Computers, Materials & Continua
DOI: 10.32604/cmc.2023.028843
Article

Tech Science Press

Privacy Data Management Mechanism Based on Blockchain and Federated Learning

Mingsen Mo[1], Shan Ji[2], Xiaowan Wang[3,*], Ghulam Mohiuddin[4] and Yongjun Ren[1]

[1]Engineering Research Center of Digital Forensics of Ministry of Education, School of Computer, Nanjing University of Information Science & Technology, Nanjing, 210044, China
[2]Zhengde Polytechnic, Nanjing, 211106, China
[3]Xi'an University of Posts & Telecommunications, Xi'an, 710061, China
[4]Department of Cyber Security at VaporVM, Abu Dhabi, 999041, United Arab Emirates
*Corresponding Author: Xiaowan Wang. Email: mms101200@163.com

Received: 19 February 2022; Accepted: 19 April 2022

Abstract: Due to the extensive use of various intelligent terminals and the popularity of network social tools, a large amount of data in the field of medical emerged. How to manage these massive data safely and reliably has become an important challenge for the medical network community. This paper proposes a data management framework of medical network community based on Consortium Blockchain (*CB*) and Federated learning (*FL*), which realizes the data security sharing between medical institutions and research institutions. Under this framework, the data security sharing mechanism of medical network community based on smart contract and the data privacy protection mechanism based on *FL* and alliance chain are designed to ensure the security of data and the privacy of important data in medical network community, respectively. An intelligent contract system based on Keyed-Homomorphic Public Key (*KH-PKE*) Encryption scheme is designed, so that medical data can be saved in the *CB* in the form of ciphertext, and the automatic sharing of data is realized. Zero knowledge mechanism is used to ensure the correctness of shared data. Moreover, the zero-knowledge mechanism introduces the dynamic group signature mechanism of chosen ciphertext attack (*CCA*) anonymity, which makes the scheme more efficient in computing and communication cost. In the end of this paper, the performance of the scheme is analyzed from both asymptotic and practical aspects. Through experimental comparative analysis, the scheme proposed in this paper is more effective and feasible.

Keywords: Data management; blockchain; federated learning

1 Introduction

With the accelerating process of digitization of medical systems in various countries, medical data shows an exponential upward trend [1]. Therefore, the fields of public medical management, online patient access and medical data management have become the research topics in academic and medical circles. The Internet has largely improved the traditional pattern of medical interrogation, thus forming a medical network community dominated by patients, research institutions and medical institutions. The community is conducive to alleviating the contradiction between patients and society. Different measures taken by relevant industries in various countries have proved this situation to a certain extent. In Estonia [2], a digital medical infrastructure was created to form a healthy community network of citizens and health care stakeholders (providers, insurance companies, etc.). Through this way, it is of great significance to study the reform of medical model under the new situation. The recent 2019 New Coronavirus (also known as 2019-nCoV and COVID-2019) pandemic also demonstrated the necessity to establish a medical network community. Through the management of patient medical data in various countries, research institutions can conduct data visual analysis based on big data [3], and provide government and other-decision-making institutions with real-time research and judgment and make these institutions able to predict the epidemic situation, which greatly reduces the momentum of public panic about the epidemic situation [4].

In the medical network community, the patient's personal information data and condition information will be recorded to form big data [5]. Big data has been applied in many fields and plays a particularly critical role in the decision-making process of many hospital institutions [6]. For the security needs of data management, medical institutions have the responsibility to take confidentiality measures for patients' private data [7,8]. If being leaked, this information may be biased. *FL* keeps all sensitive data in the local organization to which the data belongs, which makes it possible to combine fragmented medical data sources with privacy protection. Although the training method of *FL* exchanging model parameters without exchanging specific data can effectively protect users' privacy, *FL* still faces some security risks. In the training process of *FL* model, though the local original data of each user will not be disclosed, but it still have limitation that if there are 'dishonest', 'honest and curious' servers or malicious clients, the user's local data information may still be deduced from the updated model parameters, that is, reasoning attack, poisoning attack and attack of Generative Adversarial Networks (*GAN*) and other types of attacks [9]. Therefore, the privacy protection ability of important data under *FL* environment is insufficient.

