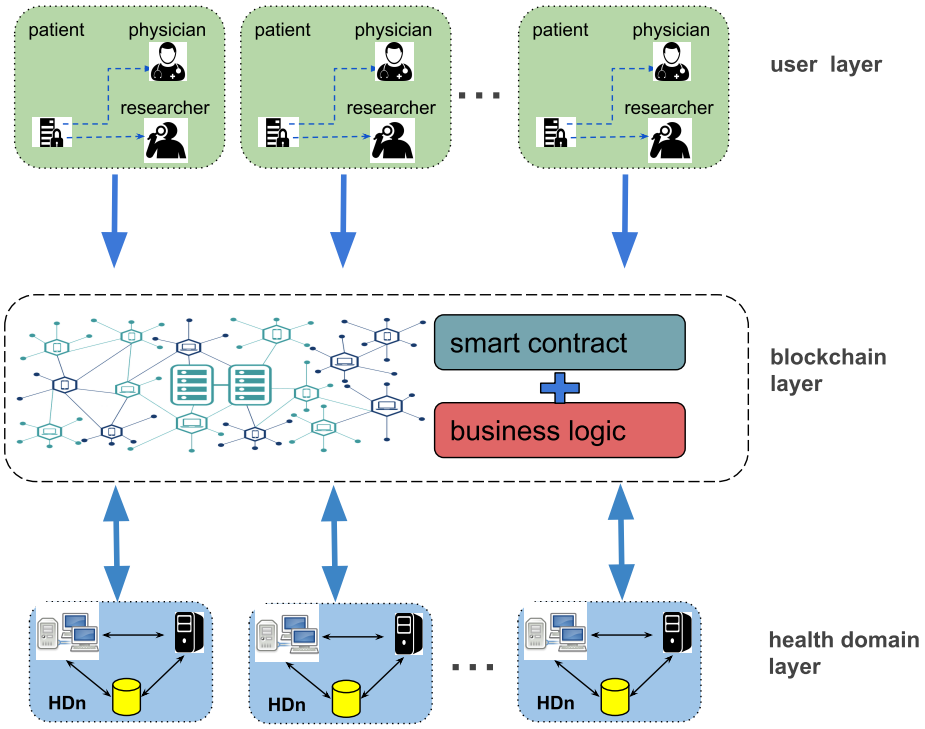

Fig. 2: Architecture of blockchain-based medical data sharing

doctors and medical researches from different healthcare organizations.

Blockchain-based medical information sharing is an ongoing field that requires a vast amount of techniques to cooperate with one another in order to achieve HIPAA compliant data sharing. In the future, new architectures and security and privacy-related cryptographic primitives may appear and can be seamlessly integrated with blockchain. Here, we briefly discuss some challenges in blockchain-based medical data sharing that need further investigation and exploration.

1. Query on scattered medical databases.

Traditionally, EMR data are organized in SQL-based relational databases for storage, and queries are performed within an independent administrative domain where the database resides. However, in a blockchain setting where various healthcare institutes are connected through the blockchain, it is inconvenient to make such a SQL query due to data stewardship and network boundaries.

Most existing schemes choose to store encrypted metadata (e.g., query strings in [47, 48] and secure indices in [53, 57]) indicating data locations on chain. When a client wants to perform a global query on all connected databases, a further challenge would be in how to efficiently perform the query on all independently managed databases simultaneously and get an aggregated query result. This problem remains unaddressed in existing schemes. A possible solution is to let some partially centralized servers distributed in the network to collect and aggregate parallelly computed queries and return the aggregated esult to the querier. However, strong security and recover mechanisms may need to be carefully deployed on these servers to protect them from denial-of-service attacks.

2. Finer-grained access and privacy control.

Currently, there are some methods[30, 48] that have adopted advanced cryptographic primitives (e.g., attribute-based encryption) to enforce strict and flexible access control for medical data access. Specifically, they focus on the construction of access policies with rich semantics, which, of course, is necessary in access policy customization. However, the differentiation of various EMR fields in sensitivity is also of critical importance for privacy control. A naive approach is to segment a record into multiple parts according to sensitivities and encrypt each part with a different key, however, which complicates the task of key management when the separation is fine-grained. To address this problem, some key derivation mechanisms[24, 26] can be integrated with access control policies to facilitate key management.

3. Compatibility of security mechanisms among healthcare domains.

Since each healthcare institute can be regarded as an independent domain equipped with its own security and privacy mechanism, it is difficult to predict the extent to which these mechanisms will be compatible with each other. Furthermore, one should also consider how to address the compatibility problem caused by different or even contradictory data privacy laws of various states or nations.

