Hindawi
Security and Communication Networks
Volume 2022, Article ID 8996815, 12 pages
https://doi.org/10.1155/2022/8996815

Research Article

Blockchain Data Sharing Query Scheme based on Threshold Secret Sharing

Lu Chen,[1,2,3] **Xin Zhang**,[1,2,3] **and Zhixin Sun**[1,2,3]

[1]*Engineering Research Center of Post Big Data Technology and Application of Jiangsu Province, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
[2]*Research and Development Center of Post Industry Technology of the State Posts Bureau (Internet of Things Technology), Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
[3]*Engineering Research Center of Broadband Wireless Communication Technology of the Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*

Correspondence should be addressed to Zhixin Sun; sunzx@njupt.edu.cn

Received 11 December 2021; Accepted 1 March 2022; Published 6 April 2022

Academic Editor: Yuling Chen

Blockchain is a distributed ledger that combines technologies such as timestamp, cryptography, consensus mechanism, and peer-to-peer network. In the field of data recording and management, the blockchain data query scheme based on smart contracts consumes a lot of resources, and blockchain platforms that do not support smart contracts cannot achieve convenient data query. This study proposes a blockchain data sharing query scheme based on threshold secret sharing. The secret elements used to query data are shared through the Blakley space plane equation to limit the rights of the inquirer, ensuring the security of blockchain data query. At the same time, the Blakley space plane equation coefficient matrix is used to segment the data to be uploaded to the blockchain. It solves the problem that the data cannot be directly stored in the block due to their large size. It facilitates data uploading to the blockchain. The experimental results show that the additional time consumption of the secret sharing and recovery, data segmentation, and reconstruction of this scheme is much less than the block generation time. Therefore, this solution will not affect the normal operation of blockchain applications and can improve the security and the fault tolerance rate of data query.

1. Introduction

"Bitcoin: A Peer-to-Peer Electronic Cash System" [1], published in 2008, first introduced the concept of blockchain. This study describes the architectural concept of Bitcoin based on the peer-to-peer network [2], cryptographic mechanism [3], timestamp [4], and consensus mechanism [5]. At the same time, the concept of "chain of blocks" is introduced. The birth of Bitcoin has promoted the development of digital cryptocurrency. At this time, the focus of people's attention is still on the "currency" attribute of Bitcoin, rather than the blockchain technology at the core of Bitcoin. After 2015, the emergence and subsequent rapid development of Ethereum [6] and Hyperledger [7] have changed this phenomenon. This made the underlying blockchain technology more widely known and studied by more people. At present, the mainstream definition of blockchain in the industry is a combination of key technologies such as timestamps, cryptographic mechanisms, peer-to-peer transmission, and distributed consensus. It is a decentralized shared ledger with collective maintenance, immutability, security, and credibility [8].

According to different trust construction methods, blockchain can be divided into permissioned blockchain and nonpermissioned blockchain [9]. According to the different degrees of openness [10], the permission blockchain can be divided into consortium blockchain and private blockchain. The common application areas of blockchain can be divided into digital cryptocurrency, data recording and management, information security, and other fields [11].

Applications in the field of data recording and management include data storage, data authentication, information sharing, and copyright protection. Some researchers use blockchain for logistics data management. The blockchain-based logistics data security storage system stores logistics records in the blockchain to ensure that the entire logistics process can be audited. At the same time, the throughput of the blockchain-based IoT system is improved through the group-based POW mechanism [12]. Blockchain can also be used for public auditing of data, enabling the secure sharing of recorded information and the traceability of user identities [13]. Some researchers used the characteristics of blockchain technology to achieve efficient sharing of government information resources and digital rights management [14, 15]. The core of the application of blockchain in data recording and management scenarios is data, and these applications need to provide blockchain data query services, so the security of data queries is very important.

The smart contract is a computer program that can run premade rules on a distributed ledger. It can execute and verify the complex behavior of distributed nodes [16]. Smart contracts can be executed more effectively in a decentralized and trustless environment like the blockchain. At present, many blockchain applications have adopted smart contract-based methods to provide blockchain data query services. The query methods based on the smart contract deploy the data access rules directly into the blockchain smart contract after negotiation and approval. When the conditional parameters carried in the request meet the access rules, the smart contract will continue to execute, without relying on nonblockchain systems. However, the query of data in blockchain applications is often only associated with a specific node. The smart contract-based approach requires each node to run the deployed contract, and the resource consumption of blockchain applications is relatively large.

At the same time, in the digital encryption currency field, the data size that the block body of the early blockchain platform can accommodate does not match the needs of nondigital encryption currency scenarios. Therefore, some blockchain-based data applications can only store relevant data indexes on the blockchain, and the original complete data need to be stored on the off-chain storage system. Literature [17] proposed a medical consortium blockchain system based on the PBFT consensus mechanism. The consortium blockchain in the system is responsible for storing the ID of the transaction order corresponding to the traditional database table, and the specific transaction order and original medical data still need to be saved to the existing medical information system. In the electronic medical record sharing model [18], the desensitized electronic medical record will be stored off-chain, and the off-chain index of the desensitized medical record will become the leaf node of the Merkel tree in the block.

Therefore, in view of the blockchain data query requirements in the data recording and management scenario, this study builds a secure and reliable blockchain data sharing query scheme based on the threshold secret sharing mechanism and erasure codes. The main contributions of this study are as follows:

(1) A blockchain data sharing query scheme based on the Blakley threshold secret sharing mechanism is proposed. The secret elements used to query data are shared through the Blakley space plane equation, limiting the rights of blockchain data inquirers, and it further ensures the security of blockchain data query.

(2) An erasure code-based data uploading method is proposed. The coefficient matrix of the Blakley space plane equation is used as the erasure code encoding matrix to segment the data. This method not only solves the problem that the data cannot be directly stored in the block because of the large size of the block but also improves the fault tolerance rate during query.

(3) Simulation experiments and analysis of the scheme are carried out. The results show that the rate of secret sharing and recovery, data segmentation, and reconstruction of this scheme is higher than that of block generation. This scheme does not affect the normal operation of the blockchain system; at the same time, it improves the security of data query.

2. Related Work

Many researchers design blockchain data query methods based on smart contracts. In the blockchain-based digital archive protection and sharing system proposed in reference [19], the consortium blockchain manages digital archives and users, the public blockchain regularly stores block snapshots of consortium blockchain, and the private IPFS stores ciphertext digital archives. The management tasks of the members of the internal consortium blockchain are undertaken by the digital identity management contract. Literature [20] designed a medical data sharing platform based on blockchain. When the access statement sent by the data requester matches the consent statement specified by the data provider, the smart contract will continue to execute the medical data sharing process. In the medical privacy data sharing query model proposed in reference [21], the access list contract is deployed by the medical institution to frame the private data collection belonging to the medical research category. At the same time, users also have the right to remove institutions from their own data authorized access list. Literature [22] constructed a paid data query scheme based on smart contracts, which built data access clauses into smart contracts along with data hashes. Only users who meet the access terms can get the key in the contract and the data hash in the trusted environment and finally retrieve the data.

