Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 6685762, 12 pages
https://doi.org/10.1155/2021/6685762

Research Article

A Blockchain-Based Medical Data Sharing Mechanism with Attribute-Based Access Control and Privacy Protection

Yingwen Chen,[1] Linghang Meng,[1] Huan Zhou,[1] and Guangtao Xue[2]

[1]*College of Computer, National University of Defense Technology, Changsha 410073, China*
[2]*School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*

Correspondence should be addressed to Huan Zhou; huanzhou@nudt.edu.cn

Received 11 December 2020; Accepted 10 June 2021; Published 1 July 2021

Academic Editor: Yaguang Lin

The rapid development of wearable sensors and the 5G network empowers traditional medical treatment with the ability to collect patients' information remotely for monitoring and diagnosing purposes. Meanwhile, the health-related mobile apps and devices also generate a large amount of medical data, which is critical for promoting disease research and diagnosis. However, medical data is too sensitive to share, which is also a common issue for IoT (Internet of Things) data. The traditional centralized cloud-based medical data sharing schemes have to rely on a single trusted third party. Therefore, the schemes suffer from single-point failure and lack of privacy protection and access control for the data. Blockchain is an emerging technique to provide an approach for managing data in a decentralized manner. Especially, the blockchain-based smart contract technique enables the programmability for participants to access the data. All the interactions are authenticated and recorded by the other participants of the blockchain network, which is tamper resistant. In this paper, we leverage the K-anonymity and searchable encryption techniques and propose a blockchain-based privacy-preserving scheme for medical data sharing among medical institutions and data users. To be specific, the consortium blockchain, Hyperledger Fabric, is adopted to allow data users to search for encrypted medical data records. The smart contract, i.e., the chaincode, implements the attribute-based access control mechanisms to guarantee that the data can only be accessed by the user with proper attributes. The K-anonymity and searchable encryption ensure that the medical data is shared without privacy leaking, i.e., figuring out an individual patient from queries. We implement a prototype system using the chaincode of Hyperledger Fabric. From the functional perspective, security analysis shows that the proposed scheme satisfies security goals and precedes others. From the performance perspective, we conduct experiments by simulating different numbers of medical institutions. The experimental results demonstrate that the scalability and performance of our scheme are practical.

1. Introduction

Data sharing is crucial for promoting the research of disease tracking and treatment. For instance, the sharing of substantial medical data can help government agencies make correct decisions in public health or help medical research institutions conduct scientific research to promote the progress of medical science. The fight against to current epidemic of COVID-19 also proves that efficient medical information sharing among different institutions can effectively trace the disease and accelerate the vaccine development. The application of wearable medical devices and health-related mobile apps alleviate the difficulty of personal medical data collection and retrieving, but for the following stages of data management, the security and privacy of the data have become the top concern at the same time.

Traditional approaches are based on clouds to store data remotely and different institutions share the data in a centralized manner. The top two challenges for these methods are privacy and security.

For the privacy issue, any data user should not infer a specific patient private information, including the name, address, and phone number, from querying the stored medical records. If the privacy of the patient is not well protected during data sharing, the patient would be reluctant to share their medical data [1]. In addition, for sharing purposes,

the data is stored remotely on clouds. The patients and medical institutions have to trust that the cloud provider would not leak out the data, if the data is not encrypted. However, the encrypted data would hinder the sharing process.

K-anonymity [2] and differential privacy [3] are two commonly adopted techniques to protect data privacy. Comparing both, differential privacy is relatively new and got more attention in recent years due to its strong privacy guarantee. Differential privacy is usually achieved through adding noise to attributes or values [4], i.e., through simply adding or deleting some less important data in the original dataset. Though differential privacy defines an extremely strict attack model for guaranteeing privacy, the noise data it introduces may affect the statistical characteristics of the data. However, for the medical research purpose, the statistical characteristics are more important to study a disease instead of focusing on some individual patients. Therefore, the K-anonymity [2] technique is more suitable for preprocessing the medical data, which obscures some sensitive data fields without affecting the original data. The accuracy of the statistical results on the data, therefore, can be preserved. On the other hand, the medical data are encrypted and stored on untrusted clouds. Searchable encryption is a potential solution which allows the server to search on encrypted data without knowing the content of the data [5].

For the security issue, cloud-based data platforms have problems of a single point of failure, vulnerability, and inefficiency. Besides, cloud-based data sharing relies on third-party services and there may exist stealing, leakage, tampering, or misusing of data. Although existing cryptography-related solutions have solved some problems of the cloud, the single point of failure problem cannot be solved. On the other hand, the access control for medical data management is also centralized and usually based on roles. However, the RBAC (role-based access control) model requires configuring complex rules to restrict the accessibility of different types of data users. Especially, the process of configuring and updating rules is vulnerable that the attack may leverage the loophole and easily elevate privileges to obtain the grant for the entire dataset, due to the centralized data storage.

Blockchain is a distributed ledger, which has the characteristics of decentralization, tamper resistance, and reliability. Thus, blockchain is a potential solution to replace centralized cloud storage. The blockchain-based smart contract provides a decentralized manner to authenticate the data access request among participants of different institutions. The ABAC (attribute-based access control) model can be further leveraged to simplify configurations for restricting the data accessibility according to the users' assigned attributes.

In this paper, we propose a privacy-preserving scheme based on the blockchain for medical data sharing. To tackle the two challenges mentioned above, the contributions of our developed system are as follows:

(i) We adopt the K-anonymity technique to preprocess the data for privacy preserving

(ii) We design the scheme based on searchable encryption for storing the encrypted medical data on clouds and enable the keyword search in a privacy-preserving manner

(iii) We develop smart contracts based on Hyperledger Fabric, and realize the secure keyword search and the attribute-based access control model

(iv) We implement a prototype system with smart contracts based on a chaincode of Hyperledger Fabric (the URL of source code: https://github.com/mythsand/privacy-preserving-medical-data) and conduct experiments with simulating different numbers of institutions

(v) We analyze the security properties and evaluate the computational overhead

The rest of the paper is organized as follows. Section 2 explains the preliminaries related to this paper, including K-anonymity, bilinear pairings, blockchain, and access control. Then, we formulate the problem with explaining the security goal and notions. The system design and technique details are introduced in Section 4. Security analysis and experimental studies are demonstrated in Section 5 and Section 6, respectively. Section 7 presents the related work and Section 8 finally concludes the paper.

2. Preliminaries

2.1. K-Anonymity. K-anonymity was the first model proposed to protect data privacy through syntax [2]. Its main idea is to make reidentification infeasible by hiding K objects in the same group. That is to say, K-anonymity requires that each record in the anonymized data cannot be distinguished from other at least K-1 records. Quasi-identifier attributes, e.g., social security numbers, state liquor identification cards, drivers' licences, and even passports or national identity cards, are concerned. Therefore, no identity in the K-anonymous dataset will be linked to fewer than K records. That is, the probability of correct reidentification is at most $1/K$. Several definitions of K-anonymity related to this paper are explained below. Here, we assume that all the information is in a table S, which contains multiple tuples.