4. Software-defined networking is needed to facilitate domain management. The SDN controller provides a central point of control to distribute policy information. However, centralized control by one entity has the disadvantage of creating a central point of attack. Moreover, the programmability associated with the SDN platform adds security risks. Therefore, properly and securely implementing a SDN controller to cooperate with blockchain and facilitate the management and collaboration among various healthcare domains is of great

importance. This should simplify the management of existing legacy healthcare systems to let them be easily added to the new blockchain-based architectures.

V. Conclusion

Medical information sharing without violating security and privacy regulations has long been a challenging topic. This paper reviews related solutions in this area, including cloud-based approaches, blockchain-based approaches, and SDN-based approaches. We observed that security and privacy protection of medical information covers confidentiality, integrity, and authenticity of data in transit and at rest, access and privacy control, etc. Therefore, a practical approach for medical data sharing may need to integrate many different 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 (permissioned or permissionless) for medical data sharing. Moreover, there are still some problems calling for further investigation and exploration in blockchain-based medical data management. We shed a light on these challenges by pointing out potential research directions and methodologies that may further secure and facilitate the sharing of healthcare information.

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, 1738965, 1450996, and 1541434).

References

[1] "Data leakage events," https://informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/.

[2] https://www.hipaajournal.com/healthcare-data-breach-statistics/, 2018.

[3] "Summary of the hipaa security rule," https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/, 2017.

[4] "General data protection regulation," https://eugdpr.org/the-regulation/, 2016.

[5] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health information science and systems*, vol. 2, no. 1, p. 3, 2014.

[6] C. Esposito, A. De Santis, G. Tortora, H. Chang, and K.-K. R. Choo, "Blockchain: A panacea for healthcare cloud-based data security and privacy?" *IEEE Cloud Computing*, vol. 5, no. 1, pp. 31–37, 2018.

[7] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," http://bitcoin.org/bitcoin.pdf, 2009.

[8] A. Gervais, G. O. Karame, K. Wüst, V. Glykantzis, H. Ritzdorf, and S. Capkun, "On the security and performance of proof of work blockchains," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. ACM, 2016, pp. 3–16.

[9] Ethereum, "Proof of stake faq," https://github.com/ethereum/wiki/wiki/Proof-of-Stake-FAQ, 2014.

[10] C. Cachin, "Architecture of the hyperledger blockchain fabric," in *Workshop on Distributed Cryptocurrencies and Consensus Ledgers*, vol. 310, 2016.

[11] J. Sousa, E. Alchieri, and A. Bessani, "State machine replication for the masses with bft-smart," 2013.

[12] J. Sousa, A. Bessani, and M. Vukolic, "A byzantine fault-tolerant ordering service for the hyperledger fabric blockchain platform," in *48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 2018, pp. 51–58.

[13] M. Castro and B. Liskov, "Practical byzantine fault tolerance," in *OSDI*, vol. 99, 1999, pp. 173–186.

[14] ——, "Practical byzantine fault tolerance and proactive recovery," *ACM Transactions on Computer Systems (TOCS)*, vol. 20, no. 4, pp. 398–461, 2002.

[15] G. Wood, "Ethereum: a secure decentralised generalised transaction ledger. ethereum project yellow paper 151 (2014)," 2014.

[16] Ethereum, "Solidity programming documentation," https://solidity.readthedocs.io/, 2016.

[17] R. Raja, "Chaincode on the go: Smart contracts on the hyperledger fabric blockchain," http://medium.com/coinmonks/chaincode-on-the-go-smart-contracts-on-the-hyperledger-fabric-blockchain-82dd61b3c669, 2017.

[18] K. M. Khan and Q. Malluhi, "Establishing trust in cloud computing," *IT professional*, no. 5, pp. 20–27, 2010.

[19] K. Ren, C. Wang, and Q. Wang, "Security challenges for the public cloud," *IEEE Internet Computing*, vol. 16, no. 1, pp. 69–73, 2012.

[20] S. Nepal, R. Ranjan, and K.-K. R. Choo, "Trustworthy processing of healthcare big data in hybrid clouds," *IEEE Cloud Computing*, vol. 2, no. 2, pp. 78–84, 2015.

[21] T. Dillon, C. Wu, and E. Chang, "Cloud computing: issues and challenges," in *24th IEEE International Conference on Advanced Information Networking and Applications (AINA)*. Ieee, 2010, pp. 27–33.

[22] "Architecting for hipaa security and compliance on amazon web services," https://d1.awsstatic.com/whitepapers/compliance/AWS-HIPAA-Compliance-Whitepaper.pdf, 2018.

