J. COMBIN. MATH. COMBIN. COMPUT. 124 (2025) 189–205

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Application of blockchain technology for data integrity and privacy protection in distributed networks

Yihui Deng[1], Sanxiang Xiao[2,✉]

[1] *Experimental Training Center, Guangzhou College of Applied Science and Technology, Guangzhou, Guangdong, 511300, China*
[2] *School of Computing, Guangzhou College of Applied Science and Technology, Guangzhou, Guangdong, 511300, China*

ABSTRACT

Blockchain technology has the characteristics of data anti-tampering and anti-forgery, which can provide solution ideas for the secure storage and transmission of data in distributed networks. The study applies blockchain technology to data auditing, constructs an aggregated signature based on conditional identity anonymization to protect user privacy, simplifies the auditing computation by using homomorphic hash function, and deploys three kinds of smart contracts on the blockchain to design a blockchain-based data integrity auditing scheme. For the privacy protection problem, a blockchain privacy protection model based on differential privacy is constructed by integrating the differential privacy policy into the blockchain smart contract layer. The experimental results show that the data integrity auditing scheme has superior blockchain storage cost and time overhead, and the average time overhead under different dynamic operations is below 30ms. The privacy protection model also exhibits high efficiency, with encryption and decryption times of 0.075s and 0.063s, respectively, under the largest data file, and a significant speed advantage in all phases of operation. The proposed scheme in this paper meets the needs of data integrity and privacy protection, and can provide efficient services for users.

Keywords: blockchain technology, differential privacy algorithm, privacy protection

1. Introduction

With the advent of the big data era, the growing volume of data has also brought new challenges to the secure storage and sharing of data. In the past, users were accustomed to storing their personal

✉ Corresponding author.
E-mail address: sanxiang_003@126.com (S. Xiao).

Received 19 July 2024; Accepted 18 December 2024; Published Online 16 March 2025.
DOI: 10.61091/jcmcc124-12

data in third-party data storage platforms, which were managed and maintained by centralized servers. However, the third-party platform is semi-trustworthy, users cannot control the platform's use of data stored in the platform, there is a possibility of data misuse, and the centralized platform is prone to data loss due to a single point of failure [7, 20, 27]. Therefore, the traditional centralized data management model lurks a huge risk of privacy leakage and is in urgent need of transformation.

In contrast, distributed network is a network environment composed of computer nodes distributed in different geographical locations [15]. Data is jointly maintained by multiple independent storage servers in the network, which not only avoids the single point of failure problem in the centralized storage system, reduces the risk of user data loss, and improves the reliability of the data management system [19, 14, 5]. In addition, nodes outside the network can also be configured to become a member of the distributed network with strong availability and scalability [3, 25]. The peer-to-peer network represented by blockchain network is a typical distributed network widely used in the field of user data security protection research.

Blockchain is a distributed database technology realized through a variety of technologies such as public key cryptography algorithms, hashing algorithms, consensus mechanisms, and distributed storage technologies [6, 18]. The distributed database of blockchain has more security advantages compared to the centralized system. In blockchain system, even if a specific node fails, the data is still guaranteed to be complete and not tampered with [22, 13]. The structure without third-party intermediaries also promotes data security and integrity as each transaction in a blockchain is based on a consensus established by the nodes of the entire blockchain network and one no longer needs to assess the trustworthiness of intermediaries or other participants in the network [10, 8, 1].

Many data in today's society have non-negligible commercial value, and the security of user data should not completely hope on the third-party application platform, a large number of scholars research on privacy data protection schemes. Literature [16] introduces the privacy security risk of cloud computing, on the basis of which it compares numerous privacy security protection techniques, including access control techniques, attribute-based encryption techniques, etc., and analyzes the characteristics and scope of application of typical schemes supported by technology. Literature [21] addresses the online social network (OSN) privacy protection problem, proposes to construct a security prediction model based on neural network, hybrid recursive genetic algorithm and radial basis function, and adopts attribute-based encryption scheme to encrypt the preprocessed OSN information, and further improves the security of the privacy data using particle swarm optimization algorithm. Literature [9] utilizes homomorphic encryption scheme to achieve secure data aggregation of ciphertexts in elected central nodes, which are generated in Device-to-device network environment by relying on the reliability-based central node election mechanism ordering. Literature [26] proposes a privacy-preserving scheme for social networks (PPSSN) based on categorical attribute encryption, which balances the privacy and security of data distribution by categorizing users and designing access control with different users and permissions, and also utilizes the buddy data caching mechanism to further reduce the decryption cost.

In the field of information security, blockchain technology's decentralization and other characteristics and can be very good in solving the crisis of trust in user privacy data security sharing problems. Literature [23] proposes blockchain-based edge computing technology, which realizes both the security protection and integrity checking of data in the cloud, as well as wider multi-party secure computing, while introducing the Paillier cryptosystem, which reduces the computational burden of the terminals under the premise of guaranteeing the operational efficiency. Literature [4] shows that the chained block structure of blockchain provides tamper-resistant data storage and sharing func-

tions and is based on a trusted consensus mechanism that enables it to verify the security of data, however, blockchain still has some privacy issues, and the anonymity and transaction privacy protection of blockchain in the existing cryptographic defense mechanisms are discussed. Literature [11] designs a federated blockchain privacy protection scheme (PDPChain) based on the improved Paillier homomorphic encryption mechanism, which encrypts and stores distributed private cluster data based on fine-grained access control of ciphertext policy attribute-based encryption, and is suitable for storing and sharing large amounts of data. Literature [2] designed a blockchain privacy protection model based on the DEPLEST algorithm, which maintains the local database storage and computational power within the limits of individual user's device, and ultimately protects the user's sensitive information through the distributed blockchain and passes the non-sensitive information to the main system to manage the size of the blockchain. Literature [17] proposes a two-stage privacy protection mechanism using the transparency of blockchain technology, firstly using double perturbation local difference privacy algorithm to perturb the location information of the worker, and secondly using edge cloud computing to upload the sensory data of the edge nodes to the blockchain to feed back to the requester, which achieves both the integrity of the sensory data and the privacy protection purpose. Literature [12] evaluates the role of blockchain-based distributed access control system in the user privacy protection problem, and proposes the concept of fog computing and federated chain, which effectively solves the single-point-of-failure problem of data storage by encrypting the data on the edge nodes while providing dynamic and fine-grained access control for the data to achieve privacy protection. Literature [24] emphasizes that the use of centralized access control mechanisms in the cloud can easily lead to tampering or leakage of sensitive data, so a blockchain access control framework AuthPrivacyChain is proposed, which not only blocks illegal access from hackers and administrators, but also protects authorized privacy.