Definition 1. (quasi-identifier attribute set). A quasi-identifier is a minimal set of attributes in table S that can be combined with external information records to reidentify personal information. This paper assumes that quasi-identifiers are known based on empirical data and epistemology.

Definition 2. (equivalent class). The equivalent class in table S indicates that each tuple is the same as several other tuples.

Definition 3. (K-anonymity property). Table S is a set of determined values of the attribute group in K that is anonymized to appear at least K times in S, i.e., each of the equivalent classes is at least K in size.

2.2. Bilinear Pairings. The scheme of our encrypted search is constructed based on bilinear maps, the definition of which is described as follows:

2.2.1. Bilinear Map. For two cyclic groups G_1 and G_2 of order p, there is a bilinear map e between them: $e : G_1 * G_1 \longrightarrow G_2$. The map relation satisfies following three properties:

(1) Computability: given $g_1, g_2 \in G_1$, algorithms to compute $e(g_1, g_2) \in G_2$ can finish within a polynomial time

(2) Bilinearity: for any integers $x, y \in [1, p]$, $e(g^x, g^y) = e(g, g)^{xy}$

(3) Nondegeneracy: if g is a generator of G_1, then $e(g, g)$ is a generator of G_2. In other words, this can be simplified as $e(g, g) \neq 1$

The size of G_1, G_2 is determined by the security parameter.

2.2.2. Decisional Bilinear Diffie-Hellman (DBDH) Assumption. Suppose an adversary chooses random a, b, c, $z \in \mathcal{Z}_p$, the DBDH assumption [6] means that there is no adversary, who can distinguish the tuple $(A = g^a, B = g^b, C = g^c, Z = e(g, g)^{abc}$ from the tuple $(A = g^a, B = g^b, C = g^c, Z = e(g, g)^z$, within a probabilistic polynomial time with a nonnegligible advantage.

2.3. Blockchain and Smart Contract. The consortium blockchain provides a permissioned design, which is different from the private blockchain and the public blockchain. It does not require the same level of strict control and restriction as the private blockchain. Meanwhile, it is not completely decentralized as the public blockchain. The consortium blockchain is maintained and operated by several trusted nodes. Besides, the consortium blockchain has the advantages of promoting openness and collaboration, marvelous data control, and node management and speeding up the operation of the system. In this paper, we build a medical data sharing system based on Hyperledger Fabric. Hyperledger is an open source project under the Linux Foundation and is supported by companies such as IBM, Intel, and Sap. Hyperledger Fabric is one of the implemented blockchains. This consortium blockchain has the advantages of high throughput, low latency, and scalability. It is a popular choice in the industrial blockchain scenario.

2.4. Access Control. Attribute-based access control (ABAC) is a type of access control technology that considers attributes, objects, permission, and environment as input. It determines whether to grant authorization by examining whether the object contains the proper attributes. ABAC can provide fine-grained access control, which can support a large number of input decision sets, define many possible rules, and express many strategies, but with only limited computing consumption and attributes. Such flexibility can decouple the relationship between the subject and the object. For instance, an employee is given attributes in the subject's attribute set, e.g., a nurse in a certain hospital. Objects are given attributes when they are created, e.g., medical data of patients with diabetes. Object owners will define attributes to set access control rules at the beginning of the creation, e.g., all nurses in the hospital can access the medical data of patients with diabetes. Under attribute-based access control, access decisions can be modified by simply changing the attribute value without affecting the relationship between a single subject and an object. Hence, ABAC can provide more dynamic and flexible access control management capabilities, reducing long-term maintenance costs. In addition, ABAC can be performed without any knowledge of new subjects, which means that there is no need to modify the existing rules.

3. Problem Formulation

3.1. Problem Scenario. There are four main roles in our problem scenario, data owner, medical institution, and data user. The respective responsibility and operations of these roles are as follows.

(i) Data owners, i.e., patients, share their medical data with medical institutions, including personally identifiable information and medical data

(ii) Medical institutions extract medical data keywords, perform keyword search, and conduct access control

(iii) Data users are entities that need to be authorized to perform keyword searches and obtain the required data, such as research institutions, insurance companies, or government departments. Data users need to be authorized first by access control and then can perform a keyword search on medical data

3.2. Security Goal. Here, we address the security goals as follows in our scheme.

3.2.1. Privacy Preserving. With all these patient personal information and medical data involved in this scheme, the major challenge is to preserve the privacy of all patients. No matter who wants to approach the patients' health data, the identity information must be controlled or limited.

3.2.2. Security Search. If data users want to obtain patients' health data, they have to get through the access control mechanism. Meanwhile, the searching and query process should not leak information related to the keywords.

3.2.3. Data Integrity and Reliability. In addition to data privacy issues, there are data security issues. The patient's data should not be tampered and deleted after uploading, i.e., the integrity and reliability of the data should be guaranteed.

3.3. Notions

(i) SID: the patient identifier, which can be used to denote one specific patient, such as the social security number

(ii) QID: quasi-identifier, such as zip code, address, age, gender, and birthday

(iii) M: the medical institution collection, denoted as a set of m medical institutions $M = (M_1, M_2, \cdots, M_m)$

(iv) D_i: the plain medical data text of M_i, denoted as a set of n data $D_i = (D_{i,1}, D_{i,2}, \cdots, D_{i,n})$

(v) C_i: the ciphertext of medical data; medical data can be encrypted by a medical institution, denoted as $C_i = (C_{i,1}, C_{i,2}, \cdots, C_{i,n})$

(vi) W: the keywords of medical data that can represent a medical data, denoted as a set of u keywords $W = (w_1, w_2, \cdots, w_u)$

(vii) $\widehat{W}$: encrypted keyword of W, medical institutions' encrypted keyword collection, denoted as $\widehat{W} = (\widehat{w}_1, \widehat{w}_2, \cdots, \widehat{w}_u)$

(viii) $\tilde{W}$ represents the queried keywords and the subset of the keywords W, denoted as a set of q keywords $\tilde{W} = (w_1, w_2, \cdots, w_q)$

(ix) T: the trapdoor for $\tilde{W}$, denoted as $T = (T_{w_1}, T_{w_2}, \cdots, T_{w_q})$

(x) K_{pub}: the public key

(xi) K_{priv}: the private key

(xii) λ: the security parameter

4. System Design and Technique Details

To tackle the issue of privacy protection and access control for sensitive medical data, we propose and develop a system based on the consortium blockchain platform with K-anonymity and searchable encryption techniques. The technologies adopted in this paper are designed and selected for medical data and medical data application scenarios. Medical data mainly has two kinds of privacy characteristics: data privacy and identity privacy. And the application scenarios of medical data are the coexistence of multiple medical institutions, similar to the P2P network in the computer network. The following is a detailed analysis and explanation.

(1) For data privacy, this paper extracts keywords from medical data and further encrypts and searches the keywords with a searchable encryption technology, which not only protects the privacy of data but also ensures the availability of data

(2) As for identity privacy, the current technologies to protect identity privacy include differential privacy and K-anonymity. Based on the characteristics of medical data, K-anonymity is chosen in this paper. The main consideration is that differential privacy will change the original statistical data characteristics of medical data, and that is more important to medical data. K-anonymity can retain the statistical characteristics of data. Therefore, we choose K-anonymity

(3) For the application scenario of medical data sharing, this paper adopts the consortium blockchain. Each medical institution is acting as one node of the consortium blockchain. The consortium blockchain is in line with the medical data sharing scenario of a peer-to-peer network and authorized access. On the contrary to the public blockchain, the data stored and managed by the consortium blockchain can keep being secured and private, instead of being totally transparent

In this section, we first introduce the system overview and then zoom into the detailed techniques adopted.