[23] E.-J. Goh, H. Shacham, N. Modadugu, and D. Boneh, "Sirius: Securing remote untrusted storage." in *NDSS*, vol. 3, 2003, pp. 131–145.

[24] R. A. Popa, J. R. Lorch, D. Molnar, H. J. Wang, and L. Zhuang, "Enabling security in cloud storage slas with cloudproof." in *USENIX Annual Technical Conference*, vol. 242, 2011, pp. 355–368.

[25] A. Kumbhare, Y. Simmhan, and V. Prasanna, "Cryptonite: a secure and performant data repository on public clouds," in *5th IEEE International Conference onCloud Computing (CLOUD)*. IEEE, 2012, pp. 510–517.

[26] H. Jin, K. Zhou, H. Jiang, D. Lei, R. Wei, and C. Li, "Full integrity and freshness for cloud data," *Future Generation Computer Systems*, vol. 80, pp. 640–652, 2018.

[27] M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, "Scalable

and secure sharing of personal health records in cloud computing using attribute-based encryption," *IEEE transactions on parallel and distributed systems*, vol. 24, no. 1, pp. 131–143, 2013.

[28] L. Jianghua, H. Xinyi, and J. K. Liu, "Secure sharing of personal health records in cloud computing: Ciphertext-policy attribute-based signcryption," *Future Generation Computer Systems*, vol. 52, pp. 67 – 76, 2015, special Section: Cloud Computing: Security, Privacy and Practice.

[29] B. Fabian, T. Ermakova, and P. Junghanns, "Collaborative and secure sharing of healthcare data in multi-clouds," *Information Systems*, vol. 48, pp. 132–150, 2015.

[30] R. Guo, H. Shi, Q. Zhao, and D. Zheng, "Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems," *IEEE Access*, vol. 776, no. 99, pp. 1–12, 2018.

[31] S. Narayan, M. Gagné, and R. Safavi-Naini, "Privacy preserving ehr system using attribute-based infrastructure," in *Proceedings of the 2010 ACM workshop on Cloud computing security workshop*. ACM, 2010, pp. 47–52.

[32] M. Barua, X. Liang, R. Lu, and X. Shen, "Espac: Enabling security and patient-centric access control for ehealth in cloud computing," *International Journal of Security and Networks*, vol. 6, no. 2-3, pp. 67–76, 2011.

[33] L. Chen and D. B. Hoang, "Novel data protection model in healthcare cloud," in *2011 IEEE International Conference on High Performance Computing and Communications*. IEEE, 2011, pp. 550–555.

[34] M. Terrovitis, N. Mamoulis, and P. Kalnis, "Privacy-preserving anonymization of set-valued data," *Proceedings of the VLDB Endowment*, vol. 1, no. 1, pp. 115–125, 2008.

[35] Y. Xu, K. Wang, A. W.-C. Fu, and P. S. Yu, "Anonymizing transaction databases for publication," in *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2008, pp. 767–775.

[36] T. Li, N. Li, J. Zhang, and I. Molloy, "Slicing: A new approach for privacy preserving data publishing," *IEEE transactions on knowledge and data engineering*, vol. 24, no. 3, pp. 561–574, 2012.

[37] B. Zhou, J. Pei, and W. Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data," *ACM Sigkdd Explorations Newsletter*, vol. 10, no. 2, pp. 12–22, 2008.

[38] I. J. Vergara-Laurens, L. G. Jaimes, and M. A. Labrador, "Privacy-preserving mechanisms for crowdsensing: Survey and research challenges," *IEEE Internet of Things Journal*, vol. 4, no. 4, pp. 855–869, 2017.

[39] C. Dwork, "Differential privacy," *Encyclopedia of Cryptography and Security*, pp. 338–340, 2011.

[40] J. Soria-Comas, J. Domingo-Ferrer, D. Sánchez, and S. Martínez, "Enhancing data utility in differential privacy via microaggregation-based k-anonymity," *The VLDB JournalThe International Journal on Very Large Data Bases*, vol. 23, no. 5, pp. 771–794, 2014.

[41] D. Sánchez, J. Domingo-Ferrer, S. Martínez, and J. Soria-Comas, "Utility-preserving differentially private data releases via individual ranking microaggregation," *Information Fusion*, vol. 30, pp. 1–14, 2016.

[42] J. Gardner and L. Xiong, "Hide: an integrated system for health information de-identification," in *2008 21st IEEE International Symposium on Computer-Based Medical Systems*. IEEE, 2008, pp. 254–259.