The study is based on blockchain technology to design data integrity auditing method and privacy protection method. On the one hand, an aggregate signature algorithm is constructed by combining user anonymous identity and homomorphic hash function to realize conditional identity privacy protection, and the homomorphic feature is utilized to reduce the burden of auditing computation, and at the same time, three kinds of smart contracts are deployed, which are in charge of recording the metadata and executing the auditing tasks, to build the data integrity auditing method. Several comparison algorithms are also selected to analyze their storage costs and time overheads under different dynamic operations to explore the performance of conditional identity anonymous data auditing methods. On the other hand, the differential privacy mechanism is integrated into the smart contract layer of the blockchain network to realize the process of automatically invoking the chain code to add noise to the data during user uploading. The perturbation of the original data is realized by adding random Laplace noise to the numerical data and adding random response noise conforming to the definition of differential privacy to the binary data in non-numerical data for random flip. The encryption and decryption time analysis of different sizes of data and different complexity of access policies are carried out respectively, and comparative experiments of the algorithms under different stages are conducted to measure the privacy protection performance of the privacy protection method in this paper.

2. Application of blockchain technology in network information security

Blockchain technology, as an innovative distributed database technology, has an increasingly obvious application value in secure information storage and transmission. As blockchain has the unique

performance of decentralization, non-tampering and high security, it brings a brand-new solution and concept to the traditional information security problem.

First of all, blockchain technology can effectively solve the problem of data leakage, through the use of distributed ledger and encryption algorithms, so that blockchain technology to ensure that the stored data is not interfered with a single point of failure or malicious attacks. All data is strictly encrypted and only allowed to be accessed and modified under specific conditions. This greatly reduces the risk of illegal access or theft of data and improves information security.

Secondly, blockchain technology can enhance the security of data transmission. Traditional data transmission methods often face the risk of interception and tampering, etc. Blockchain technology ensures the integrity and authenticity of the transmitted data through encryption algorithms and consensus mechanisms. The entire transmission data is recorded in the blockchain and jointly verified and maintained by a number of nodes within the network. This ensures that the data is not subject to malicious modification and forgery, and enhances the reliability and security of information transmission.

In addition, blockchain technology has a long-term preservation and backup function, because the blockchain is distributed storage, so the data will be decentralized storage to a number of nodes, and in each node to save a complete copy of the data. This decentralized storage ensures that the data is durable, reliable, and not lost even when some nodes fail or are attacked. At the same time, the blockchain automatically backs up and restores data, reducing the risk of data loss.

3. Blockchain-based data integrity auditing program

With the widespread use of cloud computing technology and the explosive growth of data volumes, it has become particularly important to effectively safeguard the integrity and privacy of outsourced data. However, most of the existing data auditing methods rely on third-party auditors, a practice that not only increases the risk of data leakage and possible malicious behaviors of auditors, but also fails to meet the rising demand for data protection. To address the above challenges, this chapter proposes a blockchain-based conditional identity anonymization data auditing scheme.

3.1. System model

The system model of the blockchain-based data auditing scheme is shown in Figure 1, including four core entities: data owner (DO), key generation center (PKG), blockchain (BC), and cloud service provider (CSP). The specific four core entities are described as follows:

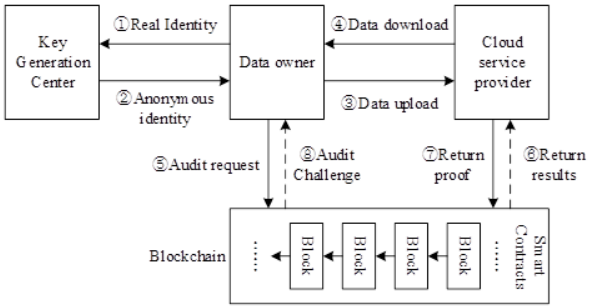


Fig. 1. Data audit system model based on block chain

DO: The user first generates the corresponding digital signature and integrity verification auxiliary

information for the outsourced data file. Subsequently, it is uploaded to the CPS and the integrity verification auxiliary information is uploaded to the blockchain via a secure channel. Finally, DO deletes the local copy to save storage space.

PKG: Responsible for generating anonymous identity and corresponding key pairs based on DO's real identity in the initialization phase.

CSP: responsible for storing the outsourced data and their digital signatures, and handling integrity audit challenges from the BC, returning response messages containing audit proofs to the BC.

BC: The blockchain is maintaining a shared ledger among the participants of the decentralized network. The BC automatically records data integrity information and performs periodic cloud data audits via smart contracts.

3.2. Adversary modeling

The blockchain-based conditional identity anonymization data integrity auditing scheme aims to achieve reliable auditing of data integrity and to protect the identity privacy of the data owner from possible misbehavior by malicious third-party auditors. Therefore, the following types of attacks are considered in the data auditing process: forgery attacks, privacy attacks and data recovery attacks.

3.3. Program construction

This program contains $Setup$, $Key\,Gen_{PID}$, Sig, $Challenge$, and $Verify$, five algorithms in total.

3.3.1. System initialization. Setup : The algorithm is executed by PKG by entering security parameters ξ into the system, generating two multiplicative cyclic groups of order prime p, G_1, G_2, setting up a bilinear mapping: $G_1 \times G_1 \to G_2$. Let g be the generating element of the group G_1, setting up two collision-proof hash functions, $H_1 : G_1 \times G_1 \times \{0,1\}^* \to \{0,1\}^\rho$, $H_2 : G_1 \times \{0,1\}^\rho \to Z_p^*$, where ρ denotes the length of the bits of the anonymous identity. And a homomorphic hash function, $H_3 : Z_p \to G_1$. The PKG randomly selects element $x \leftarrow Z_p^*$ as the master private key, computes the master public key $mpk = g^x$, and randomly selects element $\omega \leftarrow G_1$. Finally, the PKG securely and secretly saves x and discloses the system parameters:

$$params = (e, G_1, G_2, g, p, \omega, mpk, H_1, H_2, H_3)\,. \tag{1}$$

3.3.2. Generating anonymous identities and signing private keys. $Key\,Gen_{PID}$: The algorithm is executed by the PKG to generate an anonymous identity PID and a corresponding signing private key SK_{PID} for the data owner DO based on his real identity $ID \in \{0,1\}^\rho$. First, the PKG randomly selects $s \leftarrow Z_p^*$ and computes the anonymous identity PID of the DO using the master private key x. Specifically, PID consists of two parts:

$$PID = (PID_a, PID_b)\,, \tag{2}$$

$$PID_a = g^s, \tag{3}$$

$$PID_b = ID \oplus H_1\left(PID_a^x \,\|mpk\|\, Time\right), \tag{4}$$

where Time denotes the effective period for which the DO can maintain anonymity.