4.1. System Overview. Figure 1 shows the architecture overview of our system. We introduce the consortium blockchain as the middleware to perform as the trust layer. After patients upload data, i.e., medical data, to medical institutions. The medical institutions need to preprocess the data with a K-anonymity technique to blur some sensitive data fields which can probably reflect the patient's identity. Then, the keywords are extracted from the dataset and form the index. We design a scheme based on searchable encryption to encrypt the dataset and the index, i.e., the keyword. The entire encrypted dataset is uploaded to the cloud and managed by the medical institution itself. The encrypted data are uploaded to the consortium blockchain platform, which is constructed by different medical institutions. Each medical institution acts as a participant node in the platform. We develop smart contracts and deploy them on the blockchain platform. The smart contracts provide all the related interfaces for medical institutions and data users to leverage. One of the interfaces is designed for data users to query with the keyword index. According to our searchable encryption scheme, the corresponding results can be fetched back to the data user from clouds without decrypting the secured data. The searching process is also implemented by the smart contract. Besides, the smart contract also provides the interface of access control based on the attributes of data users.

In this paper, K-anonymity and searchable encryption are different from traditional application scenarios. The implementation of K-anonymity in this paper is different from the traditional way of centralized processing and centralized storage. The process of the K-anonymity algorithm adopted in this paper is processed by distributed nodes. The results of processing are stored on the consortium blockchain, and all nodes in the blockchain network can access them. As for searchable encryption, in traditional application scenarios, the processing and calculation modes are centralized. In this paper, the searchable encryption scheme based on the blockchain platform is implemented with smart contracts, including the calculation process of trapdoor generation and keyword matching, which are also in a decentralized manner.

The consortium blockchain network in this paper is based on blockchain nodes, and each medical institution acts as one node of the blockchain network. For the newly joined nodes, i.e., medical institutions, they need to apply for certificates from the authority node and then join the blockchain network with certificates. Among them, the authority node is the root of trustworthiness in the consortium blockchain, which can be played by the authoritative management department of the medical institution.

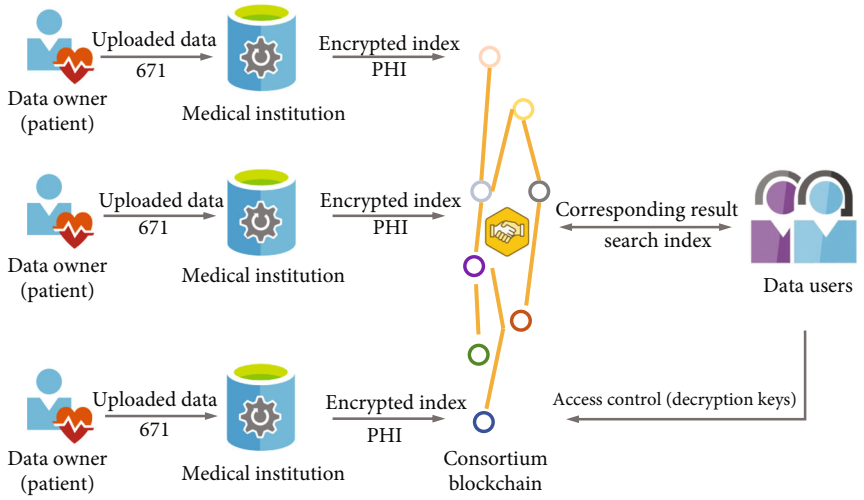

Figure 1: The system architecture overview.

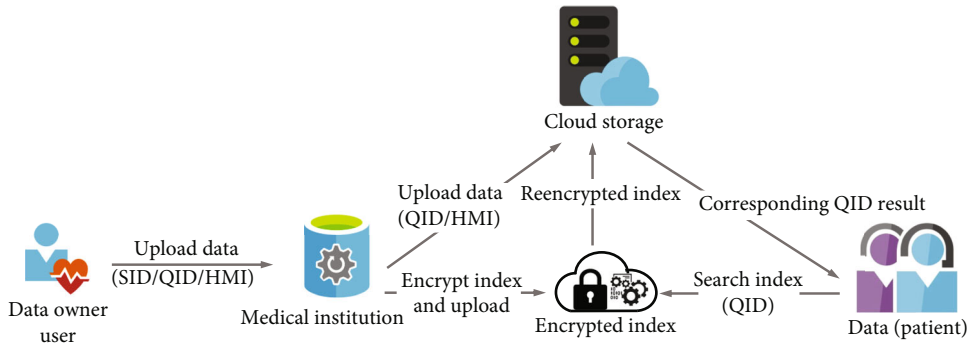

Figure 2: The process of medical data preprocessing, uploading, and querying with searchable encryption.

Figure 1 shows the entire architecture with multiple medical institutions. For the following parts in this section, we zoom into the details and explain how the techniques are leveraged. Hence, for a specific medical institution, Figure 2 demonstrates how the medical institution preprocesses the data and uploads the data to a remote cloud. The personal SID is firstly removed and then the K-anonymity is adopted to blur the QID. Afterwards, the preprocessed data is encrypted and uploaded to the remote cloud. Finally, data users can perform queries without decrypting the data. The related technique is described as follows.

4.2. K-Anonymity. Mondrian multidimensional partitioning [7] is a K-anonymous multidimensional partitioning algorithm with K-anonymous processing in two steps. In the first step, the multidimensional regions covering all the domain space attributes are defined, i.e., the partition stage constructs kd-trees [8]. The second step is to construct functions for data recoding.

The partitioning algorithm is described in Algorithm 1. In the algorithm, each dimension selects the dimension and the value of the partition. In the literature of kd-trees, one approach is to use the median as the value of the partition. The partition is completed to get k-groups. and each k-group spontaneously includes at least k records. Each k-group is then generalized. Thereby, the QID of each group is the same.

Input: Table S to be partitioned.
Output: Partitioning result.
1: Anonymize(partition)
2: **if** (no allowable multidimensional cut for partition) **then**
3: return ϕ: partition $\longrightarrow$ summary
4: **else**
5: dim $\longleftarrow$ choose_dimension()
6: fs $\longleftarrow$ frequencySet(partition, dim)
7: splitVal $\longleftarrow$ find.median(fs)
8: lhs $\longleftarrow$ t $\in$ partition: t.dim $\leq$ splitVal
9: rhs $\longleftarrow$ t $\in$ partition: t.dim > splitVal
10: return Anonymize(rhs) $\cup$ Anonymize(lhs)
11: **end if**

Algorithm 1: Mondrian partitioning algorithm.

For the selection of parameter K in the algorithm, the principle we follow is to get the value of the K-anonymity parameter K after the practical test in the practical application scenarios, so as to protect the patient's identity privacy and not make the system query results too redundant.