The above research all use smart contracts for data query, which requires each node of the blockchain to run deployed smart contracts. Therefore, the data query methods based on smart contracts will make the resource consumption of blockchain applications larger.

In addition to smart contract-based methods, many scholars at home and abroad currently adopt blockchain data query schemes based on cryptographic mechanisms. Huang et al. [23] stored the access strategy of attribute-based

encryption (ABE) and user attributes in the database. The metadata were uploaded to the blockchain system, and the blockchain was used to ensure data security. The solution uses attribute-based encryption to control the access and query of blockchain data. In order to improve the security of data sharing between enterprises, Wang et al. [24] proposed a shared query and access model based on ABE and blockchain, which is composed of two blockchains responsible for the internal and interenterprise. The model uses attribute-based encryption to protect internal data access and data sharing between enterprises. Thwin and Vasupongayya [25] designed a medical data sharing query scheme based on proxy reencryption. The medical ciphertext data were uploaded to the chain through the gateway server after reencryption. The cloud storage service saves medical data, and the private chain stores medical metadata and raw data access logs to improve the confidentiality and security of personal medical data. Truong et al. [26] adopted prefix encryption technology for hierarchical structure scenarios to improve the security of data queries between IoT devices. In this scheme, the data requester first uses the ciphertext data encrypted by the prefix to apply to the data provider. After obtaining the consent of the data provider, the requester obtains the key sent by the data provider from the key authority and decrypts the ciphertext.

The above research all use cryptographic mechanisms to control access rights and protect the query of blockchain data. But due to the characteristics of the blockchain, it is determined that the blockchain data need to be shared and inquired more. This requires the data owner to know the specific inquirers in advance to implement effective authority control. Therefore, the solutions in the above research cannot quickly respond to the inconsistency between the actual inquirers and the preset inquirers. So, they are not flexible enough.

The query method based on threshold secret sharing can determine the approximate query permission range in advance. So, it is suitable for sharing query scenarios and is more flexible. At the same time, only when the number of people applying for reconstruction exceeds the preset threshold, the data can be successfully queried, which limits the rights of the inquirer and improves security. In the medical privacy data sharing scheme based on the consortium chain [27], medical institutions distribute key shares through the Shamir secret sharing mechanism to medical users who propose to share patient medical privacy data. The system then restores the key through enough key shares uploaded by medical users. Finally, medical users can query the private data. Xu et al. [28] proposed a verifiable secret sharing and multi-party coordination mechanism that can reduce the complexity and cost of interaction in a blockchain-based cyber-physical system. The user node sends an application to the witness node responsible for verifying the $1/n$ core shares distributed by the sensor node. Only if it obtains valid responses from more than n witness nodes, it can obtain permission to view the data. Sohrabi et al. [29] introduced a master node responsible for storing user key shares distributed by smart contracts. After the nodes that meet the query conditions successfully obtain the ciphertext data, they need to continue to send the key share request. The user can view the data only after receiving a certain valid key share response. Literature [28, 29] distributed the key to multiple places through a secret sharing mechanism, but only one data query request can be served. When the number of people applying for query is n, the system needs to transmit the key share set n times. So, these methods have a large amount of transmission. Table 1 summarizes the features of existing blockchain data query schemes.

Therefore, this study proposes a blockchain data sharing query scheme based on the Blakley threshold secret sharing mechanism. This scheme adopts the mode of multiple key shares and multiple query users to reduce transmission costs. It improves the security of data queries while improving the fault tolerance of the system through the erasure code mechanism.

3. Preliminary Knowledge

3.1. Threshold Secret Sharing. Given positive integers t and $n(n \geq t)$, the secret information S is divided into n subsecrets S_i and they are distributed to n different people, where $1 \leq i \leq n$. These people are respectively denoted as $\{p_1, p_2, \ldots, p_n\}$. When the number of subsecrets is greater than or equal to t, the secret S can be reconstructed; conversely, when the number of subsecrets is less than t, S cannot be reconstructed. At this time, the positive integer t becomes the threshold value, and the above process is called (t, n) threshold secret sharing.

Shamir proposed a threshold scheme based on algebraic Lagrange interpolation polynomials in 1979, namely, Shamir secret sharing [30]. First, the prime number q is determined, the finite field F_q is selected, and the secret information S is made to satisfy $S \in F_q$. Then, different nonzero elements $x_1, x_2, \ldots, x_n$ are chosen on F_q. Then, $t-1$ elements $a_1, a_2, \ldots, a_{t-1}$ are chosen on F_q, and a polynomial $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots a_{t-1} x^{t-1}$ is constructed, where a_0 is the secret S. $f(x_i)$ is obtained through the polynomial, where $1 \leq i \leq n$. Then, $(x_i, f(x_i))$ is sent to the corresponding p_i. When there are t people trying to reconstruct the secret S, they need to solve the following equation:

$$f(x) = \sum_{i=1}^{t} f(x_i) \prod_{1 \leq j \leq i, j \neq i} \frac{x - x_j}{x_i - x_j} \bmod q. \tag{1}$$

Then, $S = f(0)$ can be calculated.

Blakley constructed another secret sharing scheme through spatial geometry [31]. It regards the secret S as a point in the t-dimensional space and divides S into n $(t-1)$-dimensional linearly independent spaces. Then, any t linearly independent $t-1$ dimensional spaces can determine a unique point, which is the secret S. Therefore, only $t \times n + n$ elements need to be selected in the finite field F_q to form a t-element linear equation set:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1t}x_t = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2t}x_t = b_2, \\ \cdots, \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nt}x_t = b_n. \end{cases} \tag{2}$$

Table 1: Feature comparison of blockchain data query schemes.