[43] K. El Emam, F. K. Dankar, R. Issa, E. Jonker, D. Amyot, E. Cogo, J.-P. Corriveau, M. Walker, S. Chowdhury, R. Vaillancourt *et al.*, "A globally optimal k-anonymity method for the de-identification of health data," *Journal of the American Medical Informatics Association*, vol. 16, no. 5, pp. 670–682, 2009.

[44] P. Belsis and G. Pantziou, "A k-anonymity privacy-preserving approach in wireless medical monitoring environments," *Personal and ubiquitous computing*, vol. 18, no. 1, pp. 61–74, 2014.

[45] G. Loukides, J. Liagouris, A. Gkoulalas-Divanis, and M. Terrovitis, "Disassociation for electronic health record privacy," *Journal of biomedical informatics*, vol. 50, pp. 46–61, 2014.

[46] M. Wang, Z. Jiang, Y. Zhang, and H. Yang, "T-closeness slicing: A new privacy-preserving approach for transactional data publishing," *INFORMS Journal on Computing*, vol. 30, no. 3, pp. 438–453, 2018.

[47] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "Medrec: Using blockchain for medical data access and permission management," in *International Conference on Open and Big Data (OBD)*. IEEE, 2016, pp. 25–30.

[48] H. Yang and B. Yang, "A blockchain-based approach to the secure sharing of healthcare data," in *Proceedings of the Norwegian Information Security Conference*, 2017.

[49] 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, pp. 218–225, 2016.

[50] 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, pp. 44–59, 2017.

[51] Q. Xia, E. B. Sifah, K. O. Asamoah, J. Gao, X. Du, and M. Guizani, "Medshare: Trust-less medical data sharing among cloud service providers via blockchain," *IEEE Access*, vol. 5, pp. 14 757–14 767, 2017.

[52] 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, pp. 136–146, 2018.

[53] K. Peterson, R. Deeduvanu, P. Kanjamala, and K. Boles, "A blockchain-based approach to health information exchange networks," in *Proceedings of the NIST Workshop Blockchain Healthcare*, vol. 1, 2016, pp. 1–10.

[54] J. Zhang, N. Xue, and X. Huang, "A secure system for pervasive social network-based healthcare," *IEEE Access*, vol. 4, pp. 9239–9250, 2016.

[55] X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li, "Integrating blockchain for data sharing and collaboration in

mobile healthcare applications," in *28th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE, 2017, pp. 1–5.

[56] C. McFarlane, M. Beer, J. Brown, and N. Prendergast, "Patientory: A healthcare peer-to-peer emr storage network v1." *Entrust Inc.: Addison, TX, USA*, 2017.

[57] A. Zhang and X. Lin, "Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain," *Journal of medical systems*, vol. 42, no. 8, p. 140, 2018.

[58] G. Zyskind, O. Nathan *et al.*, "Decentralizing privacy: Using blockchain to protect personal data," in *IEEE Security and Privacy Workshops (SPW)*. IEEE, 2015, pp. 180–184.

[59] H. Zhao, Y. Zhang, Y. Peng, and R. Xu, "Lightweight backup and efficient recovery scheme for health blockchain keys," in *13th IEEE International Symposium on Autonomous Decentralized System (ISADS)*. IEEE, 2017, pp. 229–234.

[60] T.-T. Kuo and L. Ohno-Machado, "Modelchain: Decentralized privacy-preserving healthcare predictive modeling framework on private blockchain networks," *arXiv preprint arXiv:1802.01746*, 2018.

[61] L. Wu, Y. Zhang, Y. Xie, A. Alelaiw, and J. Shen, "An efficient and secure identity-based authentication and key agreement protocol with user anonymity for mobile devices," *Wireless Personal Communications*, vol. 94, no. 4, pp. 3371–3387, 2017.

[62] S. Wang, J. Wang, X. Wang, T. Qiu, Y. Yuan, L. Ouyang, Y. Guo, and F.-Y. Wang, "Blockchain-powered parallel healthcare systems based on the acp approach," *IEEE Transactions on Computational Social Systems*, no. 99, pp. 1–9, 2018.

[63] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *International conference on the theory and applications of cryptographic techniques*. Springer, 2004, pp. 506–522.

[64] T. Hardjono and A. Pentland, "Verifiable anonymous identities and access control in permissioned blockchains," *arXiv preprint arXiv:1903.04584*, 2019.

[65] P. Li, C. Xu, Y. Luo, Y. Cao, J. Mathew, and Y. Ma, "Carenet: building regulation-compliant home-based healthcare services with software-defined infrastructure," in *IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 2017, pp. 373–382.

[66] ——, "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, 2017, pp. 69–72.