Security and privacy of data management are priorities [10]. In order to make up for the lack of privacy protection of important data in *FL*, we use *KH-PKE* scheme to hide patient data information and data sources. At the same time, *FL* algorithms can reduce the frequency of data communication and reduce the risk of exposing patient data. Therefore, when considering data security, we use the smart contract system based on the above-mentioned *KH-PKE* solution instead of the joint server to realize that medical data can be uploaded to the *CB* and realize the automatic sharing of medical data. In each round of federation, smart contracts collect data from medical and research institutions and return aggregated parameters for all parties to automatically update the Machine Learning model. In view of the openness and distributed execution of smart contracts, all parties can also verify the correct execution of the steps in an open manner at any time. The goal is to construct a zero-knowledge proof scheme for lightweight devices under *FL* without excessively increasing the scale of proof, so as to realize medical data management. It hides the addresses of the sender and receiver to achieve anonymity and privacy protection. In non-interactive zero knowledge (*NIZK*), the dynamic group signature scheme based on *CCA* anonymity we used makes the *NIZK* scheme more secure and efficient in computing and communication cost. In the end, we combined theory and practice to analyze, and simulated and implemented the program on a personal computer.

We organize the rest of this article as follows. The second section introduces the related work, and the third section mainly constructs the data security management framework of medical network community based on *CB* and *FL*. The fourth section mainly introduces the data security sharing mechanism of medical network community based on smart contract. The fifth section mainly constructs the data privacy protection mechanism based on *FL* and *CB* and compares it with the previous schemes. The sixth section is the conclusions of this paper.

2 Related Work

2.1 Data Management in Consortium Blockchain

The ideal medical data management scheme should meet the following basic requirements, namely security and privacy protection [11], data access [12], access control and unified standards [13]. Therefore, the scheme proposed in this paper should meet the following requirements.

1. Security and privacy protection: no one may illegally use medical data. The program should be able to ensure that the data resists illegal attacks.
2. Data access: after obtaining the authorization, the research institution can view all relevant medical records, and the research institution can access the previous medical information under the authorization of the medical institution.
3. Access control: only medical institutions can manage patient data, that is, no one can obtain historical data without the consent of the medical institution.
4. Unified standards: unified data management standards should be adopted in the model to balance the overall stability of the system.

2.2 Data Management in the Consortium Blockchain with Federated Learning

The powerful, robust, flexible and secure functions of *FL* data cognition is helpful to solve the problem of the reliability, security, privacy protection of *CB* data management and optimize the storage data cost of blockchain data sharing.

Providing a trusted mechanism for all participants of *FL* through blockchain. The parameters of the *FL* model can be stored in the blockchain to ensure its security and reliability. *FL* has the characteristics of distributed intelligence and data privacy protection. It combines with the *CB* to complement each other's advantages and improve the overall security of the system.

Some researchers use blockchain as a secure distributed ledger, which provides a potential solution for cross system management of medical information. Li et al. [14] constructed a privacy data storage protocol based on elliptic curve using ring signature to ensure data security and user identity privacy in blockchain applications. Omar et al. [15] proposed a patient-centered medical data management system medical chain, which realized privacy protection based on blockchain. Liu et al. [16] proposed a privacy protection electronic medical records management scheme based on blockchain. The original data is stored in the cloud environment, and the data index is retained in the *CB* to reduce the risk of disclosure [17]. However, the above scheme is not suitable for the confidentiality and constantly updated data requirements in our system.

3 Data Security Management Framework of Medical Network Community

3.1 Privacy Data Protection in Federated Learning

In the process of *FL* model training, the user's local data information may still be deduced from the updated model parameters by reasoning attack, and may also be subject to poisoning attack, and attack of *GAN* and other types of attacks. At the same time, *FL* does not detect and verify the participants. Malicious participants may attack and destroy the *FL* training process by providing false model parameters, resulting in training failure. Therefore, the privacy protection ability of important data in *FL* is insufficient.