The PKG then computes a signing private key corresponding to the anonymized identity PID using the master private key:

$$SK_{PID} = (s + x)\, H_2\left(PID\right). \tag{5}$$

Finally, the PKG transmits $(PID, SK_{PID}, Time)$ to the DO over a secure channel.

3.3.3. Generating digital signatures. The algorithm is executed by the DO to generate the corresponding digital signature based on the data file that the DO itself is outsourcing, as well as the corresponding auxiliary integrity audit verification information. First, DO divides data F into n data blocks, specifically denoted as $F[1], F[2], \cdots, F[i] \in Z_p^* (i = 1, 2, \cdots, n)$, with file name $F_{id} \in Z_p^*$. Then, DO randomly selects $\sigma, \kappa, \gamma \in Z_p$, $t \in Z_p^*$, and computes the following information:

$$\varphi_1 = \sigma \cdot \kappa \cdot \gamma, \varphi_2 = \sigma^2 \cdot \kappa^2 \cdot \gamma, \cdots, \varphi_n = \sigma^n \cdot \kappa^n \cdot \gamma. \tag{6}$$

$$Y = g^t. \tag{7}$$

And generates an auxiliary integrity verification validation message:

$$IVA = \{\sigma, \kappa, \gamma, Y, H_3(\varphi_1), H_3(\varphi_2), \cdots, H_3(\varphi_n)\}. \tag{8}$$

In addition, the DO computes its digital signature Sig_i for each data block using the signing private key SK_{PID}:

$$R_i = (H_3(\varphi_i) \cdot H_3(PID))^t, \tag{9}$$

$$T_i = \left(\omega^{F[i]} H_2(F_{id} \| PID \| i)\right)^{SK_{PID}}, \tag{10}$$

$$Sig_i = R_i \cdot T_i. \tag{11}$$

Finally, the DO uploads the data block and digital signature $\left(\{F[i]\}_{1 \le i \le n}, \{Sig_i\}_{1 \le i \le n}, Y\right)$ to the CSP, and uploads the auxiliary audit information IVA and the anonymized identity PID to the blockchain's storage contract for preservation.

3.3.4. Generating challenge information. The algorithm is executed by the BC, which generates an audit challenge message based on the data recorded on the chain when it receives a request from a DO authorizing it to audit the integrity of the outsourced data stored in the CSP. First, a challenge contract deployed on the blockchain executes randomly selecting a subset $J = \{j_1, j_2, \cdots j_m\}$ containing m elements from set $\{1, 2, \cdots, n\}$, assigning a random coefficient $v_j \leftarrow Z_p$ to each subset j in subset J, and sending the challenge message $chal = \left\{(j, v_j)_{j \in J}\right\}$ to the CPS.While waiting for a response from the CPS, the following pre-computation is executed based on the auxiliary integrity audit information IVA stored on the chain:

$$\lambda = H_3(PID)^{\sum_{j=j_1}^{j_m} v_j} \cdot H_3(\gamma)^{\sum_{j=j_1}^{j_m} \cdot \sigma_j \cdot \kappa^j v_j}. \tag{12}$$

3.3.5. Data integrity audit. Verify: This algorithm is executed by the CSP, which generates the corresponding integrity audit proof response message upon receiving the challenge message $chal = \left\{(j, v_j)_{j \in J}\right\}$. First, the CPS picks a random value $\alpha \leftarrow Z^p$ for blinding the data block to prevent leakage and computes the proof:

$$\mu = \omega^{\alpha}. \tag{13}$$

The CPS aggregates the appropriate data blocks based on the challenged information:

$$\eta = \alpha^{-1} \sum_{j=h}^{j_m} v_j F[j] + \mu. \tag{14}$$

Then, the aggregated signature is computed:

$$Sig = \prod_{j=h}^{j_m} Sig_j^{v_j}. \tag{15}$$

Finally, CPS sends the integrity audit proof response message $\{\mu, \eta, Sig\}$ to the blockchain.

After BC receives the integrity audit proof response message $\{\mu, \eta, Sig\}$, it utilizes λ to determine whether the integrity verification equation is valid or not:

$$e(g, Sig) = e(Y, \lambda) \cdot e(Y, \lambda) \cdot e\left((PID_a \cdot mpk)^{H_2(PID)}, \mu^{p-\mu} \cdot \prod_{j=j_1}^{j_m} H_2(Fid \| PID \| i)^{v_j}\right). \tag{16}$$

If the validation equation holds, the data outsourced by DO to CPS is complete. Otherwise, the outsourced data is incomplete.

3.3.6. Smart contract deployment. In this scheme, the programmability, transparency, and immutability of smart contracts, as well as their ability to automate the handling of complex logic and mathematical computations, play a key role. In view of this, this scheme designs three types of smart contracts: storage contract, challenge contract and audit contract and deploys them on the blockchain in order to achieve the intended functions and goals. Among them, the storage contract realizes the storage and retrieval of auxiliary integrity audit information by defining two main functional interfaces, storeData and getData, respectively.

3.4. Experimental evaluation and results

In this section, the performance of the data integrity auditing scheme in this paper is evaluated by doing comparison with other schemes. Scheme 1 is a multi-copy data integrity verification scheme based on spatio-temporal chaos, Scheme 2 is a multi-copy data integrity verification scheme based on identity signature, and Scheme 3 is a blockchain cloud storage integrity auditing scheme based on T-Merkle hash tree. The three schemes are compared in terms of storage cost of blockchain and time overhead required under different dynamic update operations to verify the performance of the schemes in this paper.

3.4.1. Blockchain storage costs. There are many blockchain-based data integrity auditing schemes, but most of the schemes store the data or the tags of the data blocks on the blockchain, which leads to a high memory overhead of the blockchain. The scheme in this paper is related to smart contracts, which includes the storage and retrieval of smart contracts.