	Type of blockchain	Application scenarios	Data query technology adopted	Advantages	Disadvantages
Literature [19]	Consortium blockchain, public blockchain	Digital archive management	Smart contract	Queries are automatically executed when user requests meet preset rules and do not rely on nonblockchain systems.	Each node needs to run the deployed contract, and the resource consumption of blockchain applications is large.
Literature [20]	Consortium blockchain	Medical data sharing	Smart contract		
Literature [21]	Consortium blockchain	Medical privacy data sharing	Smart contract		
Literature [22]	Public blockchain	Paid query of data	Smart contract		
Literature [23]	Public blockchain	Data security sharing	Attribute-based encryption	Easy to add an access permission control mechanism, which makes access permission control more fine-grained.	Inability to quickly respond to situations where the actual inquirers do not match the preset inquirers.
Literature [24]	Private blockchain, consortium blockchain	Enterprise data sharing	Attribute-based encryption		
Literature [25]	Private blockchain	Medical data sharing	Proxy reencryption		
Literature [26]	Consortium blockchain	IoT device data query	Prefix encryption		
Literature [27]	Consortium blockchain	Medical privacy data sharing	Secret sharing	Determine the general query permission scope in advance, which is suitable for shared query scenarios. Data can be queried only when the number of users applying for reconstruction exceeds the preset threshold.	There is no data reconstruction mechanism. Unable to recover data when there is a failed node. Only one data query request can be served. When the number of inquiries is large, the system transmission volume is large.
Literature [28]	Not specified	Cyber-physical system	Secret sharing		
Literature [29]	Public blockchain	Cloud data sharing	Secret sharing		

The above equations can be denoted as $AX = B$. In the coefficient matrix A, any t rows are required to be linearly independent, and the secret S must be the only solution of the equation set. Therefore, when there are t people trying to recover the secret S, they only need to solve the equation set containing these t linearly independent lines to solve the secret S.

3.2. Erasure Codes. Erasure codes [32] originated in the field of communications. They are used to correct errors in data transmission. In the field of data storage, the original data D can be divided into t data blocks. Then, they are coded with erasure codes, to obtain a total of n data blocks and redundant blocks. Any t redundant blocks can recover the original data D according to the nature of erasure codes.

Reed-Solomon code [33] (Reed-Solomon, referred to as RS code) is a horizontal encoding that performs polynomial operations on the elements of the Galois field $\mathrm{GF}(2^w)$. Common RS coding matrices are Vandermonde matrix and Cauchy matrix. The construction of Vandermonde matrix [33] is relatively simple. In Galois field $\mathrm{GF}(2^w)$, Vandermonde matrix undergoes elementary transformation, and the first t rows are transformed into the identity matrix E. However, although the addition on the Galois field $\mathrm{GF}(2^w)$ can be converted into an exclusive OR, the multiplication needs to rely on the discrete logarithm operation. Therefore, the RS code based on the Vandermonde matrix has a large amount of calculation. The Cauchy matrix [34] can convert the multiplication in the Galois field $\mathrm{GF}(2^w)$ to the $\mathrm{GF}(2)$ binary operation and turn the matrix multiplication into a simple binary XOR operation.

4. System Scheme

4.1. System Model. The system model of blockchain data sharing query scheme based on threshold secret sharing mainly involves three entities, namely, data user, shared user, and blockchain storage system.

Data User. The owner of the data to be uploaded. The data will be uploaded to the blockchain storage system after being encrypted and segmented.

Shared User. In order to ensure that the data uploaded by the user who owns data are complete and correct, a certain number of shared users are required to work together to share data.

Blockchain Storage System. A storage system that stores data uploaded by data users. If a data query request is received, the blockchain storage system will return the corresponding data result.

The system model of this scheme is shown in Figure 1. First, data users will need to upload data to the blockchain storage system for encryption operations. Next, data segmentation and erasure code encoding are performed on the encrypted data. Then, the data encoded by the erasure code are uploaded to the blockchain storage system to complete the storage of the data.

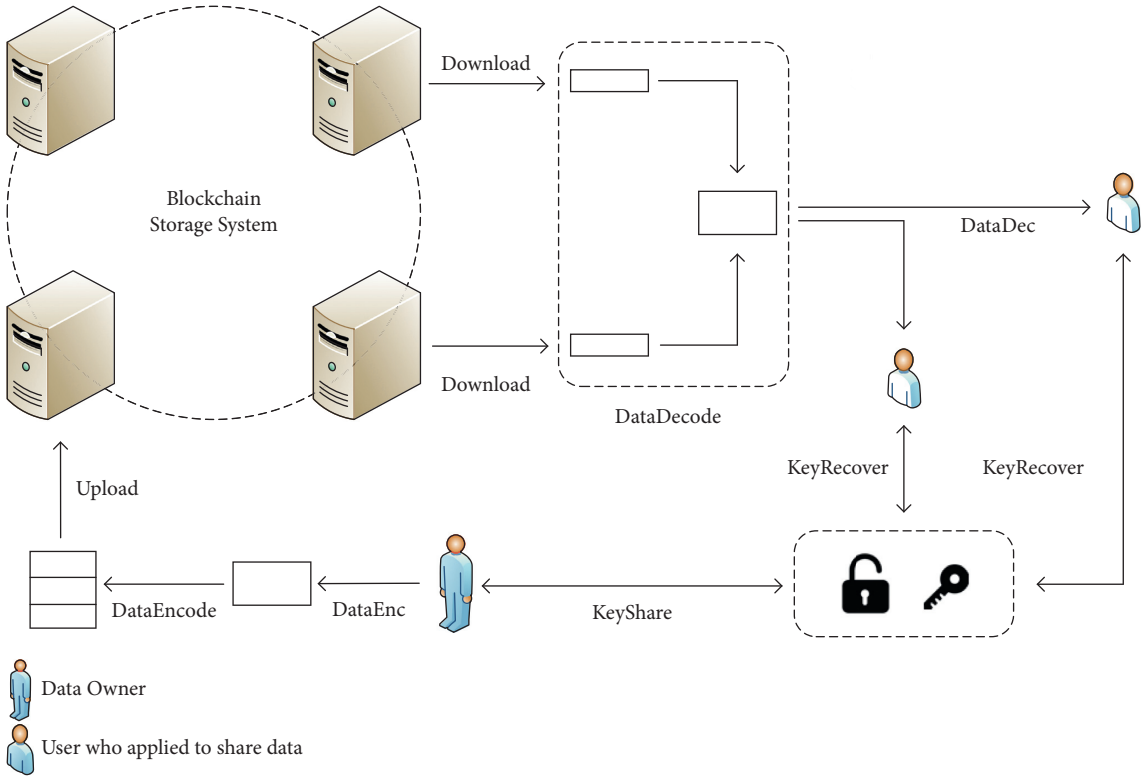

Figure 1: System model.