In the medical network community, data security is the premise of medical big data management. *FL*'s common privacy protection technologies, such as secure multi-party computing and differential privacy, are not enough to ensure data security [18,19]. At the same time, if privacy data was leaked, it would cause serious harm to patients. However, the development trend of medical big data is sharing and opening. Therefore, we should strengthen the ability of privacy protection in *FL* on the premise of ensuring the effective sharing of training data among participants, so as to give full play to the value of data management.

In order to make up for the lack of privacy protection of important data in *FL*, a zero-knowledge proof scheme suitable for lightweight devices is constructed to realize medical data management. In *NIZK* scheme, *CCA* anonymous dynamic group signature mechanism [20] is more secure and efficient in computing and communication cost. The signature mechanism can be used in the *CB* to provide unlinkability and anonymity, so as to improve the ability of data management and privacy protection in the medical *CB*. At the same time, in order to consider the security of private data, a smart contract system for the safe sharing of medical data is constructed based on the *KH-PKE* scheme to realize that the medical data hidden and stored in medical institutions and research institutions are stored on the *CB* for automatic data sharing operation to meet the needs of more High privacy requirements.

3.2 Medical Network Community Based on Consortium Blockchain and Federated Learning

Consortium Blockchain network: as a part of the blockchain [21], the *CB* has more advantages in efficiency and flexibility than the public blockchain, and can provide better privacy protection. It uses the distributed and tamper proof characteristics of blockchain to realize the secure storage and sharing of medical and health data and reduce the managementcost. as shown in Fig. 1.

Federated learning network: providing a trusted mechanism for all participants of federated learning through *CB*. On an artificial neural network predictive model (*ANN*) [22], use weight parameters and biases to train medical data and upload the trained data to a smart contract for data sharing. Keeping medical data on the blockchain throughout the data sharing environment. At the same time, the global model is stored and shared using the blockchain, thus avoiding a potential single point of failure on a central server.

Entity: including research institutions and medical institutions. The mechanism of *KH-PKE* scheme is used to directly store ciphertext and perform some calculations on the *FL* and blockchain. There is no need to make any change to the *FL* and *CB* itself, so as to provide confidentiality and privacy protection in the process of data management between research institutions and medical institutions.