Table 1: Patient data.

Age	Sex	Zip code	Disease
25	Male	53711	Flu
25	Female	53712	Hepatitis
26	Male	53711	Bronchitis
27	Male	53710	Broken arm
27	Female	53712	AIDS
28	Male	53711	Hang nail

Table 2: A 2-anonymity example.

Age	Sex	Zip code	Disease
[25–26]	Male	53711	Flu
[25–27]	Female	53712	Hepatitis
[25–26]	Male	53711	Bronchitis
[27–28]	Male	[53710–53711]	Broken arm
[25–27]	Female	53712	AIDS
[27–28]	Male	[53710–53711]	Hang nail

4.2.1. Before Processing. For example, supposing that the patient data structure is shown in Table 1, QID includes the age, sex, and zip code. These attributes that appear in private personal data may also appear in public datasets. If the two sets of data are linked together, the patient's private data may be leaked. Therefore, these attributes need to be processed by the K-anonymous algorithm to avoid privacy leakage.

4.2.2. After Process. The multidimensional anonymization of patients is shown in Table 2. It shows that with the condition of 2-anonymity, each record has another one record, whose QID attributes are exactly the same.

4.3. Searchable Encryption. In this paper, we assume that there are several medical institutions with different keys that participate. Then, we are faced with the problem of how to search the different key encrypted keywords in multiple medical institutions. To achieve a secure search for multiple medical institutions, we have adopted a secure search scheme that satisfies the following three conditions:

(i) Different medical institutions encrypt keywords with their own keys

(ii) The data user does not need to know the key when generating the trapdoor

(iii) After retrieving the trapdoor, the cloud server can search for the corresponding data content through the keywords without knowing the specific value, as the keywords are encrypted by multiple medical institutions

The cloud server mentioned in this section refers to the computing and storage resources provided by the public cloud service provider (CSP), to form a running unit and provide services to clients. In addition, the middle server is provided by a specific trusted organization to provide the system with completely credible and reliable services, such as secret keys and crucial information storage.

Figure 2 shows the detailed process of using searchable encryption to manage the data. It is worth mentioning that the encrypted index is stored in the consortium blockchain and the encrypted dataset is uploaded to a remote cloud managed by a third party. In order to explain the procedure shown in Figure 2, we use an example to illustrate the details of this scheme. Medical institution i, i.e., M_i, needs to encrypt D_i into C_i with its own key before sharing medical data D_i. At the same time, in order for the data user to be able to perform search, the medical institution needs to extract the keyword $w_{i,h}$ from the document and send the encrypted keyword $\widehat{w}_{i,h} = (E_a{}', E_0)$ to the middle server. The middle server is further encrypted, $E_a{}'$ to E_a, and obtains $\widehat{w}_{i,h} = (E_a, E_0)$; then, the result is sent to the cloud server. Next, assume data user U wants to search for a document related to the keyword $w_h{}'$. Basically, he needs to generate a trapdoor $T'_{w_h{}'}$ and upload it to the middle server. The middle server then reencrypts the trapdoor $T'_{w_h{}'}$ to obtain $T_{w_h{}'}$, while generating secret data S_a. Then, $T_{w_h{}'}$ and S_a are uploaded to the cloud server. The cloud server finally calculates $\widehat{e}(E_o, T_3) = \widehat{e}(E_o, T_1) \cdot \widehat{e}(S_a, T_2)$ for keyword search.

4.3.1. Encryption Construction. The construction is based on a bilinear map. We define g to be the generators of the cyclic groups, G_1 and G_2, whose orders both are p. $\widehat{e}$ is a bilinear map $\widehat{e} : G_1 \times G_1 \longrightarrow G_2$. In the process of encryption construction, the random key generation algorithm generates different keys for different inputs. $k_{m1} \in \mathbb{Z}_p^+$, $k_{m2} \in \mathbb{Z}_p^+$, $k_{i,w} \in \mathbb{Z}_p^+$, $k_{i,d} \in \mathbb{Z}_p^+ \longleftarrow (0,1)^*$. k_{m1} and k_{m2} are the private keys of the middle server; $k_{i,w}$ and $k_{i,d}$ are the private keys used to encrypt keywords and data of medical institution M_i, respectively. $H(\cdot)$, located in $\mathbb{Z}_p^+$, is a hash function.

4.3.2. Keyword Encryption. The keys of different medical institutions are different in this system, and the ciphertext generated each time for the same keyword is different. Therefore, even if the key is lost, the data cannot be leaked, since the cloud server cannot obtain any information about the keyword. For the hth keyword of the medical institution M_i, i.e., $w_{i,h}$,. the process of encryption calculation is as follows.

$$\widehat{w}_{i,h} = \left(g^{k_{i,w} \cdot r_o \cdot H(w_{i,h})}, g^{k_{i,w} \cdot r_o}\right), \tag{1}$$

where r_o is a number generated randomly each time, which is leveraged to calculate $E'_a = g^{k_{i,w} \cdot r_o \cdot H(w_{i,h})}$ and $E_o = g^{k_{i,w} \cdot r_o}$.

The medical institution submits $\widehat{w}_{i,h} = (E'_a, E_o)$ to the middle server, and the middle server reencrypts E'_a with its own keys, k_{m1} and k_{m2}, to obtain E_a, as follows.

$$E_a = \left(E'_a \cdot g^{k_{m1}}\right)^{k_{m2}}. \tag{2}$$

Finally, the middle server submits the $\widehat{w}_{i,h} = (E_a, E_o)$ to the cloud server. In the entire process, the middle server is always unable to know the specific value of the keyword.

4.3.3. Trapdoor Generation. In the scheme that we propose, data users do not need to know the key of the medical institution and the trapdoors generated for the same keyword each time are different. The trapdoor is generated in two steps. First, the data user generates a trapdoor based on the search key and the random number and then submits the trapdoor to the middle server. Second, the middle server reencrypts the trapdoor. Here, we assume that the data user wants to search for the keyword $w_h{}'$ and the encryption is calculated as follows.

$$T'_{w_h{}'} = \left(g^{H(w_{h'})\cdot r_u}, g^{r_u}\right), \tag{3}$$

where r_u is a random number generated randomly each time. After receiving the $T'_{w_h{}'}$, the middle server generates a random number r_m and reencrypts $T'_{w_h{}'}$ as follows.

$$T_{w_h{}'} = \left(g^{H(w_{h'})\cdot r_u\cdot k_{m1}\cdot k_{m2}\cdot r_m}, g^{r_u\cdot k_{m1}}, g^{r_u\cdot k_{m1}\cdot r_m}\right). \tag{4}$$

Let us make $T_1 = g^{H(w_{h'})\cdot r_u\cdot k_{m1}\cdot k_{m2}\cdot r_m}$, $T_2 = g^{r_u\cdot k_{m1}}$, $T_3 = g^{r_u\cdot k_{m1}\cdot r_m}$, i.e., $T_{w_h{}'} = (T_1, T_2, T_3)$. Finally, the middle server submits $T_{w_h{}'}$ to the cloud server.