In order to ensure the consistency of data, shared users need to cooperate with each other to obtain and share data. That is, shared users need to apply to data users first. After the data user receives the application, the key is divided into multiple key shares through the Blakley secret sharing technology and then distributed to the shared user who requests. Shared users obtain the minimum required number of data coding blocks from the blockchain storage system, and then, the shared users work together to complete the reconstruction of the coding block. The key is recovered based on the multiple key shares held by shared users. Then, they complete the decryption of the data and obtain the original data.

The formal definition of this scheme is as follows:

(1) KeyGen(λ). Enter the security parameter λ, the algorithm output key K, and the key K is used to encrypt the original data;

(2) DataEnc(K, D): Input the key K, the original data D, and the algorithm outputs the encrypted ciphertext ED;

(3) MatrixGen(ID, n, t): Enter the data user ID, and the algorithm uses different data user ID to determine different matrix factors. Then, generate different $n \times t$-order nonsingular matrices M;

(4) DataEncode(M, ED): Input matrix M, ciphertext data ED, and output code block $C = \cup_{i=0}^{n-1} C_i$ after encoding;

(5) KeyShare(K, M, n, t): Input the key K, matrix M, the number of shared users n, the threshold value t, and output the corresponding key share equation x;

(6) KeyRecover(x): Input the key share equation x and output the key K;

(7) DataDecode(M, C): Input coding matrix M, ciphertext coding data C, and output ciphertext data ED;

(8) DataDec(K, ED): Enter the key K, the ciphertext data ED, and decrypt the original data D.

The security requirements of this solution are as follows:

(1) When the shared users do not provide the geometric space equations that meet the preset number requirements, then they cannot recover the key correctly.

(2) When the shared user obtains the spatial geometric equation used to recover the secret, the equation can be used multiple times to recover the secret.

4.2. Scheme Construction

4.2.1. Data Upload and Storage Stage. First, the data user executes the KeyGen algorithm to generate the key K that will be used later. Then, using the DataEnc algorithm, the original data D to be uploaded to the blockchain storage system is encrypted with the key K generated in the previous step. Then, data user uses the MatrixGen algorithm to generate a nonsingular matrix according to ID. The matrix will be used as an erasure code encoding matrix to encode the encrypted ciphertext data and the Blakley secret sharing

space geometric equation coefficient matrix. The data user uses the erasure code encoding matrix M as the input of the algorithm, DataEncode to encode and slice the ciphertext data, and ED to obtain the slice $\cup_{i=0}^{n-1} C_i$. Finally, the encoded block is sent to the blockchain storage system. The blockchain system stores the encoded ciphertext data C. The implementation of the data upload algorithm is shown in Figure 2.

4.2.2. Key Distribution Stage. Data users need to distribute keys to shared users. Both the Cauchy matrix and the Vandermonde matrix are nonsingular matrices. But the Cauchy matrix has a smaller computational time complexity than the Vandermonde matrix. The data user executes the MatrixGen algorithm to generate the Cauchy matrix M, and the matrix M is used as the coefficient matrix of the t-dimensional space geometric equation. The coefficient matrix M and a certain point K are used in the t-dimensional space to determine the unique t-dimensional space plane equation. Assuming that the current secret is shared with n users in total, then t of the n shared users can recover the secret K. The implementation of the shared user secret sharing algorithm is shown in Figure 3. I_i is the coefficient of the $t-1$-dimensional space equation, that is, the subsecret.

4.2.3. Key Recovery and Data Access Stage. In the key recovery and data access stage, the shared user first issues an access application and obtains the encoded ciphertext data from the blockchain storage system. However, the data are encrypted and encoded. So, the details of the data cannot be viewed directly. At this point, it is a binary string without any semantics for the viewer.

When many shared users trying to recover the ciphertext-encoded data C, they need to cooperate to recover the t-dimensional space plane equation set $(AK - B = 0)$ through each user's own $t-1$-dimensional space plane equation. When the number of shared users reaches the number t presented by the data uploading user, the rank of the coefficient matrix of the aforementioned spatial plane equation system is equal to t, and the coefficient matrix can be reversed. This scheme uses the same matrix as the spatial plane equation coefficient matrix and the erasure code encoding matrix. At this time, the shared users can cooperate with each other to decode the ciphertext-coded data C through the inverse matrix. So, they obtain the ciphertext data ED. Then, the space plane equation of rank t is used to calculate the certain point K in the t-dimensional space, and this point is the key. Then, the original data D can be decrypted by the key. The shared user decoding and decryption algorithms are shown in Figure 4.

5. Experiment and Analysis

Based on Hyperledger Fabric1.0 and the erasure code processing library reedsolomon of Golang language, this section implements a blockchain-based threshold secret sharing query scheme on a server with Intel(R) Xeon(R) Platinum 8163 CPU @ 2.50 GHz CPU and 32G memory.

Algorithm 1 Algorithm for Data Upload and Storage

Input: λ, D, n, t, ID
Output: *null*
1: function UploadAndStoreData (λ, D, n, t, ID)
2: $K \leftarrow KeyGen\ (\lambda)$
3: $ED \leftarrow DataEnc\ (K, D)$
4: split data ED into $ED_0, ED_1, \ldots, ED_{t-1}$ linearly
5: $M \leftarrow MatrixGen\ (ID, n, t)$, let $a_{i,j}$ as the element of Matrix M
6: for $i = 0 \rightarrow n - 1$ do
7: init $C_i = 0$
8: for $j = 0 \rightarrow t - 1$ do
9: $C_i \mathrel{+}= a_{i,j} \times ED_j$
10: end for
11: end for
12: let C as $\{C_0, C_1, \ldots, C_{n-1}\}$
13: send C to Blockchain Storage System
14: *StoreDataInBlockStorageSystem* (C)
15: end function

Figure 2: Data upload and storage process.

Algorithm 2 Algorithm for Key Distribution

Input: ID, n, t, K
Output: $I_0, I_1, \ldots, I_{n-1}$
1: function DistributeKey (ID, n, t, K)
2: $M \leftarrow MatrixGen\ (ID, n, t)$, let a_{ij} as the element of Matrix M
3: let $K = \{k_0, k_1, \ldots, k_{t-1}\}$
4: for $i = 0 \rightarrow (n - 1)$ do
5: init $B_i = 0$, $I_i = [\ ]$
6: for $j = 0 \rightarrow (t - 1)$ do
7: $B_i \mathrel{+}= a_{i,j} \times k_j$
8: I_i .append $(a_{i,j})$
9: end for
10: I_i .append (B_i)
11: end for
12: return $I_0, I_1, \ldots, I_{n-1}$
13: end function

Figure 3: Key distribution process.