In ethereum, each transaction consumes a certain amount of gas. In this paper, a series of tests are conducted, and the gas consumption of smart contracts with different number of files (from 16 to 2048) is shown in Figure 2. The experimental results show that the number of consumed gases for smart contract storage and retrieval is maintained around 150,000 and 340,000, and the gas consumption of the contract is independent of the number of files stored. Therefore the blockchain storage overhead of the conditional identity anonymized data auditing scheme based on blockchain in this paper is extremely small.

3.4.2. Time overhead analysis. First, the experiment compares the average time overhead spent by the three schemes with the schemes in this chapter when inserting different amounts of data. The time overhead during the data insertion operation is shown in Figure 3, where the horizontal

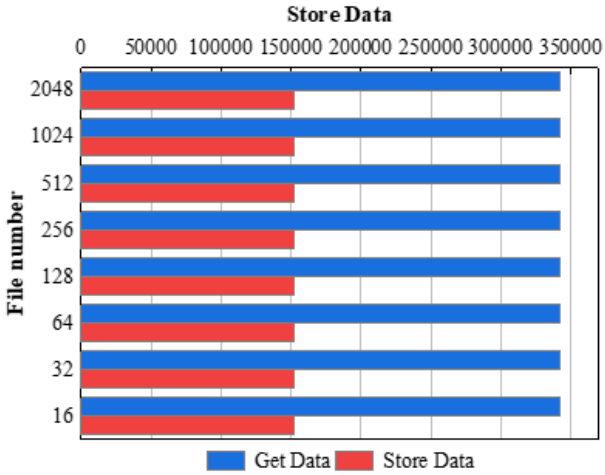


Fig. 2. The amount of intelligent contract gas consumption in different files

coordinates indicate the number of different inserted data and the vertical coordinates indicate the time overhead of the corresponding data amount for the insertion operation.

The experimental results show that the required time overhead tends to grow with the increase of inserted data. The average time overhead of this chapter's scheme with different numbers of inserted data is overall smaller than the other three schemes, with an average time overhead of around 3∼30ms. The average time overhead spent by scheme 3 is increasing with the amount of inserted data, and the average overhead time increases from 13ms to 35ms, and the average time overhead spent by schemes 1 and 2 does not float much at different amounts of inserted data, and stabilizes at about 38ms and 40ms, respectively.

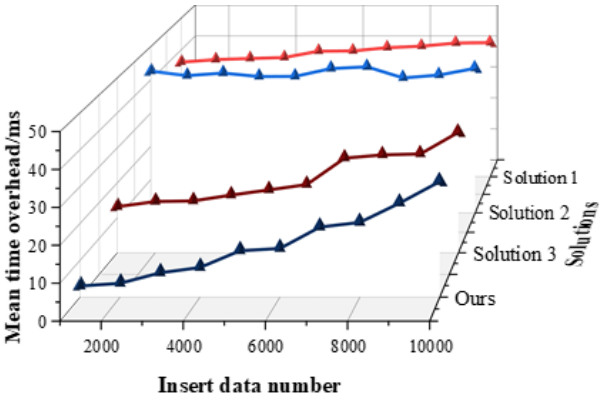


Fig. 3. Time overhead of data insertion operation

In the experiments of deletion operations, the average time overhead spent on deleting different amounts of data by the four schemes is also compared with the schemes in this chapter. The time overhead during data deletion operation is shown in Figure 4, where the horizontal coordinates indicate the different number of deleted data and the vertical coordinates indicate the time overhead of the corresponding data volume of the deletion operation.

From the experimental results, it can be seen that the average time overhead of this chapter's scheme at different numbers of deleted data has a significant advantage over the other schemes, and the average time overhead spent in scheme 3 increases with the increase of the amount of deleted data. The average overhead time of scheme 1 to scheme 3 and the blockchain based data integrity auditing scheme in this paper in the experiment are 38.18ms, 41.93ms, 21.26ms and 11.86ms respectively.

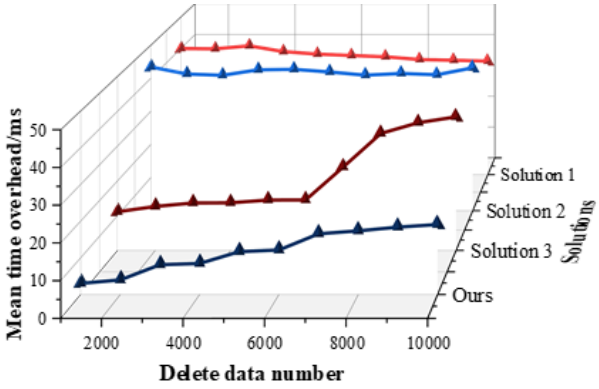


Fig. 4. Time overhead of data deletion operation

Subsequently, the experiments compare the average time overhead spent by the three schemes with the schemes in this chapter when updating different data volumes. The time overhead during the data deletion operation is shown in Figure 5, where the horizontal coordinate indicates the different number of updated data, and the vertical coordinate indicates the time overhead of the corresponding data volume of the update operation. The average overhead time of the blockchain-based data integrity auditing scheme in this paper is 2.5-20ms, and the average overhead time of scheme 1∼scheme 3 with different number of updated data is 37-40ms, 40-43ms, and 6.5-30ms, respectively.Similarly to the above experiments, it is concluded that the scheme of this chapter spends the least amount of time in updating different amount of data. The average time overhead spent by the schemes in this chapter gradually increases with the increase of the amount of data updated, but the schemes in this chapter are still relatively superior in terms of the overall average time overhead.

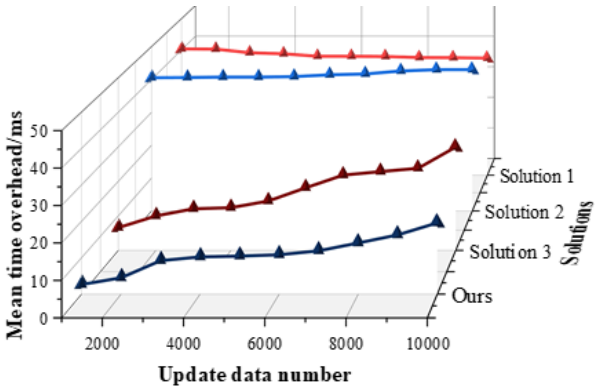


Fig. 5. Time overhead of data update operation

4. Blockchain privacy protection method based on differential privacy

Blockchain technology is used to build auditable and tamper-proof data storage solutions, especially for datasets with stringent security requirements. However, there are still challenges in protecting personal privacy information in databases, as any node can openly access personal privacy data stored on the blockchain due to the open and transparent nature of the blockchain. Therefore, there is a need to design a blockchain data sharing network architecture that satisfies the privacy and security requirements of multiple parties.