4.3.4. Keyword Matching. In the scheme that we proposed in this paper, the cloud server stores encrypted data and keywords for all medical institutions. The middle server needs to transfer a secret data $S_a = g^{k_{m1}\cdot k_{m2}\cdot r_m}$ to the cloud server. After receiving the search request, the cloud server performs a global search to match all stored keywords, in order to obtain the corresponding medical data. The searching process is described as follows. First of all, the cloud server performs the following calculations, after getting trapdoors, $T_{w_h{}'}$ and (E_a, E_o).

$$\widehat{e}(S_a, T_2) = \widehat{e}\left(g^{k_{m1}\cdot k_{m2}\cdot r_m}, g^{r_u\cdot k_{m1}\cdot r_m}\right) = \widehat{e}(g, g)^{r_m\cdot k_{m1}\cdot k_{m2}\cdot r_u\cdot k_{m1}}. \tag{5}$$

Then, the cloud server judges whether w_h equals to $w_h{}'$, according to the following equation.

$$\begin{aligned}
\widehat{e}(E_a, T_3) &= \widehat{e}\left(\left(g^{k_{i,w}\cdot r_o\cdot H(w_{i,h})}\cdot g_{k_{m1}}\right)^{k_{km2}}, g^{r_u\cdot k_{m1}\cdot r_m}\right) \\
&= \widehat{e}(g, g)^{(k_{i,w}\cdot r_o\cdot H(w_{i,h})+k_{m1})\cdot r_u\cdot k_{m1}\cdot r_m} \\
&= \widehat{e}(g, g)^{k_{i,w}\cdot r_o\cdot H(w_{i,h})\cdot r_u\cdot k_{m1}\cdot r_m}\cdot \widehat{e}(S_a, T_2) \\
&= \widehat{e}\left(g^{k_{i,w}\cdot r_o}, g^{H(w_{i,h})\cdot r_u\cdot k_{m1}\cdot r_m}\right)\cdot \widehat{e}(S_a, T_2) \\
&= \widehat{e}(E_o, T_1)\cdot \widehat{e}(S_a, T_2).
\end{aligned} \tag{6}$$

4.4. Consortium Blockchain and Attribute-Based Access Control. The consortium blockchain is the crucial part of the entire scheme. All new nodes must get the certification from the fabric-CA before participating in the consortium blockchain. The consortium blockchain implements a chaincode of searchable encryption and ABAC. At the same time, the blockchain exposes interfaces for users to access blockchain data, including access control interfaces and data interfaces. The functions implemented by the consortium blockchain are as follows:

(i) The chaincode implements the function for uploading encrypted data to the ledger of the blockchain

(ii) The chaincode provides the function of searchable encryption

(iii) The chaincode provides ABAC to manage the permissions for user access

For the ABAC part, policy management and access control are separated. Policies may vary with the actual scenario, for instance, by increasing or decreasing the number of attributes to accommodate larger or smaller scenarios. The key part of ABAC is the attribute which can be defined as $A \in \{S, O, P, E\}$. The definition of each field is as follows:

(i) A indicates the attribute. Each attribute has an identifier

(ii) S indicates the subject's attributes, i.e., the identity and characteristics of the subject which can perform the access request, for instance, the entity's name, age, and occupation

(iii) O indicates the object's attributes, i.e., the information related to the accessed resource, for instance, resource type, service location, and protocol

(iv) P indicates the permission, i.e., the operation of the subject which can be performed on the object, for instance, reading, writing, and executing

(v) E indicates the environment, i.e., the environment information when the access request is initiated, for instance, time and location

The attribute-based access control policy can be defined as $\{S \wedge \text{or} \vee O \wedge \text{or} \vee P \wedge \text{or} \vee E\}$, which indicates the access control rules of the subject to access the object. It expresses the required attribute set for accessing the protected resources. The above expression means that the XOR relationship among the subject, object, permission, and environment constitutes the access grants of the access control mechanism.

The attribute-based access control request can be defined as $\{A \wedge O \wedge P \wedge E\}$, as mentioned above, which is a set of attributes. It indicates the operation of the subject on the object under the environment. When users access the ABAC system, the attribute set constitutes the access request operation, which is used as the input parameter to initiate the access request of the access control mechanism in the system.

5. Security Analysis

In this section, we discuss that our scheme satisfies the following security goals.

5.1. Privacy Preserving. Patient's personal information and medical data are protected by searchable encryption and attribute-based access control. Medical institutions encrypt medical data, and the consortium blockchain provides access control, which is a proper way to avoid privacy disclosure to malicious entities.

5.2. Security Search. Data users need to obtain authorization of access control before they can search on the consortium blockchain. Therefore, malicious entities cannot initiate searches after access control, which guarantees the security of the data.

5.3. Data Integrity and Reliability. The patients' medical data with encrypted keywords are uploaded by the medical institutions to the consortium blockchain. Thus, during the process of data storage and transmission, no one can modify or read data without the authorization of the medical institutions.

5.4. Scheme Properties. As for scheme properties, we compare the properties of our scheme with [9, 10] of the cloud-based one and [11] of the blockchain-based one. As shown in Table 3, our scheme can meet all the scheme properties which are vital properties of medical data sharing schemes.

6. Experimental Study

In this section, we implement the proposed scheme on the Hyperledger Fabric platform and evaluate its performance. Especially, we simulate different numbers of medical institutions to construct the blockchain platform for testing the performance and scalability of our system. For testing purposes, we leverage the docker to build a consortium blockchain platform with Hyperledger Fabric, which is contrusted using CloudsStorm [12] in the Cloud environment. The programming language is Python3.7 and Go1.12; the Fabric edition is 1.4. For the underlying server, we set up the environment with a virtual machine from the ExoGENI [13] cloud. The entire virtual machine is only for this purpose. The configuration of the virtual machine is the type of "XOXLarge," which contains 4 CPU cores and a 12 GB RAM.

6.1. Computational Cost. In order to evaluate this scheme quantificationally, we have also conducted some experiments based on our scheme. We provide the details of the computational cost of functions in Table 4.

In our scheme, the *KeyGen* function is responsible for generating the public key and private key. And, the function *Enc* is used to encrypt the keywords of medical data. The trapdoor can be generated by the function of *TdGen*. Finally, the function of *Search* is for searching the expected keyword and the result can be "True" or "False."

Due to the fact that the computational cost of these algorithms is related to keyword numbers, we test our algorithms by setting keyword numbers as $w = 10$, 50, and 100. As shown in Table 4, we can find out that the time cost of all functions, including *KeyGen*, *EncIndex*, *TdGen*, and *Search*, increases with the size of keyword amounts linearly.

Among these four main operations, three of them are implemented in the smart contracts of the consortium blockchain, including *Enc*, *TdGen*, and *Search*. Hence, to test the system performance for these on-chain operations, we measure the computational cost of these operations varying with different frequencies of operation requests, i.e., the medical data accessing rate in TPS (transactions per second). The number of keywords to be processed is set to once in each transaction, i.e., for each data accessing request. Five medical institutions are simulated here to construct the 5-node blockchain network. The experimental results are shown in Figure 3. It shows that the computational cost of all the on-chain operations is linear with the medical data accessing rate, which demonstrates that our system is scalable.

Table 3: Scheme properties.