5.1. Correctness Analysis. This section will analyze the correctness of shared users' key recovery in this scheme. The data user that needs to upload to the blockchain storage system is assumed D and the private key is assumed K. Then, the plaintext is encrypted to get $\mathrm{ED} = \mathrm{DataEnc}(K, D)$. According to the user's globally unique ID, the corresponding matrix factor is generated, and then, the $n \times t$-th order nonsingular matrix M is generated: $M = \mathrm{MatrixGen}(\mathrm{ID}, n, t)$. The $\mathrm{DataEncode}(E, \mathrm{MD})$ algorithm is executed on the ciphertext data ED, and the $C = \cup_{i=0}^{n-1} C_i$ is uploaded to the blockchain storage system. When a sharing user applies for secret sharing, the data user constructs n $t-1$-dimensional space plane equations through the matrix M and the key K. The sharer can determine the unique intersection point K through any t $t-1$-dimensional space planes. There are two possibilities for the number of shared users n_s and t who applied for viewing:

Algorithm 3 Algorithm for Key Recovery and Data Access

Input: $I_0, I_1, \ldots, I_{k-1}$
Output: D
function RecoverKeyAndAccessData ($I_0, I_1, \ldots, I_{k-1}$)
  download the encoded ciphertext C from Blockchain Storage System
  init $A = [\,]$, $B = [\,]$;
  for $i = 0 \rightarrow (k-1)$ do
    for $j = 0 \rightarrow (t-1)$ do
      $a_{i,j} = I_{i,j}$
    end for
    $b_i = I_{i,t}$
  end for
  $r = \text{rank}(A)$
  if $r \geq t$ then
    let $p_{i,j}$ as the element of Matrix A^{-1};
    for $i = 0 \rightarrow (t-1)$ do
      for $j = 0 \rightarrow (n-1)$ do
        $ED_i = p_{i,j} \times C_j$
        $K_i = p_{i,j} \times B_j$
      end for
    end for
    $D = DataDec(K, ED)$
    return D
  end if
  return *null*
end function

Figure 4: Key recovery and data access process.

(1) $t \leq n_s \leq n$. At this time, the number of users applying for shared secrets n_s is greater than the minimum threshold t and less than the initially set number of secret shares n. Although at this time $n_s \leq n$, but in the coefficient matrix, any t rows in n rows are linearly independent. From the knowledge of linear algebra, there are t rows in any n_s rows of the coefficient matrix that are linearly independent. That is, among the n_s shared users, any t shared users can use their own space plane equations to determine the unique intersection point K.

(2) $n_s > n$. If n_s shared users $(n_s > n)$ share n $t-1$-dimensional space plane equations (a $t-1$-dimensional space plane equation will be held by multiple shared users), there is no guarantee that the shared users will hold exactly t linearly independent spatial plane equations when trying to recover the key. Therefore, all shared users are divided into $k = [n_s/n]$ groups. The data distribute n space plane equation coefficient matrices to each group of shared users. There are n shared users in each group. Only t linearly independent $t-1$-dimensional space plane equations can recover the secret. At this time, there are still $n_s - k \cdot n$ shared users who have not received the shared secret share. The data user first sends the $n_s - k \cdot n$ spatial plane equation coefficient matrices to $n_s - k \cdot n$ shared users. If $n_s - k \cdot n \geq t$, $n_s - k \cdot n$ shared users can directly recover the keys; if $n_s - k \cdot n < t$, then the data user can hold the remaining $t - (n_s - k \cdot n)$ space plane equations as a virtual shared user.

5.2. Security Analysis. In this scheme, the shared users restore the original data through erasure code decoding and Blakley secret sharing. For Blakley secret sharing, only when the number of shared users with the correct secret share exceeds the minimum threshold t, the unique solution of the full-rank t-ary linear equation system can be solved, and any number of users less than t cannot get the unique intersection point correctly. For the $n \times t$-order erasure code matrix, when $(n-t)$ pieces of data are lost or tampered with, the original data can still be recovered through the erasure code decoding mechanism. When the number of lost or tampered data fragments exceeds $(n-t+1)$, it cannot be recovered. This scheme is a blockchain application for data recording and management scenarios. The open, transparent, and nontamperable characteristics of the blockchain make this probability almost nonexistent. This can only happen when the adversary has more than 51% of the computing power, but the cost of this is far greater than the value of the information itself.

5.3. Efficiency Analysis

5.3.1. Successful Reconstruction Rate. This scheme uses erasure code as the data slicing method. When shared users need to reconstruct data, the corresponding code package can be obtained from the blockchain storage system. The current blockchain storage system is assumed to have a total of p nodes. The storage system does not adopt a full copy redundancy scheme due to the limitation of node storage capacity, and each node only stores part of the data locally. When the data user uploads the data for storage, the erasure coding matrix is an $n \times t$-order Cauchy matrix; that is, the ciphertext data ED are first divided into t data slices and then encoded and converted into n coded slices. The average storage capacity of each node in the current system is assumed to be only $q\,(q \leq n)$ and the performance of each node is evenly distributed in the system. Then, the probability that the data can be directly restored through the local storage of the blockchain storage system node is as follows:

$$P_1 = 1 - \frac{C_n^{t-1} F(t-1, p, q) + C_n^{t-2} F(t-2, p, q) + \cdots + C_n^q F(q, p, q)}{(C_n^q)^p}. \tag{3}$$

Among them, $F(\text{num}, p, q)$ indicates that the current p nodes (the storage limit of each node is q) can completely store the number of combinations of num different blocks.

If the data uploaded by the data user are sliced and uploaded directly without encoding, then the probability that the data can be directly recovered from the blockchain storage system is as follows:

$$P_2 = 1 - \frac{C_t^{t-1} F(t-1, p, q) + C_t^{t-2} F(t-2, p, q) + \cdots + C_t^q F(q, p, q)}{\left(C_t^q\right)^p}. \tag{4}$$

When $n \geq t$, $C_n^x > C_t^x$, $0 < x < t$, $P_1 > P_2$ can be obtained. Therefore, the appropriate erasure code matrix size and node parameters can provide a high error tolerance rate.

5.3.2. Processing Rate. In this section, we will analyze the correctness of the shared user recovery key in this scheme. The three common key lengths are chosen as follows: 128 bit, 256 bit, and 512 bit. The current secret threshold is assumed to be $t = 10$, and the number of the shared people is set to 11 to 25. The simulation results are shown in Figure 5. When the secret length is 128 bit and the number of shared users is 25 (that is, any 10 users out of 25 users can recover the secret), the time spent on key distribution and recovery is 9.93 ms and 2.53 ms, respectively. When the secret length is 256 bit and 512 bit, the corresponding key distribution time is 10 ms and 20.73 ms, respectively. When the length of the secret is longer and the number of people sharing the secret is larger, the distribution time of the secret is also longer, but it remains at the millisecond level.