4.1. Blockchain network model

Aiming at the data privacy protection needs in blockchain, this paper designs a blockchain privacy protection model based on differential privacy, and the blockchain privacy protection model is shown in Figure 6. Users can carry out user operations through the data upload terminal and data access terminal of the client module, and make upload and access requests through the Web service with the middleware module. The smart contract module, on the other hand, automatically processes the target user's request according to the differential privacy chain code function and permission control chain code function installed in the blockchain network. The public data and identity certificates are stored in the public database of the blockchain, which can provide services to all nodes in the network, while the noise data is stored in the private ledger of the Hyperledger Fabric, and only the privileged nodes can use the private data through the authentication by the privilege control function.

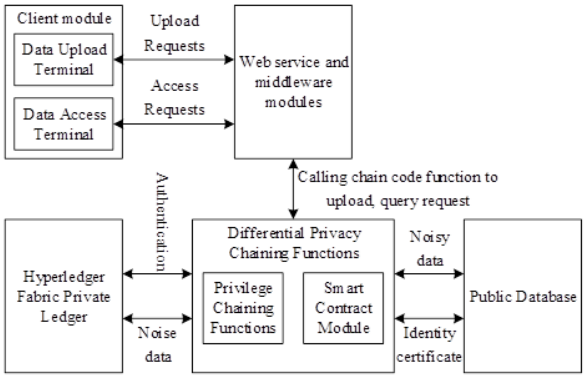


Fig. 6. Blockchain data privacy protection model based on differential privacy

4.2. Differential privacy algorithm

Differential privacy algorithm design for data sharing removes attributes such as name, identification and other attributes for anonymization when an individual or organization uploads information to the blockchain network for data sharing, and noise is added to the data by the differential privacy mechanism in the smart contract to scramble the uploaded personal data to prevent differential attacks.

4.2.1. Laplace random noise. By integrating the differential privacy mechanism into the blockchain smart contract layer, it realizes the process of automatically calling the chain code to add noise to the data in the process of user uploading data. By adding random Laplace noise to numerical data, the perturbation of the original data is realized, which ensures the privacy of user data while the data features are not damaged, and the querier can still analyze the relevant data.

Laplace noise function is for numerical data to add a random number in line with the Laplace distribution: take the random variable $\alpha \sim UNI\,(0,1)$ to meet the uniform distribution, and brought into the inverse function of the Laplace cumulative distribution function, then the noise value can be obtained to meet the conditions of the formula is as follows.

$$F^{-1}(x) = \begin{cases} \lambda \ln(2\alpha) & \alpha < 1/2, \\ \mu - \lambda \ln(2\alpha) & \alpha > 1/2. \end{cases} \tag{17}$$

If the uniform distribution $\alpha \sim UNI(-0.5, 0.5)$ is taken, the above segmented function can be expressed in the form of an equation where the *sign* function is used to obtain the positive and negative of the parameter and the *abs* function is used to obtain the absolute value. The noise value is the:

$$F^{-1}(x) = \mu - \lambda * sign(\alpha) * \ln(1 - 2 * abs(\alpha)). \tag{18}$$

Add the computed random Laplace noise values to the data as in:

$$M(D) = f(D) + Lap\left(0, \frac{\Delta f}{\varepsilon}\right). \tag{19}$$

The privacy budget ε is inversely proportional to the size of the added noise value, the smaller the privacy budget, the larger the added noise, the higher the privacy protection strength of the data, and the lower the data usability, and vice versa, the larger the privacy budget, the smaller the added noise, the lower the protection strength of the data but the higher the usability. According to the privacy requirements of a specific dataset, adding a privacy budget that meets the requirements can balance the privacy protection utility level and data availability.

4.2.2. Stochastic response. The random response is a noise function that performs random flipping, for example, the "s" attribute in the personal information of the data used is that the individual has two values of "yes" and "no", a noise that satisfies $\varepsilon-$ differential privacy is added to it, and the flip probability p is calculated through the given privacy budget ε to achieve the purpose of protecting personal privacy data, and the availability of data is retained, and the user can calculate the maximum likelihood estimate of each value in the original data through ε for data analysis.

Suppose that the sample of questions and answers are counted and the number of "s" attributes is counted. The proportion of true answers is given as π. Assume that the number of people who answered "yes" is n_1 and the number of people who answered "no" is n_2, and there are:

$$\begin{cases} P_r[x_i = "yes"] = \pi * p + (1 - \pi) * (1 - p) \\ P_r[x_i = "no"] = (1 - \pi) * p + \pi * (1 - p) \end{cases} \tag{20}$$

Unbiased estimation using the method of great likelihood:

$$L = (\pi * p + (1 - \pi) * (1 - p))^{n1} * ((1 - \pi) * p + \pi * (1 - p))^{n1}. \tag{21}$$

Obtain an unbiased estimate of π for $\tilde{\pi}$:

$$\tilde{\pi} = \frac{p - 1}{2p - 1} + \frac{n_1}{(2p - 1)n}. \tag{22}$$

Estimated number of "s" attributes:

$$\tilde{n} = n * \pi = \frac{p - 1}{2p - 1} n + \frac{n_1}{2p - 1}. \tag{23}$$

Noise addition using random flipping can effectively protect the original data from theft, and the overall features of the original dataset can still be obtained by unbiased estimation of the data processing.

4.2.3. Responding to queries. Algorithm (17) will set the corresponding sensitivity and privacy budget parameters according to the type of data to be uploaded, and the publisher of the data will standardize the attributes and format of the data to be uploaded, set the noise adding function in the attribute part of the user's sensitive information, and do not add the data perturbation to the rest of the attribute information as a way of maximizing the usability of the data.

Users can verify the data by sending verification requests to the endorsing nodes to prevent their uploaded private data from being tampered by other attackers, and the consensus property of the blockchain itself also requires the process of data verification. In algorithm (18), the endorsing node sends its identity information and the data identifier to be verified to the blockchain network through the client, and the smart contract retrieves and calls the noisy data in the private database of the corresponding identifier of the verified data according to the user's authority, and returns it to the verifier through denoising to complete the security verification of the data and ensure the data consistency of the blockchain network.

4.2.4. Privacy analysis. Assuming that the DPNA algorithm computes the model with a noise privacy budget of ε, then the DPNA algorithm satisfies $\varepsilon-$ differential privacy.