	Liu [9]	Wang [10]	Azaria [11]	Proposed scheme
Blockchain	N	N	Y	Y
Access control	Y	Y	N	Y
Privacy preserving	Y	Y	N	Y
Searchable encryption	N	Y	N	Y

Table 4: Computational cost (in milliseconds).

	KeyGen	*Enc*	*TdGen*	*Search*
$w = 10$	71.49	115.66	83.39	620
$w = 50$	359.70	575.78	432.45	3510
$w = 100$	712.46	1168.90	880.17	6830

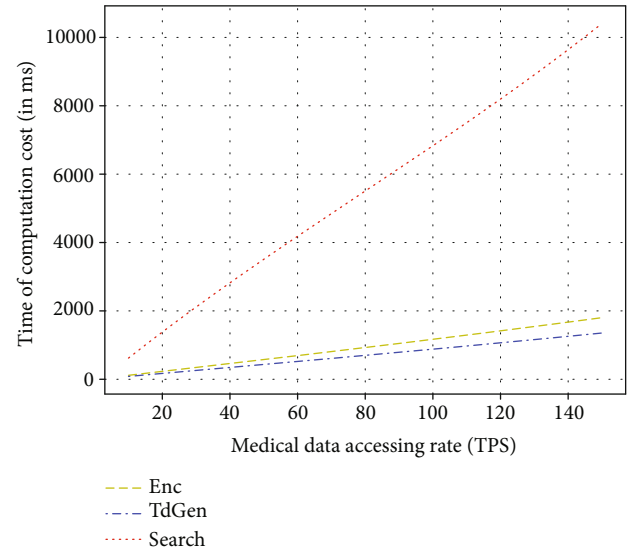

Figure 3: The computational cost of on-chain operations varying with different data access rates.

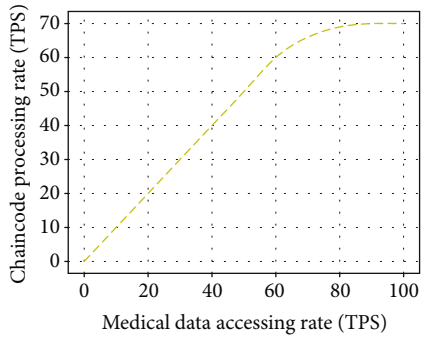

Figure 4: The performance of the keyword searching operation with searchable encryption under the scenario of 5 medical institutions.

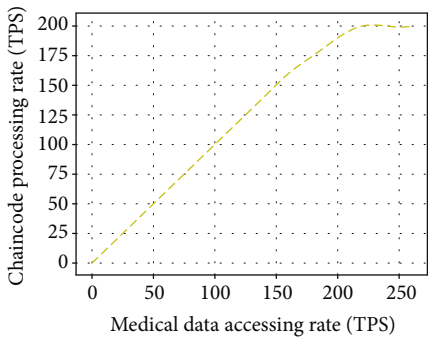

Figure 5: The performance of access control with the ABAC model under the scenario of 5 medical institutions.

6.2. Performance of the Smart Contract. In this subsection, we mainly test the performance of two types of chaincodes, i.e., smart contracts. One is designed for storing and searching the dataset, and the other is for attribute-based access control (ABAC). For this experiment, we simulate that there are five medical institutions, i.e., five nodes construct the experimental blockchain platform. Figure 4 shows the chaincode processing performance of the searching operation with searchable encryption when increasing the medical data accessing rate. In this scenario, the processing rate of the chaincode for searching operations increases linearly according to the increased medical data accessing rate. The system is able to handle all the requests when the accessing rate is not high. However, when the rate of accessing requests comes to 70 TPS, the processing rate cannot increase anymore. It demonstrates that the throughput of our system with 5 medical institutions is around 70 TPS. For the chaincode of the ABAC model, the performance is similar as shown in Figure 5. But the throughput for the ABAC chaincode is better than the chaincode with searching operations, which is around 200 TPS. As the medical information sharing is mainly for research purposes, the amount of requests is not at a large scale. Hence, the performance of our system is acceptable.

On the other hand, the memory consumption of the chaincode for realizing searchable encryption is also measured. The measurements are performed within the node where the chaincode for encryption is invoked. For testing,

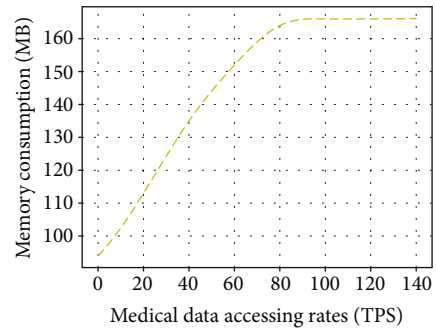

Figure 6: The memory consumption of the chaincode for performing operations of searchable encryption.

we input the system with different rates of medical data accessing requests. The experimental results shown in Figure 6 also demonstrate that the system is able to handle the accessing rate blow of approximately 80 TPS. When the accessing rate further increases, the memory consumption stays steady afterwards. The reason is that the system has reached its capacity. Meanwhile, it is worth mentioning that the maximum memory consumption for running searchable encryption chaincodes is around 170 MB. It is, therefore, practical for each medical institution to operate a server for running the chaincode and participant of the blockchain network.

6.3. Scalability. In this section, we test the scalability of our implemented prototype. We still mainly test two types of chaincodes, i.e., for searchable encryption and ABAC. Since these two parts are crucial to our system, the scalability of these chaincodes can determine the scalability of the entire system. Hence, we assume various scenarios, under which there are different numbers of medical institutions. To check the trend, we simulate 5, 10, 15, and 20 medical institutions. It means that the scales of the blockchain platform are 5, 10, 15, and 20. Then, we increase the medical data accessing rate as system inputs until the chaincode processing rate is getting steady. The procedure is similar to the experiment conducted in Section 6.2. In this way, the capacity of system throughput is achieved.

Figures 7 and 8 show the experimental results of chaincodes for searchable encryption and ABAC, respectively. For both of these two types of chaincodes, the descending rates of their performance are not even linear to the increased number of blockchain participants, i.e., medical institutions. However, with the configurations of our experiments, the system capacity has not decreased much for each type of chaincode when the number of simulated medical institutions increases from 5 to 20 (70 TPS to 60 TPS and 200 TPS to 165 TPS, respectively). In practice, not all the medical institutions construct a single federation to share medical data; the scalability of our system is still feasible. However, for a large-scale federation with many institutions, the system still needs to be further optimized.

7. Related Work

The data privacy and security are hot topics in recent years. Some papers consider the issue related with different stages

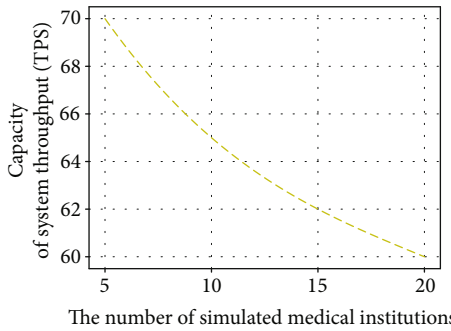

Figure 7: The capacity of system throughput varying with simulating different numbers of medical institutions for searchable encryption chaincodes.