Before the data owner uploads the data and before the data inquirer views the data, the data need to be segmented and reconstructed. Therefore, the performance of the erasure code-based data sharing mechanism in our solution is related to the execution performance of the blockchain application. The original data size is set to 100 M, $t = 10$ (that is, the original ciphertext data is divided into 10 parts). Now, the effect of the number of different codes n on the encoding and decoding processing time is tested. The number of code blocks is set to 11–20. When $n = 11$, the processing time was the shortest, encoding and decoding took 225.8 ms and 156.6 ms, respectively, and the encoding and decoding rates were 442.86 MB/s and 638.57 MB/s, respectively. As shown in Figure 6, the larger the number of encoding blocks, the longer the encoding and decoding time. As the number of redundant data fragments increases (that is, the matrix size continues to increase), the encoding and decoding rate gradually decreases, as shown in Figure 7.

Therefore, our solution uses the Blakley space plane equation coefficient matrix to perform erasure-coded data fragmentation processing, which reduces the data size, and this does not significantly affect the speed and performance of blockchain applications in data recording and management scenarios.

5.4. Adaption Analysis. This section will analyze the adaptability between our scheme and the blockchain system. As shown in Figure 8, the relationship between our scheme and blockchain architecture [8] mainly includes the data layer and network layer. In our scheme, coding algorithm, data segmentation, and reconstruction modules are added to the data layer. At the same time, the data divided into t slices will generate n slices after erasure coding. There are redundant $(n - t)$ data fragments. This mechanism can increase the storage cost of a piece of data by (n/t) times.

At the network layer, the scheme adds computing and network resources of data users, shared users, and nodes of blockchain storage system. This improves the correct reconstruction rate of data to be stored in the system and the security of data sharing. It is assumed that the ciphertext data ED is encoded by the $n \times t$-order erasure matrix and uploaded to the blockchain storage system. That is, the ciphertext data are segmented into t slices, and the amount of data in each slice is d. Then, the amount of data transferred is shown in the following formula:

$$O_t = k \cdot d, \quad t \leq k \leq n. \tag{5}$$

As shown in the blue area in Figure 9, the amount of data O_t to be transmitted is between $n\ d$ and $t\ d$, and the shared user can successfully recover the original data. At the same time, the number of verification times is linearly related to the number of fragments when the data fragments are cut and encoded.

According to the above content, the improvements made in this scheme increase the cost of the transmission and verification mechanism at the network layer, but do not change their internal operation mechanism. Similarly, if the block structure is not modified, the erasure code mechanism introduced will only increase the block storage cost without affecting the existing block storage structure. According to the analysis in Section 5.3, our scheme can provide a high reconstruction success rate without affecting the speed and performance of blockchain applications in data recording and management scenarios. Changes in the data layer and network layer did not break the original mechanism but only increased the cost. Therefore, our solution does not affect the decentralized, sequential data, collective maintenance, programmability, and security and trusts features of blockchain. Therefore, our scheme is completely suitable for the blockchain storage system.

5.5. Comparison. As summarized in Table 2, we compare and analyze our scheme and the other two blockchain data query schemes. The query mode of literature [28] and

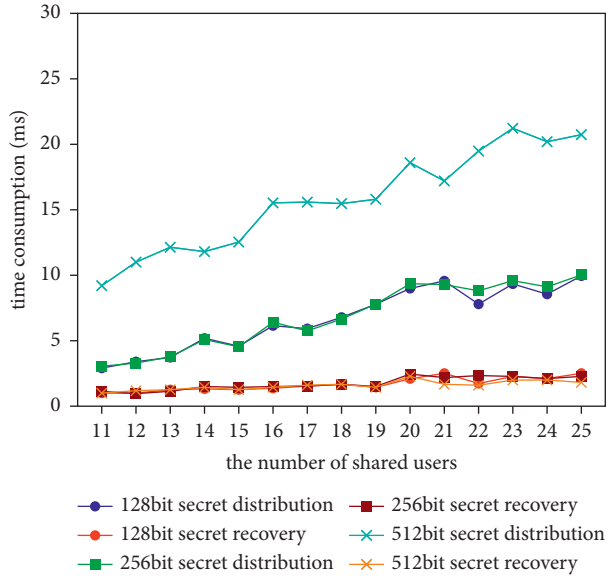

Figure 5: Secret distribution and recovery time.

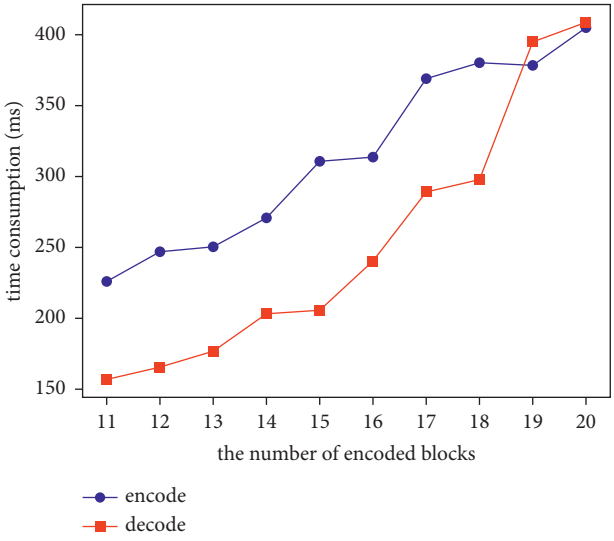

Figure 6: Erasure code encoding and decoding time.

literature [29] is a multi-share, single-user mode. That is, the key shares are transferred to n nodes or objects. The user applies for and obtains more than t valid key shares to recover the data, and our scheme adopts a multi-share, multi-user mode, which directly distributes the key shares to the data inquirers who have passed the review of the data applicant. As long as the number of data inquirers applied for viewing exceeds t or the data owner agrees, the key can be reconstructed. Therefore, the single transmission volume of our scheme is $O(t)$ level, which is higher than that $O(1)$ level of the multi-share, single-user query mode. But the total transmission volume of our scheme is only $n \times (t + 1)$, lower than the total transmission volume $n \times 2t$ of the multi-share single-user query mode. In addition, our scheme uses the coefficient matrix of the Blakley space plane equations to segment the ciphertext data and reduce the data size to facilitate the chaining. At the same time, it increases the power of data reconstruction during query. However, the solution using Lagrange interpolation can only achieve data segmentation through additional mechanisms.