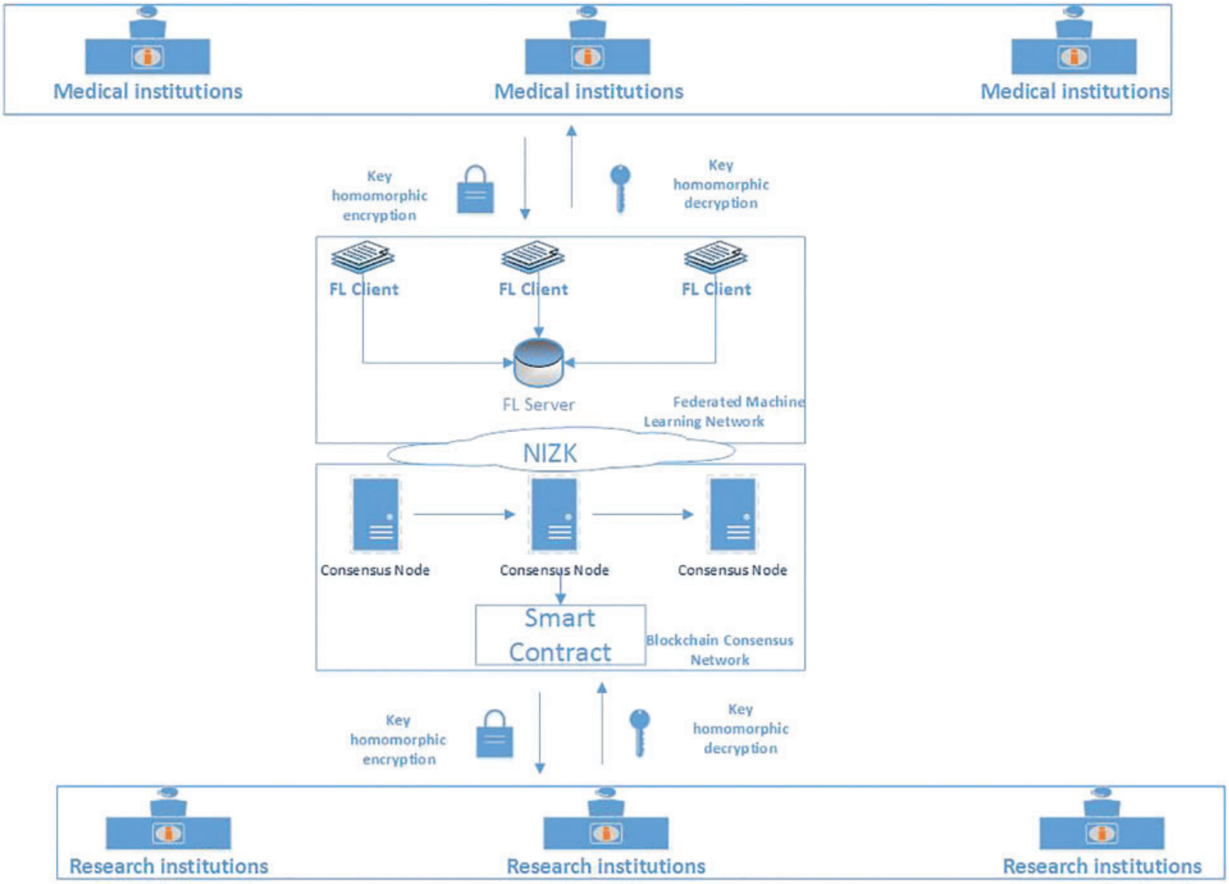


Figure 1: Medical network community framework based on Consortium Blockchain and Federated learning

4 Data Security Sharing Mechanism of Medical Network Community

CB can provide technical support for the security protocol and training method of *FL* algorithm. After the data of medical institutions are encrypted by the mechanism of *KH-PKE* scheme, the local model parameters are calculated by *FL*. Finally, upload the data to the smart contract, aggregate different data information and return the results. At the same time, smart contracts can also resist poisoning attacks and improve the security of private data.

In the medical network community, in order to consider the security of private data, public key encryption schemes are usually used to hide medical data. However, the fact that homomorphic public key encryption schemes are vulnerable to (adaptive) ciphertext selection attacks has been ignored to some extent. Theoretically, the adversary sends a homomorphic evaluation challenge ciphertext to the decryption oracle, and can immediately destroy the security. Therefore, we use the *KH-PKE* scheme to hide medical privacy data to meet higher privacy requirements. Homomorphic encryption scheme can only achieve indistinguishable encryptions under chosen plaintext attack security. Our *KH-PKE* scheme evaluates the secret key by controlling the homomorphic operation function and achieves stronger security than indistinguishable encryptions under chosen ciphertext attack(*IND-CCA1*) and is as close to indistinguishable encryptions under adaptive chosen ciphertext attack(*IND-CCA2*) as possible. Then, the structure is described under the secure under the decision linear(*DLIN*) assumption, and its security proof is given.

Definition 1 (*KH-PKE*): let M be a message space and $\odot$ be a binary operation on M. We require that for all $m_1, m_2, m_3 \in M$, $m_1 \odot m_2 \odot m_3 \in M$. The *KH-PKE* scheme (*Gen, Enc, Dec, Eval*) for homomorphic operation $\odot$ consists of the following four algorithms:

Gen: The secret key generation algorithm takes the security parameter $n \in N$ as the input and selects $h \leftarrow_{\$} G, \gamma, \alpha_1, \alpha_2 \leftarrow_{\$} Z_p$: output public key $gpk = (X_1 = g^{\gamma}, X_2 = g^{\alpha_1}, X_3 = g^{\alpha_2}, g, h)$, decrypt secret key $gsk = (x, y)$and homomorphic operation secret key sk_h.