Proof: with two neighboring datasets D and D', and let $f(\cdot)$ be the feature correlation function on the neighboring datasets, and M denote the correlation feature computed for any one record, the global sensitivity formula is as follows:

$$\Delta f_M = \max_{D,D'} \|f_M(D) - f_M(D')\|_I . \tag{24}$$

According to the definition of Laplace distribution, the probability density function of DPNA algorithm is as follows:

$$P_r[D] = \frac{1}{2b}\exp\left(-\frac{|f_M(D)-M|}{b}\right) = \frac{\varepsilon}{2\Delta f_M}\exp\left(-\frac{\varepsilon|f_M(D)-M|}{\Delta f_M}\right). \tag{25}$$

Thus the ratio of the probability density functions of the results computed by the DPNA algorithm on two neighboring datasets is as follows:

$$\begin{aligned}
\frac{P_M(D)}{P_M(D')} &= \prod_{j=1}^{d}\exp\left(\frac{\varepsilon|f_M(D)|_j - M_j|}{\Delta f_M}\right) \Big/ \exp\left(\frac{\varepsilon|f_M(D')|_j - M_j|}{\Delta f_M}\right) \\
&= \prod_{j=1}^{d}\exp\left(\frac{\varepsilon\left(|f_M(D')|_j - M_j\right) - \left(\varepsilon|f_M(D)|_j - M_j\right)}{\Delta f_M}\right) \\
&\le \prod_{j=1}^{d}\exp\left(\frac{\varepsilon|f_M(D)_j - f_M(D')_j|}{\Delta f_M}\right) \\
&= \exp\left(\frac{\varepsilon||f_M(D) - f_M(D')||_1}{\Delta f_M}\right) \\
&\le \exp(\varepsilon)
\end{aligned} \tag{26}$$

Therefore, the above satisfies the definition of differential privacy, i.e., the DPNA algorithm satisfies $\varepsilon-$differential privacy when the given privacy budget is ε. Thus, the proof is complete.

4.3. Experimental results and analysis

4.3.1. Experimental environment. All experiments in this chapter were carried out on a processor Intel 2.70GHz i7-7500U CPU, operating system 64-bit Windows 10 OS, blockchain network deployed

using Hyperledger Fabric 1.4, consisting of two organizations each consisting of two peer-to-peer nodes and two users, and relevant code written using Go. Time measurements are performed mainly for differential privacy algorithm performance.

4.3.2. Experimental results. In order to test the performance of the scheme, three experiments are designed in this section to study the encryption and decryption time of data files of different sizes, the impact of data access control policies of different complexity on the encryption and decryption computation time, and the computation time comparison of the two algorithms, which are designed as follows.

Experiment 1. The experiment sets up data files of different sizes, uses the differential privacy algorithm to perform encryption and decryption operations on the data, and takes the final computation time as the basis for studying the relationship between data size and data encryption and decryption cost.

The encryption and decryption times for data files of different sizes are shown in Figure 7. With the increase of data size, the length of encryption and decryption are all in a linear growth trend. In the actual super ledger transaction, a single block can hold a maximum of 10MB of data, so the data is a maximum of 10MB, at this time, the encryption time is about 0.075s or so, and the decryption time is about 0.063s. According to the encryption and decryption time length obtained from the experiment, this paper finds that the realization time of these two operations are within the acceptable range of the user, so this scheme has good feasibility.

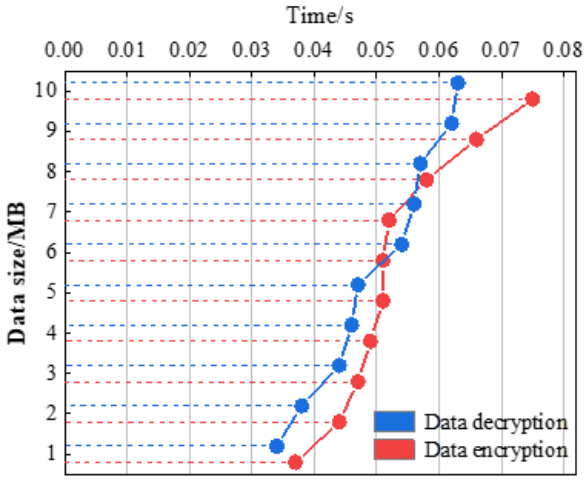
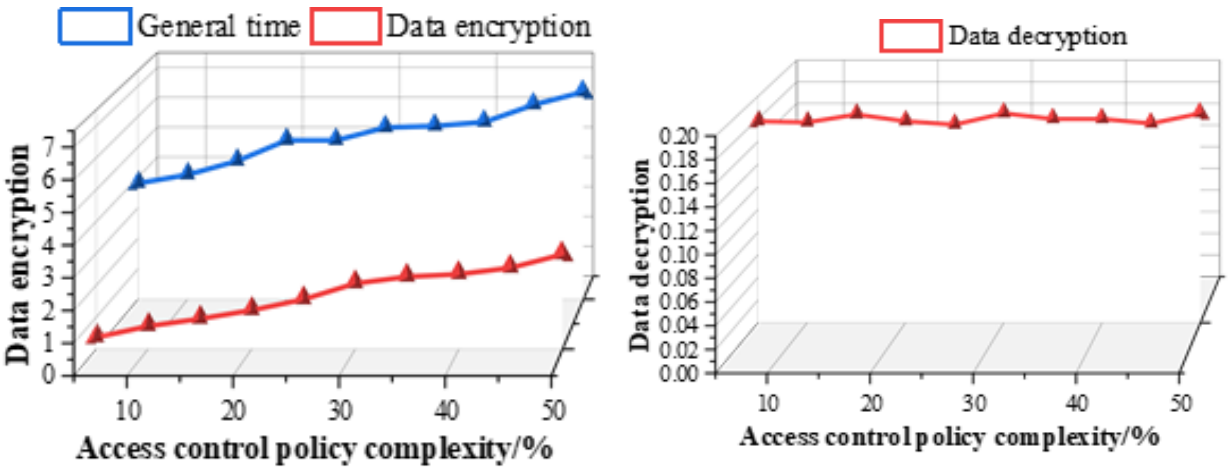


Fig. 7. The time of encryption and reconciliation of different size data files

Experiment 2. Setting different degrees of access control policies, obtaining the running time of the differential privacy algorithm for access control policies of different complexity and the overall running time of the scheme, and investigating the effect of the complexity of data access control policies on the running time.