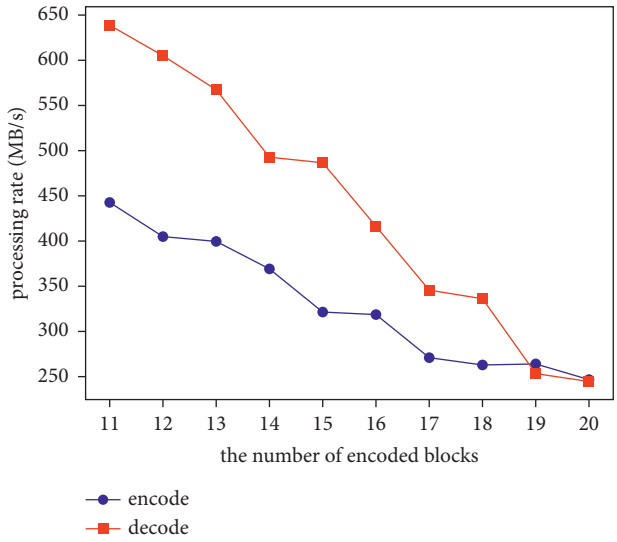

Figure 7: Erasure code encoding and decoding rate.

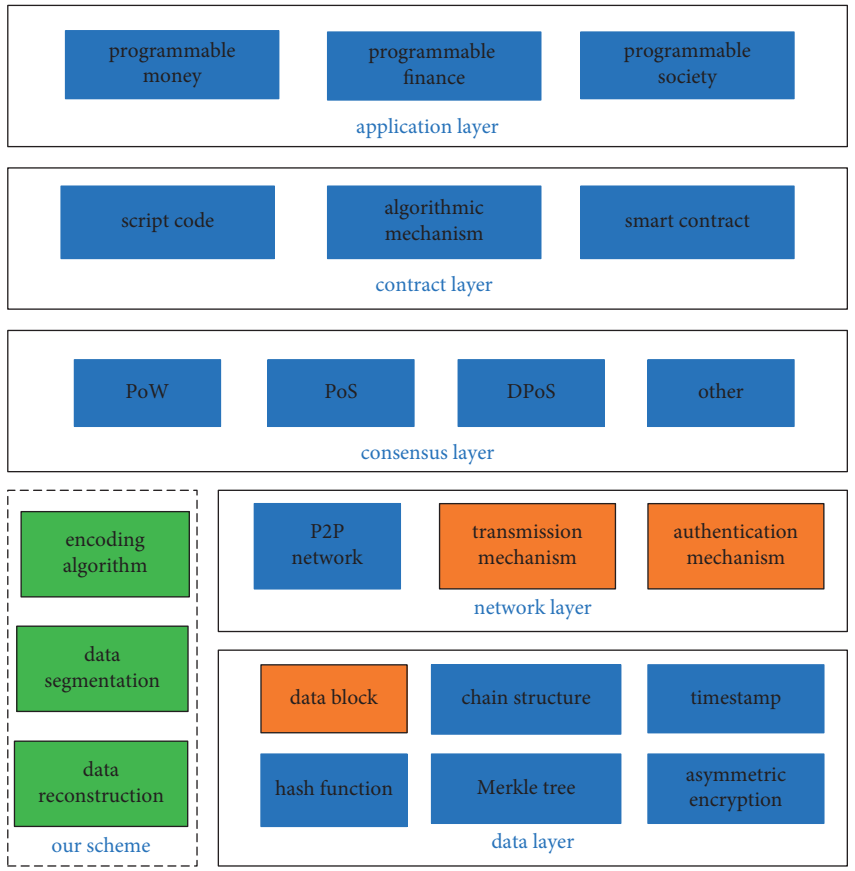

Figure 8: The relationship between this scheme and the blockchain architecture.

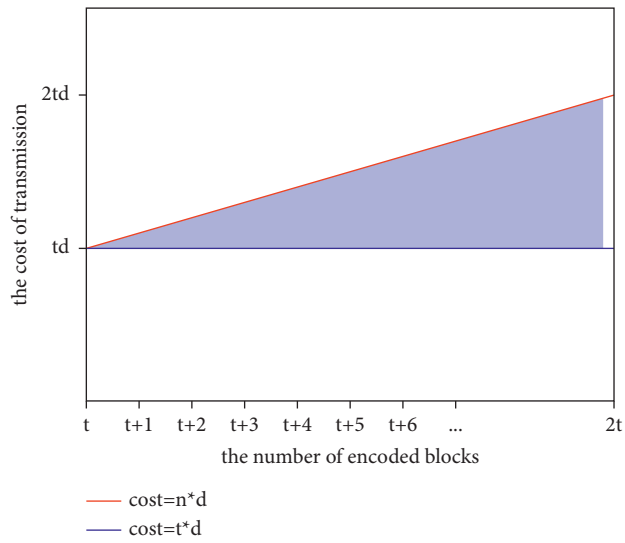

Figure 9: Network layer data transmission costs.

Table 2: Comparison of schemes.

	Our scheme	Literature [28]	Literature [29]
Query mode	Multi-share, multi-user	Multi-share, single-user	Multi-share, single-user
Single transmission volume	$O(t)$	$O(1)$	$O(1)$
Total transmission volume	$n \times (t+1)$	$n \times 2t$	$n \times 2t$
Data reconstruction	RS code	Not supported	Not supported
Depends on a specific platform	No	No	Ethereum
Applicable scene	Data recording and management	Cyber-physical system	Cloud data protection

6. Conclusion

In view of the blockchain data query requirements in the data recording and management scenario, this study analyzes the phenomenon that the data on the chain are too large to be directly stored in the block and proposes a blockchain data sharing query scheme based on threshold secret sharing. The sharing query scheme uses the Blakley space plane equation to share the secret elements used for data query. It restricts the rights of blockchain data inquirers, thus improving the security of blockchain data queries. At the same time, a method for uploading data to the blockchain based on erasure codes is proposed. It uses the Blakley space plane equation coefficient matrix as the erasure code encoding matrix to segment the ciphertext data. It not only reduces the data size but also improves the system error tolerance rate during query. The simulation experiment results show that the additional time consumption of the secret-sharing recovery and data segmentation reconstruction of this solution is much less than the cost of block generation. Therefore, it will not affect the normal operation of the blockchain application in the scene of data recording and management, and it can also improve the security of blockchain application data.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Nature Science Foundation of China (nos. 61972208 and 61672299), Postgraduate Research &Practice Innovation Program of Jiangsu Province(SJKY19_0770).

References

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

[2] Z. Yu, X. G. Liu, and G. Wang, "A survey of consensus and incentive mechanism in blockchain derived from P2P," in *Proceedings of the . 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 1010–1015, Singapore, December 2018.