[67] L. Hu, M. Qiu, J. Song, M. S. Hossain, and A. Ghoneim, "Software defined healthcare networks," *IEEE Wireless Communications*, vol. 22, no. 6, pp. 67–75, 2015.

[68] https://www.healthcareitnews.com/projects/biggest-healthcare-data-breaches-2018-so-far, 2018.

[69] "Ethereum 2.0 phase 0 – the beacon chain," https://github.com/ethereum/eth2.0-specs/blob/master/specs/core/0_beacon-chain.md#introduction.

[70] A. Fiat and M. Naor, "Broadcast encryption," in *Annual International Cryptology Conference*. Springer, 1993, pp. 480–491.

[71] J. A. Garay, J. Staddon, and A. Wool, "Long-lived broadcast encryption," in *Annual International Cryptology Conference*. Springer, 2000, pp. 333–352.

[72] D. Boneh, C. Gentry, and B. Waters, "Collusion resistant broadcast encryption with short ciphertexts and private keys," in *Annual International Cryptology Conference*. Springer, 2005, pp. 258–275.

[73] A. Shamir, "Identity-based cryptosystems and signature schemes," in *Workshop on the theory and application of cryptographic techniques*. Springer, 1984, pp. 47–53.

[74] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in *Annual international cryptology conference*. Springer, 2001, pp. 213–229.

[75] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2005, pp. 457–473.

[76] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in *Proceedings of the 13th ACM conference on Computer and communications security*. Acm, 2006, pp. 89–98.

[77] M. Chase, "Multi-authority attribute based encryption," in *Theory of Cryptography Conference*. Springer, 2007, pp. 515–534.

[78] M. Chase and S. S. Chow, "Improving privacy and security in multi-authority attribute-based encryption," in *Proceedings of the 16th ACM conference on Computer and communications security*. ACM, 2009, pp. 121–130.

[79] A. Lewko and B. Waters, "Decentralizing attribute-based encryption," in *Annual international conference on the theory and applications of cryptographic techniques*. Springer, 2011, pp. 568–588.

[80] M. Blaze, G. Bleumer, and M. Strauss, "Divertible protocols and atomic proxy cryptography," in *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 1998, pp. 127–144.

[81] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage," *ACM Transactions on Information and System Security (TISSEC)*, vol. 9, no. 1, pp. 1–30, 2006.

[82] M. Green and G. Ateniese, "Identity-based proxy re-encryption," in *Applied Cryptography and Network Security*. Springer, 2007, pp. 288–306.

[83] G. S. Poh, J.-J. Chin, W.-C. Yau, K.-K. R. Choo, and M. S. Mohamad, "Searchable symmetric encryption: designs and challenges," *ACM Computing Surveys (CSUR)*, vol. 50, no. 3, p. 40, 2017.

[84] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proceedings of IEEE Symposium on Security and Privacy*. IEEE, 2000, pp. 44–55.

[85] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in *Proceedings of IEEE Infocom*, 2010, pp. 1–5.

Dr. Hao Jin is a Research Associate in the Department of Electrical and Computer Engineering of the University of Massachusetts Lowell. He obtained his Ph.D. degree in Computer Science and Technology from Huazhong University of Science and Technology (China) in 2016. His research interests include cloud data security, digital forensics, auditing and accountability, and blockchain technology.

Dr. Yan Luo is a Professor of the Department of Electrical and Computer Engineering at the University of Massachusetts Lowell. He obtained his Ph.D. in Computer Science from University of California Riverside in 2005. While his research interest spans broadly computer architecture and network systems, Prof. Luos current research focuses on heterogeneous architecture and systems, software defined networking and deep learning. He and his team aim to design novel architecture and systems to facilitate programmable networking, deeply embedded sensing, and healthcare applications.

Dr. Peilong Li is a Research Assistant Professor of the Department of Electrical and Computer Engineering at the University of Massachusetts Lowell. He obtained his Ph.D. degree in Computer engineering from the UMass Lowell in 2016. His research interests include heterogeneous and parallel computer architecture, big data analytics with distributed computing framework, and data plane innovation in software defined networking.

Dr. Jomol Mathew is the Chief Research Informatics Officer at the University of Massachusetts Medical School. She received her Ph.D. in Plant and Soil Sciences from the University of Massachusetts at Amherst. Dr. Mathew oversees Data Science and Technology which includes clinical and research data integration, analytics, data visualization, and high-performance computing at the University of Massachusetts Medical School (UMMS). Dr. Mathew also directs the Clinical Informatics component of the Clinical and Translational Science Center (UMCCT) at UMMS. In 2017, she co-founded D3Health, a Center for Improving Patient and Population Health through Integration of Advanced Digital Technologies, Analytics, and Decision Support.