Enc: The encryption algorithm takes gpk and a message $m \in M$ as input to calculate $C_1 = X_1^r, C_2 = X_2^s, C_3 = g^{r+s} \cdot h^m$, and output $C = (C_1, C_2, C_3)$, where $r, s \leftarrow_{\$} Z_p$ represents the randomness of *Enc* use.

Dec: The decryption algorithm takes gsk and C as inputs, parses C into a tuple (C_1, C_2, C_3), and calculates $hm = C_3/(C_1^{\frac{1}{x}} \cdot C_2^{\frac{1}{y}})$. If the plaintext space is small, the message $m = log_h^{hm}$ can be obtained effectively and m or $\perp$ can be output.

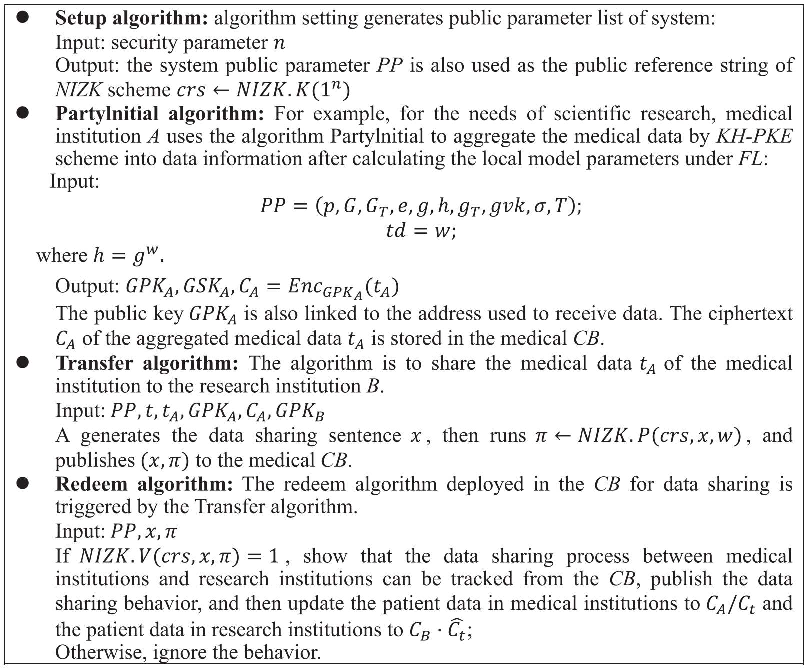
Eval: The evaluation algorithm adopts sk_h. Three ciphertext C_1, C_2, C_3 as input, output ciphertext C or $\perp$.

So far, the above *KH-PKE* scheme does not realize the homomorphic attribute, but the homomorphic function is reflected in the correctness defined below. Let *gpk* be the public key generated by *Gen* algorithm, $C_{gpk,m}$ is all ciphertext sets of $m \in M$ under the public key *gpk*,that is, $C_{gpk,m} = \{C | \exists\, r \in \{0, 1\}^* \, s.t. C = Enc(gpk, m; r)\}$.

Definition 2 (correctness). For all $(gpk, gsk, sk_h) \leftarrow Gen(1^n)$, the following two conditions are met: (1) For all $m \in M$, and all $C \in C_{gpk,m}, Dec(gsk, C) = M$. (2) For all $m_1, m_2, m_3 \in M$, all $C_1 \in C_{gpk,m_1}, C_2 \in C_{gpk,m_2}, C_3 \in C_{gpk,m_3}, Eval(sk_h, C_1, C_2, C_3) \in C_{gpk,m_1 \odot m_2 \odot m_3}$.