[3] L. Chen, F. Xiang, and Z. X. Sun, "A survey of blockchain security technologies based on attribute-based cryptography," *Acta Electronica Sinica*, vol. 49, no. 1, pp. 192–200, 2021.

[4] G. Ma, C. Ge, and L. Zhou, "Achieving reliable timestamp in the bitcoin platform," *Peer-to-Peer Networking and Applications*, vol. 13, no. 6, pp. 2251–2259, 2020.

[5] W. Wang, D. T. Hoang, P. Hu, Z. Xiong, and D. Niyato, "A survey on consensus mechanisms and mining strategy management in blockchain networks," *IEEE Access*, vol. 7, Article ID 22370, 2019.

[6] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," 2015, http://gavwood.com/Paper.pdf.

[7] E. Androulaki, A. Barger, V. Bortnikov et al., "Hyperledger fabric: a distributed operating system for permissioned blockchains," in *Proceedings of the The Thirteenth EuroSys Conference*, pp. 1–15, Porto, Portugal, April 2018.

[8] Y. Yuan and F. Y. Wang, "Blockchain: the state of the art and future trends," *Acta Automatica Sinica*, vol. 42, no. 4, pp. 481–494, 2016.

[9] S. Q. Zeng, R. Huo, T. Huang, J. Liu, S. Wang, and W. Feng, "Survey of blockchain: principle, progress and application," *Journal on Communications*, vol. 41, no. 1, pp. 134–151, 2020.

[10] G. Yu, T. Z. Nie, X. H. Li, Y. F. Zhang, D. R. Shen, and Y. B. Bao, "The challenge and prospect of distributed data management techniques in blockchain systems," *Chinese Journal of Computers*, vol. 42, pp. 1–27, 2019.

[11] A. D. Liu, X. H. Du, N. Wang, and S. Z. Li, "Research progress of blockchain technology and its application in information security," *Ruan Jian Xue Bao/Journal of Software*, vol. 29, no. 7, pp. 2092–2115, 2018.

[12] H. Li, D. Han, and M. Tang, "Logisticschain: a blockchain-based secure storage scheme for logistics data," *Mobile Information Systems*, vol. 2021, Article ID 8840399, 15 pages, 2021.

[13] J. Tian, X. Jing, and R. Guo, "Public audit scheme of shared data based on blockchain," *Communications in Computer and Information Science*, vol. 1105, pp. 327–344, 2019.

[14] L. Wang, W. Liu, and X. Han, "Blockchain-based government information resource sharing," in *Proceedings of the 2017 IEEE 23rd International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 804–809, Shenzhen, China, December 2017.

[15] X. Zhang and Y. Yin, "Research on digital copyright management system based on blockchain technology," in *Proceedings of the 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pp. 2093–2097, Chengdu, China, June 2019.

[16] L. Ouyang, S. Wang, Y. Yuan, X. Ni, F. Y. Wang, and X. Han, "Smart contracts: archit and research progresses," *Acta Automatica Sinica*, vol. 45, no. 3, pp. 445–457, 2019.

[17] C. Zhang, Q. Li, Z. H. Chen, Z. R. Li, and Z. Zhang, "Medical chain: alliance medical blockchain system," *Acta Automatica Sinica*, vol. 45, no. 8, pp. 1495–1510, 2019.

[18] S. Wu and J. Du, "Electronic medical record security sharing model based on blockchain," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, pp. 13–17, Kuala Lumpur, Malaysia, January 2019.

[19] H. B. Tan, T. Zhou, H. Zhao et al., "Archival data protection and sharing method based on blockchain," *Ruan Jian Xue Bao/Journal of Software*, vol. 30, no. 9, pp. 2620–2635, 2019.

[20] V. Jaiman and V. Urovi, "A consent model for blockchain-based health data sharing platforms," *IEEE Access*, vol. 8, pp. 143734–143745, 2020.

[21] J. R. Wang, S. Z. Yu, and R. Li, "Medical blockchain of privacy data sharing model based on ring signature," *Journal of University of Electronic Science and Technology of China*, vol. 48, no. 6, pp. 886–892, 2019.

[22] A. K. Shrestha and J. Vassileva, "User data sharing frameworks: a blockchain-based incentive solution," in *Proceedings of the 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 0360–0366, Vancouver, Canada, October 2019.

[23] S. Huang, L. W. Chen, and B. B. Fan, "Data security sharing method based on CP-ABE and blockchain," *Computer Systems & Applications*, vol. 28, no. 11, pp. 79–86, 2019.

[24] X. L. Wang, X. Z. Jiang, and Y. Li, "Model for data access control and sharing based on blockchain," *Ruan Jian Xue Bao/Journal of Software*, vol. 30, no. 6, pp. 1661–1669, 2019.

[25] T. T. Thwin and S. Vasupongayya, "Blockchain based secret-data sharing model for personal health record system," in *Proceedings of the 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)*, pp. 196–201, Krabi, Thailand, Augest 2018.

[26] H. T. T. Truong, M. Almeida, G. Karame, and C. Soriente, "Towards secure and decentralized sharing of IoT data," in *Proceedings of the .2019 IEEE International Conference on Blockchain (Blockchain)*, pp. 176–183, Atlanta, USA, July 2019.

[27] M. J. Gao and H. Q. Wang, "Blockchain-based searchable medical data sharing scheme," *Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition)*, vol. 39, no. 6, pp. 94–103, 2019.

[28] Z. Xu, J. Yang, and J. Yin, "A lightweight data sharing mechanism and multiparty computation for CPS," in *Proceedings of the.2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pp. 1–5, Antwerp, Belgium, May 2020.

[29] N. Sohrabi, X. Yi, Z. Tari, and I. Khalil, "BACC: blockchain-based access control for cloud data," in *Proceedings of the Australasian Computer Science Week Multiconference*, pp. 1–10, Melbourne VIC Australia, February, 2020.

[30] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612-613, 1979.

[31] G. R. Blakley, "Safeguarding cryptographic keys," in *Proceedings of the 1973 Managing Requirements Knowledge, International Workshop on. IEEE Computer Society*, p. 313, New York, NY, USA, June 1979.

[32] S. Johnson, "Burst erasure correcting LDPC codes," *IEEE Transactions on Communications*, vol. 57, no. 3, pp. 641–652, 2009.

[33] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 300–304, 1960.

[34] R. M. Roth and A. Lempel, "On MDS codes via Cauchy matrices," *IEEE Transactions on Information Theory*, vol. 35, no. 6, pp. 1314–1319, 1989.