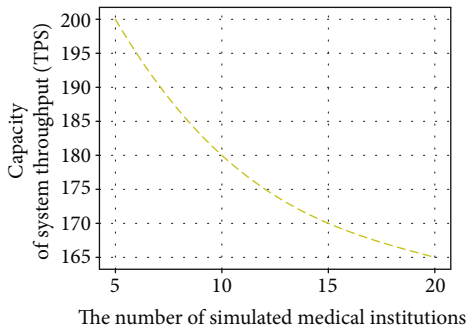

Figure 8: The capacity of system throughput varying with simulating different numbers of medical institutions for ABAC chaincodes.

of the data management, including data uploading [14], auditing [15], and sharing [16]. Other papers consider these issues in a special scenario, e.g., in a wireless environment [17]. We focus on the same issue in the field of medical data [18, 19].

Medical data usually includes the types of health reports, medical records, examination results, etc. These types of data not only have great research value but also have privacy preserving requirements. Meanwhile, the integrity of medical data requires assurance. We also need to guarantee that these data are not to be tampered with, destroyed, or deleted by anyone without grants. The accessibility of medical data must be controlled by the patients, but it cannot be modified by the patient. In addition to this, patient data should also be able to be circulated smoothly among medical institutions [20, 21].

7.1. Medical Data Privacy Preserving Based on Confusion and Anonymity. The anonymized data or differential privacy was introduced to protect the privacy of medical data [22]. Beaulieu-Jones et al. [23] used a deep neural network that introduced differential privacy to generate artificial data when sharing medical data, which solved the contradiction between data sharing and data privacy to a certain extent. However, in the context of multiple data types, the limitation of this method is that the data would be modified. Cai et al. [24] developed a differential-private framework to preserve the sensitive information for taxi companies when sharing taxi data. Sun *et* al. [25] proposed a method similar to the above, using differential privacy to process medical data, training a machine learning model, and publishing the training model instead of directly publishing private data.

7.2. Privacy Preserving of Medical Data Based on Encryption. Searchable encryption is usually leveraged in the edge environment for mobile computing [26], through being introduced into data privacy-preserving schemes. For example, Xu et al. [27] leverage a multikeyword searchable encryption technology in medical data sharing to protect data privacy. At the same time, they utilize searchable encryption to achieve the goal of sharing. In the scenario of a body equipment, the body sensor device is leveraged to encrypt the sensor data and upload to the cloud in the stage of collecting the data segment, which is similar to the scenario described by the paper [28]. Meanwhile, the searchable encryption technology is also used to share the data. Besides, this solution only requires little resource consumption, which is suitable for the mobile devices. However, the problem of sensor device credibility should also be considered.

7.3. Privacy Preserving Based on the Blockchain. Blockchain is an emerging technology to enhance the trust for the legacy systems, which has been applied in the cloud service level agreement [29], crowdsourcing [30], etc. It is also applied into aspects of data storage with privacy preserving, since the blockchain has storage characteristics of high reliability, high availability, low cost, and strong disaster tolerance. Do and Ng [31] combine searchable encryption and blockchain to ensure that data cannot be tampered with and deleted. It also adopts smart contracts to implement access control and searchable encryption, which constructs the distributed system and solves the trust problem among nodes. Similarly, the consortium blockchain combines searchable encryption [32] that uses agents to achieve searchable encryption. The consortium blockchain stores keywords and ciphertext. To a certain extent, this solution guarantees the privacy-preserving requirements of medical data, which can realize the goal that the distributed medical data cannot be tampered with. In the scheme proposed by Chen et al. [33], medical data is stored on a public cloud and the index of medical data is stored on the blockchain. Data users firstly need to be authorized when they want to obtain data, so that data owners can fully control their own data. However, it is a critical problem that the public cloud is not safe. Tian et al. [34] propose the scheme of SIFF based on the fabric blockchain; SIFF guarantees the granularity of data search. This scheme ensures the privacy, availability, and integrity of medical data. Zhang and Lin [35] propose a medical data sharing architecture combining a private blockchain, which stores full data, and a consortium blockchain, which stores keywords. At the same time, searchable encryption is leveraged to search for keywords. Encrypted medical data is obtained by grants of access control. Likewise, Azaria et al. [11] use blockchain to store medical data and access control is combined with smart contracts. Utilizing blockchain ensures that medical data cannot be tampered with and provides medical data interconnection, interoperability, and data sharing features. MedChain [36] is also a medical data sharing architecture that combines blockchain and P2P networks to tackle

efficiency issues. There are two types of nodes in MedChain; one is a super node, which performs massive calculation and storage, such as large hospitals. The other is an edge node, such as clinics and community hospitals. Although this system is not yet perfect, the scheme of MedChain still improves efficiency, privacy, and security. In addition, MedBlock [37] is also a system that uses blockchain to manage medical data. It adopts a hybrid consensus mechanism, combined with Delegated Proof of Stake (DPoS) and Practical Byzantine Fault Tolerance (PBFT), which improves efficiency. It is worth mentioning that the symmetric encryption algorithm and access control are leveraged to improve privacy and security.

8. Conclusions

In our proposed work, we have presented a consortium blockchain-based medical data sharing system using K-anonymity, keyword searchable encryption, and ABAC to achieve data privacy-preserving and security among different medical institutions. Firstly, we leverage the K-anonymity technique to preprocess the medical data for blurring the identity information. Secondly, we present a scheme for medical data sharing using searchable encryption on keywords to ensure data security and privacy-preserving. The consortium blockchain is leveraged among different medical institutions to provide the trust layer and host the smart contract. Hence, the consortium blockchain stores the encrypted keywords which are linked to the medical data of medical institutions. The encrypted medical data can then be stored safely on remote clouds. Thirdly, we design and implement attribute-based access control with a smart contract. The blockchain-based ABAC model simplifies the configurations and secures the medical data. Furthermore, we conduct a security analysis of the proposed scheme and protocol and compare them with other related work. The security analysis demonstrates that our scheme can meet the security goals of our original design. Finally, we also implement the scheme on the Hyperledger Fabric platform. Through simulating different medical data accessing rates and different numbers of medical institutions, we evaluate the computational overhead of encryption operations, the performance of implemented chaincodes, and the scalability of our prototype. The experimental studies demonstrate that our scheme and system design are feasible and practical to encourage medical data sharing among medical institutions.

Data Availability

Data are available using the following: https://github.com/mythsand/privacy-preserving-medical-data.

Disclosure

The initial version of this paper [38] was presented at the conference of "WASA: International Conference on Wireless Algorithms, Systems, and Applications."

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work is supported by the National Key Research and Development Program of China under grant 2018YFB0204301, the National Natural Science Foundation (NSF) under grant 62072306, and Open Fund of Science and Technology on Parallel and Distributed Processing Laboratory under grant 6142110200407.

References

[1] M. J. Steinberg and E. R. Rubin, *The HIPAA Privacy Rule: Lacks Patient Benefit, Impedes Research Growth*, Association of Academic Health Centers, 2009.

[2] P. Samarati and L. Sweeney, *Protecting Privacy when Disclosing Information: K-Anonymity and Its Enforcement through Generalization and Suppression*, Electronic Privacy Information Center, 1998.