Definition 3 (*KH-CCA* Security). a *KH-PKE* scheme is considered to meet *KH-CCA* security if for any *PPT* adversary *A*:

$$Adv_{KH-PKE}^{KH-CCA}(n) = \left| Pr \left[\begin{array}{c} (\mathrm{gpk}, \mathrm{gsk}, \mathrm{sk}_h) \leftarrow \mathrm{Gen}\,(1^n)\,; \\ \left(m_0^*, m_1^*, State\right) \leftarrow A^O\,(find, \mathrm{gpk})\,; \\ \beta \leftarrow_{\$} \{0, 1\}\,; \\ C^* \leftarrow Enc\left(\mathrm{gpk}, m_\beta^*\right); \\ \beta' \leftarrow A^O\,(guess, State, C^*)\,; \\ \beta = \beta' \end{array} \right] - \frac{1}{2} \right| \tag{1}$$



- **Setup algorithm:** algorithm setting generates public parameter list of system:
  Input: security parameter n
  Output: the system public parameter PP is also used as the public reference string of *NIZK* scheme $crs \leftarrow NIZK.K(1^n)$
- **Partylnitial algorithm:** For example, for the needs of scientific research, medical institution *A* uses the algorithm Partylnitial to aggregate the medical data by *KH-PKE* scheme into data information after calculating the local model parameters under *FL*:
  Input:

$$PP = (p, G, G_T, e, g, h, g_T, gvk, \sigma, T);$$
$$td = w;$$

where $h = g^w$.

  Output: $GPK_A, GSK_A, C_A = Enc_{GPK_A}(t_A)$
  The public key GPK_A is also linked to the address used to receive data. The ciphertext C_A of the aggregated medical data t_A is stored in the medical *CB*.
- **Transfer algorithm:** The algorithm is to share the medical data t_A of the medical institution to the research institution *B*.
  Input: $PP, t, t_A, GPK_A, C_A, GPK_B$
  A generates the data sharing sentence x, then runs $\pi \leftarrow NIZK.P(crs, x, w)$, and publishes (x, π) to the medical *CB*.
- **Redeem algorithm:** The redeem algorithm deployed in the *CB* for data sharing is triggered by the Transfer algorithm.
  Input: PP, x, π
  If $NIZK.V(crs, x, \pi) = 1$, show that the data sharing process between medical institutions and research institutions can be tracked from the *CB*, publish the data sharing behavior, and then update the patient data in medical institutions to C_A/C_t and the patient data in research institutions to $C_B \cdot \widehat{C}_t$;
  Otherwise, ignore the behavior.

Figure 2: Data security sharing of medical network community based on smart contract

It can be ignored in *n*, where *O* is composed of three oracle machines $Eval\ (sk_h, \cdot, \cdot)$, *RevHK* and $Dec(sk, \cdot)$. Let *D* be a list and set to$D = \{C^*\}$ after the challenge stage (*D* is set to Ø in the search stage).

- Evaluate oracle $Eval\ (sk_h, \cdot, \cdot)$: if *RevHK* has been queried before, the oracle is unavailable. Otherwise, the oracle responds to the query (C_1, C_2, C_3) with the result of $C \leftarrow Eval(sk_h, C_1, C_2, C_3)$. In addition, if $C \neq \perp$ and $C_1 \in D, C_2 \in D$ or $C_3 \in D$, the oracle updates the list through $D \leftarrow D \cup \{C\}$.
- Homomorphic key disclosures oracle*RevHK*: according to the request, the oracle machine uses sk_h responses. (this oracle is only available once.)
- Decryption prophecy $Dec(gsk, \cdot)$: if *A* queries *RevHK* and *A* obtains the challenge ciphertext C^*, the oracle is unavailable. Otherwise, the oracle will respond to the query. If $C \notin D$, the result of *C* is $Dec(gsk, \cdot)$, otherwise $\perp$ is returned.
- Lemma 1. Under the *DLIN* assumption , the above *KH-PKE* scheme meets the *IND-CCA1* security.

Proof. Suppose that the effective adversary *A* destroys the *KH-KHE* scheme in the sense of *KH-CCA* with a non-negligible probability $poly(n)$ and gives a tuple $u, v, g, s_1 = u^r, s_2 = v^s, s_3$. Decide whether to $s_3 = g^{r+s}$, we can construct a reduction algorithm *B* attacking *KH-CCA* security to break the *DLIN* assumption, as follows:

$$\text{Algorithm } B^A\ (u, v, g, s_1, s_2, s_3):$$
$$\text{Select } h \leftarrow_{\$} G, \text{ set } gpk = (u, v, g, h);$$
$$B \leftarrow gpk; RevHK \leftarrow sk_h;$$
$$\text{Input } (gpk, sk_h), \text{run } A;$$