The computation time under access control policies of different complexity is shown in Figure 8, where 8a is the data encryption time length and 8b is the data decryption time length. As the complexity of the access policy increases, the time consumption of the encryption process increases, and the overall running time of the scheme increases, and the data encryption time and the overall running time of the scheme in the experiments are below 3s and 7s, respectively. Therefore, the time overhead of the blockchain privacy protection scheme based on differential privacy is small in the whole chain phase, even when the complexity of the access control policy is as high as 50%, its

computation time is only 2.97s, and the realized time is all within the acceptable range. Moreover, in Figure 8b, the decryption speed is independent of the policy complexity. Due to the different strategy complexity of the decryption process, the decryption time consumption varies slightly with the additive access strategy complexity in the test, and the time cost stabilizes around 0.18s, which minimizes the overall time consumption.



(a) Data encryption time

(b) Data decryption time

Fig. 8. Computational time in different complexity access control strategies

Experiment 3. In order to better compare the overall time consumption of the scenarios in terms of time consumption at each stage, the overall time overhead of the test scenario is combined with the actual situation, the experiments are set to upload and download files with a size of 600 MB, respectively, using the CP-ABE algorithm, the CP-ABE-AES algorithm and the differential privacy algorithm of this paper for the overall time consumption test, and the three algorithms are used for the overall time consumption test, according to the running time of the algorithms Compare the performance advantages and disadvantages.

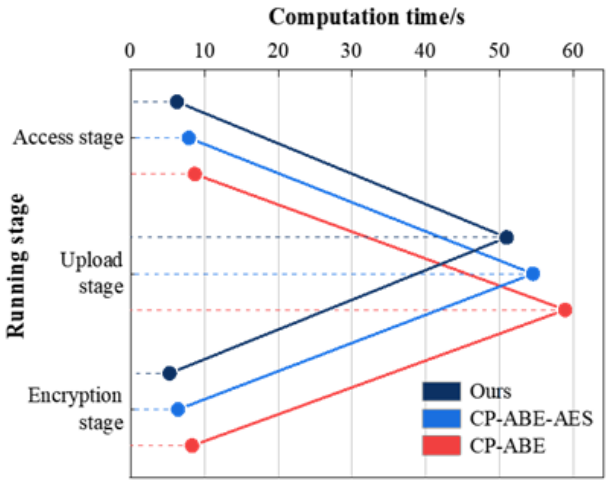


Fig. 9. The algorithms take time in different stages

The time consumed by the algorithms under different stages is shown in Figure 9. The computation time of all stages of this paper's differential privacy algorithm is smaller than the comparison algorithm. First of all, the data encryption and decryption duration of all 3 algorithms is controlled within 10s, whether it is the data encryption phase or the data access phase, the time overhead of the

differential privacy algorithm used in this scheme is less than the comparison algorithm. Therefore, the differential privacy algorithm is considered to have better performance and provide more efficient services to users. Secondly, in the whole scheme, the most time-consuming is the data transmission process, which is basically maintained at about 50∼60s for the 3 schemes, and since the file size used in this case is 600MB, it needs to be transmitted several times, and although the transmission process is affected by the bandwidth, network latency, etc., the experimental data obtained has no degradation in terms of performance performance. In conclusion, the blockchain privacy protection scheme based on differential privacy has a significant speed advantage in uploading large data files at all stages of the data.

5. Conclusion

In the context of the current digital era, data security and privacy protection have risen to be the core issues of common global concern. In this context, the study proposes a data integrity auditing scheme with conditional identity anonymization based on blockchain technology. Meanwhile, a blockchain privacy protection model based on differential privacy is proposed for the privacy protection of shared data in blockchain networks. The two methods are evaluated through experiments, and the main results are as follows:

(1) The blockchain storage overhead of the data integrity auditing scheme in this paper is small, and the time overhead under different dynamic operations is lower than other methods. In the process of inserting data, deleting data and updating data, the average time overhead of this paper's scheme is 3∼30ms, 3∼20ms and 2.5∼20ms, respectively.It confirms that this scheme outperforms other schemes in terms of blockchain storage cost and computation overhead.

(2) The encryption and decryption durations of the differential privacy-based blockchain privacy protection method are lower than 0.075s and 0.063s for different sizes of data files, the encryption durations and the total runtime in different complexity access policies are lower than 3s and 7s, respectively, and the processing times in different runtime phases are smaller than those of the comparison methods. Therefore, it can be considered that the blockchain privacy protection method based on differential privacy has better performance and can provide more efficient and convenient services.

The blockchain technology has great potential in data integrity and privacy protection, but at the same time, it also needs to pay attention to the challenges and problems it faces. In the future, with the continuous development and improvement of blockchain technology, it is believed that it will bring more extensive applications and breakthroughs in the field of network security.

About the Authors

Yihui Deng was born in Zhanjiang, Guangdong, P.R. China, in 1990. He received the bachelor's degree from Jiangxi University of Technology, P.R. China. Now, he works in Experimental Training Center, Guangzhou College of Applied Science and Technology. My main research direction is Computer networks and information security.

Sanxiang Xiao was born in October 1986. He received the Master degree from Central South University of Forestry and Technology, P.R. China. Now, he works in School of Computer Science, Guangzhou College of Applied Science and Technology. My main research direction is Computer networks and artificial intelligence.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

[1] Y. Chen, J. Li, F. Wang, K. Yue, Y. Li, B. Xing, L. Zhang, and L. Chen. Ds2pm: a data-sharing privacy protection model based on blockchain and federated learning. *IEEE Internet of Things Journal*, 10(14):12112–12125, 2021. https://doi.org/10.1109/JIOT.2021.3134755.

[2] Y. Chen, H. Xie, K. Lv, S. Wei, and C. Hu. Deplest: a blockchain-based privacy-preserving distributed database toward user behaviors in social networks. *Information Sciences*, 501:100–117, 2019. https://doi.org/10.1016/j.ins.2019.05.092.

[3] S. Fan, L. Song, and C. Sang. Research on privacy protection in iot system based on blockchain. In *Smart Blockchain: Second International Conference, SmartBlock 2019, Birmingham, UK, October 11–13, 2019, Proceedings 2*, pages 1–10. Springer, 2019. https://doi.org/10.1007/978-3-030-34083-4_1.