[3] C. Dwork, "Differential privacy: a survey of results," in *Theory and Applications of Models of Computation. TAMC 2008*, pp. 1–19, Springer, 2008.

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

[5] Y. Wu, J. Su, and B. Li, "Keyword search over shared cloud data without secure channel or authority," in *2015 IEEE 8th International Conference on Cloud Computing*, pp. 580–587, New York, NY, USA, 2015.

[6] E. Shi, J. Bethencourt, T. H. H. Chan, D. Song, and A. Perrig, "Multi-dimensional range query over encrypted data," in *2007 IEEE Symposium on Security and Privacy (SP'07)*, pp. 350–364, Berkeley, CA, USA, 2007.

[7] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Mondrian multidimensional k-anonymity," *ICDE*, vol. 6, p. 25, 2006.

[8] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic Expected time," *ACM Transactions on Mathematical Software*, vol. 3, no. 3, pp. 209–226, 1977.

[9] J. Liu, X. Huang, 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.

[10] X. Wang, A. Zhang, X. Xie, and X. Ye, "Secure-aware and privacy-preserving electronic health record searching in cloud environment," *International Journal of Communication Systems*, vol. 32, no. 8, article e3925, 2019.

[11] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "Medrec: using blockchain for medical data access and permission management," in *2016 2nd International Conference on Open and Big Data (OBD)*, pp. 25–30, Vienna, Austria, 2016.

[12] H. Zhou, Y. Hu, X. Ouyang et al., "CloudsStorm: a framework for seamlessly programming and controlling virtual infrastructure functions during the DevOps lifecycle of cloud applications," *Software: Practice and Experience*, vol. 49, no. 10, pp. 1421–1447, 2019.

[13] I. Baldine, Y. Xin, A. Mandal, P. Ruth, C. Heerman, and J. Chase, "Exogeni: a multi-domain infrastructure-as-a-service testbed," in *Testbeds and Research Infrastructure. Development of Networks and Communities*, pp. 97–113, Springer, 2012.

[14] Z. Cai and Z. Xu, "A private and efficient mechanism for data uploading in smart cyber-physical systems," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 766–775, 2020.

[15] T. Wang, Y. Mei, X. Liu, J. Wang, H.-N. Dai, and Z. Wang, "Edge-based auditing method for data security in resource-constrained internet of things," *Journal of Systems Architecture*, vol. 114, p. 101971, 2021.

[16] Z. Xu and Z. Cai, "Privacy-preserved data sharing towards multiple parties in industrial iots," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 968–979, 2020.

[17] X. Liu, M. S. Obaidat, C. Lin, T. Wang, and A. Liu, "Movement-based solutions to energy limitation in wireless sensor networks: state of the art and future trends," *IEEE Network*, vol. 35, no. 2, pp. 188–193, 2021.

[18] N. Li, N. Zhang, S. K. Das, and B. Thuraisingham, "Privacy preservation in wireless sensor networks: a state-of-the-art survey," *Ad Hoc Networks*, vol. 7, no. 8, pp. 1501–1514, 2009.

[19] A. Ukil, "Privacy preserving data aggregation in wireless sensor networks," in *2010 6th International Conference on Wireless and Mobile Communications*, pp. 435–440, Valencia, Spain, 2010.

[20] M. Li, S. Yu, K. Ren, and W. Lou, "Securing personal health records in cloud computing: Patient-centric and fine-grained data access control in multi-owner settings," in *Security and Privacy in Communication Networks. SecureComm 2010*, pp. 89–106, Springer, 2010.

[21] D. K. Mandl, P. Szolovits, and S. I. Kohane, "Public standards and patients' control: how to keep electronic medical records accessible but private," *BMJ*, vol. 322, no. 7281, pp. 283–287, 2001.

[22] M. Moussa and S. A. Demurjian, "Differential privacy approach for big data privacy in healthcare," in *Privacy and Security Policies in Big Data*, pp. 191–213, IGI Global, 2017.

[23] B. K. Beaulieu-Jones, Z. S. Wu, C. Williams et al., "Privacy-preserving generative deep neural networks support clinical data sharing," *Circulation: Cardiovascular Quality and Outcomes*, vol. 12, no. 7, p. e005122, 2019.

[24] Z. Cai, X. Zheng, and J. Yu, "A differential-private framework for urban traffic flows estimation via taxi companies," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 12, pp. 6492–6499, 2019.

[25] Z. Sun, Y. Wang, M. Shu, R. Liu, and H. Zhao, "Differential privacy for data and model publishing of medical data," *IEEE Access*, vol. 7, pp. 152103–152114, 2019.

[26] Y. Guo, F. Liu, Z. Cai, N. Xiao, and Z. Zhao, "Edge-based efficient search over encrypted data mobile cloud storage," *Sensors*, vol. 18, no. 4, p. 1189, 2018.

[27] C. Xu, N. Wang, L. Zhu, K. Sharif, and C. Zhang, "Achieving searchable and privacy-preserving data sharing for cloud-assisted e-healthcare system," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8345–8356, 2019.

[28] F. Altaf, M. Aditia, E. Saini, B. Rakshit, and S. Maity, "Privacy preserving lightweight searchable encryption for cloud assisted e-health system," in *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, pp. 310–314, Chennai, India, 2019.

[29] H. Zhou, X. Ouyang, Z. Ren, J. Su, C. de Laat, and Z. Zhao, "A blockchain based witness model for trustworthy cloud service level agreement enforcement," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pp. 1567–1575, Paris, France, 2019.

[30] S. Zhu, Z. Cai, H. Hu, Y. Li, and W. Li, "zkcrowd: a hybrid blockchain-based crowdsourcing platform," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4196–4205, 2019.

[31] H. G. Do and W. K. Ng, "Blockchain-based system for secure data storage with private keyword search," in *2017 IEEE World Congress on Services (SERVICES)*, pp. 90–93, Honolulu, HI, USA, 2017.

[32] Y. Wang, A. Zhang, P. Zhang, and H. Wang, "Cloud-assisted ehr sharing with security and privacy preservation via consortium blockchain," *IEEE Access*, vol. 7, pp. 136704–136719, 2019.

[33] L. Chen, W.-K. Lee, C.-C. Chang, K.-K. R. Choo, and N. Zhang, "Blockchain based searchable encryption for electronic health record sharing," *Future Generation Computer Systems*, vol. 95, pp. 420–429, 2019.

[34] H. Tian, J. He, and Y. Ding, "Medical data management on blockchain with privacy," *Journal of Medical Systems*, vol. 43, no. 2, p. 26, 2019.

[35] 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.

[36] B. Shen, J. Guo, and Y. Yang, "Medchain: efficient healthcare data sharing via blockchain," *Applied Sciences*, vol. 9, no. 6, p. 1207, 2019.

[37] 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.

[38] L. Meng, X. Hong, Y. Chen, Y. Ding, and C. Zhang, "K-anonymous privacy preserving scheme based on bilinear pairings over medical data," in *Wireless Algorithms, Systems, and Applications, WASA 2020*, pp. 381–393, Springer, 2020.