[4] Q. Feng, D. He, S. Zeadally, M. K. Khan, and N. Kumar. A survey on privacy protection in blockchain system. *Journal of Network and Computer Applications*, 126:45–58, 2019. https://doi.org/10.1016/j.jnca.2018.10.020.

[5] K. Gai, M. Qiu, and H. Zhao. Privacy-preserving data encryption strategy for big data in mobile cloud computing. *IEEE Transactions on Big Data*, 7(4):678–688, 2017. https://doi.org/10.1109/TBDATA.2017.2705807.

[6] L. Guo, H. Xie, and Y. Li. Data encryption based blockchain and privacy preserving mechanisms towards big data. *Journal of Visual Communication and Image Representation*, 70:102741, 2020. https://doi.org/10.1016/j.jvcir.2019.102741.

[7] R. Henry, A. Herzberg, and A. Kate. Blockchain access privacy: challenges and directions. *IEEE Security & Privacy*, 16(4):38–45, 2018. https://doi.org/10.1109/MSP.2018.3111245.

[8] B. Jia, T. Zhou, W. Li, Z. Liu, and J. Zhang. A blockchain-based location privacy protection incentive mechanism in crowd sensing networks. *Sensors*, 18(11):3894, 2018. https://doi.org/10.3390/s18113894.

[9] B. Jin, D. Jiang, J. Xiong, L. Chen, and Q. Li. D2d data privacy protection mechanism based on reliability and homomorphic encryption. *IEEE Access*, 6:51140–51150, 2018. https://doi.org/10.1109/ACCESS.2018.2869575.

[10] H. Li, L. Pei, D. Liao, G. Sun, and D. Xu. Blockchain meets vanet: an architecture for identity and location privacy protection in vanet. *Peer-to-Peer Networking and Applications*, 12:1178–1193, 2019. https://doi.org/10.1007/s12083-019-00786-4.

[11] W. Liang, Y. Yang, C. Yang, Y. Hu, S. Xie, K.-C. Li, and J. Cao. Pdpchain: a consortium blockchain-based privacy protection scheme for personal data. *IEEE Transactions on Reliability*, 72(2):586–598, 2022. https://doi.org/10.1109/TR.2022.3190932.

[12] Y. Liu, J. Zhang, and J. Zhan. Privacy protection for fog computing and the internet of things data based on blockchain. *Cluster Computing*, 24(2):1331–1345, 2021. https://doi.org/10.1007/s10586-020-03190-3.

[13] Z. Lv, L. Qiao, M. S. Hossain, and B. J. Choi. Analysis of using blockchain to protect the privacy of drone big data. *IEEE Network*, 35(1):44–49, 2021. https://doi.org/10.1109/MNET.011.2000154.

[14] Z. Min, G. Yang, A. K. Sangaiah, S. Bai, and G. Liu. A privacy protection-oriented parallel fully homomorphic encryption algorithm in cyber physical systems. *EURASIP Journal on Wireless Communications and Networking*, 2019:1–14, 2019. https://doi.org/10.1186/s13638-018-1317-9.

[15] P. J. Sun. Privacy protection and data security in cloud computing: a survey, challenges, and solutions. *Ieee Access*, 7:147420–147452, 2019. https://doi.org/10.1109/ACCESS.2019.2946185.

[16] P. Sun. Security and privacy protection in cloud computing: discussions and challenges. *Journal of Network and Computer Applications*, 160:102642, 2020. https://doi.org/10.1016/j.jnca.2020.102642.

[17] Z. Sun, Y. Wang, Z. Cai, T. Liu, X. Tong, and N. Jiang. A two-stage privacy protection mechanism based on blockchain in mobile crowdsourcing. *International Journal of Intelligent Systems*, 36(5):2058–2080, 2021. https://doi.org/10.1002/int.22371.

[18] D. Wang, J. Zhao, and Y. Wang. A survey on privacy protection of blockchain: the technology and application. *IEEE Access*, 8:108766–108781, 2020. https://doi.org/10.1109/ACCESS.2020.2994294.

[19] Y. Wang, X. Liang, X. Hei, W. Ji, and L. Zhu. Deep learning data privacy protection based on homomorphic encryption in aiot. *Mobile Information Systems*, 2021(1):5510857, 2021. https://doi.org/10.1155/2021/5510857.

[20] Z. Wang, S. Chaliasos, K. Qin, L. Zhou, L. Gao, P. Berrang, B. Livshits, and A. Gervais. On how zero-knowledge proof blockchain mixers improve, and worsen user privacy. In *Proceedings of the ACM Web Conference 2023*, pages 2022–2032, 2023. https://doi.org/10.1145/3543507.3583217.

[21] W. Wei, S. Liu, W. Li, and D. Du. Fractal intelligent privacy protection in online social network using attribute-based encryption schemes. *IEEE Transactions on Computational Social Systems*, 5(3):736–747, 2018. https://doi.org/10.1109/TCSS.2018.2855047.

[22] B. Wen, Y. Wang, Y. Ding, H. Zheng, B. Qin, and C. Yang. Security and privacy protection technologies in securing blockchain applications. *Information Sciences*, 645:119322, 2023. https://doi.org/10.1016/j.ins.2023.119322.

[23] X. Yan, Q. Wu, and Y. Sun. A homomorphic encryption and privacy protection method based on blockchain and edge computing. *Wireless Communications and Mobile Computing*, 2020(1):8832341, 2020. https://doi.org/10.1155/2020/8832341.

[24] C. Yang, L. Tan, N. Shi, B. Xu, Y. Cao, and K. Yu. Authprivacychain: a blockchain-based access control framework with privacy protection in cloud. *Ieee Access*, 8:70604–70615, 2020. https://doi.org/10.1109/ACCESS.2020.2985762.

[25] A. Yazdinejad, R. M. Parizi, A. Dehghantanha, and K.-K. R. Choo. Blockchain-enabled authentication handover with efficient privacy protection in sdn-based 5g networks. *IEEE Transactions on Network Science and Engineering*, 8(2):1120–1132, 2019. https://doi.org/10.1109/TNSE.2019.2937481.

[26] L. Zhang, L. Li, E. Medwedeff, H. Huang, X. Fu, and R. Wang. Privacy protection of social networks based on classified attribute encryption. *Security and Communication Networks*, 2019(1):9108759, 2019. https://doi.org/10.1155/2019/9108759.

[27] R. Zhang, R. Xue, and L. Liu. Security and privacy on blockchain. *ACM Computing Surveys (CSUR)*, 52(3):1–34, 2019. https://doi.org/10.1145/3316481.