When A sends ciphertext C as the decryption query, B forwards C as the decryption query of B.
$\left(m_0^*, m_1^*\right) \leftarrow A^O\left(find, gpk\right)$;

$$\text{Sampling } \beta \leftarrow_{\$} \{0,1\}, \text{ and then set } C^* = \left(s_1, s_2, s_3 \cdot h^{m_\beta}\right);$$
$$\beta' \leftarrow A^O\left(guess, State, C^*\right);$$

$$\text{If } \beta = \beta', \text{ returns } 1;$$
$$\text{Otherwise, returns } 0.$$

- If $s_3 = g^{r+s}$, then the probability of $B = 1$ is the probability of A correctly guessing the hidden bit, $poly(n) + \frac{1}{2}$.
- If s_3 is the random element in G, then $s_3 \cdot h^{m_b}$ is evenly distributed in G, irrelevant to β, so the probability of A's correct answer is $\frac{1}{2}$.

Therefore, the probability of B distinguishes the distributions $\{u, v, g, u^r, v^s, g^{r+s}\}$ and $\{u, v, g, u^r, v^s, \rho\}$ is equal to $poly(n)$. This is a non-negligible probability, which contradicts the *DLIN* assumption. This means that if the scheme meets *KH-CCA* security, then the scheme meets *IND-CCA1* security.

In the medical network community environment, assuming that A and B want to share medical data, they will publish a medical data sharing information on the medical *CB*, which is basically written as follows. The medical data ciphertext t of medical institution A is shared with research institution B, then σ is t's signature. Then, the consortium chain first verifies whether the signature is correct, that is, whether there is medical data t in A. If so, the data sharing operation will be carried out and the behavior will be published on the *CB*, otherwise the behavior will be ignored. as shown in Fig. 2.

In the above data sharing process, we can all know the medical data t shared from A to B (that is, the privacy of the shared medical data t is not guaranteed). Therefore, in Section 5, this paper introduces a data privacy protection scheme based on *FL* and *CB*, which is used to prove the security of medical data t under data sharing operation.

5 Data Privacy Protection Mechanism Based on Federated Learning and Blockchain

This section constructs an efficient *NIZK* scheme. The scheme first constructs a Σ-Protocols [23], and then use the Fiat-Shamir heuristic method to construct the *NIZK* protocol in the *FL* environment. Finally, the *NIZK* scheme for complete lightweight devices is obtained by using the set member proof protocol. At the same time, *CCA* anonymous dynamic group signature mechanism is introduced, which greatly improves the time efficiency of generating proof. The signature mechanism is more secure and efficient in computing and communication cost under the data security management framework of medical network community [24], based on *CB* and *FL*. At the same time, the *NIZK* scheme is used to realize the security sharing of medical data, and verify the correctness of medical data under the sharing operation on the smart contract. Below, we will construct the design details of a zero-knowledge proof scheme suitable for *FL*.

5.1 Preliminary Preparation

NIZK parameters have certifier A and verifier B. Suppose the plaintext space is$(0, 2^{\mathfrak{L}})$, where $\mathfrak{L} = u \times l$. First introduce the *CCA* anonymous dynamic group signature mechanism, and set the algorithms Setup and Partylnitial respectively:

Setup. Generate a bilinear group [25] $(p, G, G_T, e, g) \leftarrow G_{bp}(1^n)$. Randomly select $h \leftarrow_\$ G$. From Z_p^* randomly select $\gamma, \alpha_1, \alpha_2$, calculate $\Omega = g^\gamma, g_1 = g^{\alpha 1}, g_2 = g^{\alpha 2}$. Set $g_T = e(g, g)$, where $\vec{g}_1 = (g_1, 1, g) \in G^3$, $\vec{g}_2 = (1, g_2, g) \in G^3$ and $\vec{g}_3 = \left(g_{3,1}, g_{3,2}, g_{3,3}\right) \in \vec{g}_1 \circ \vec{g}_2 \in G^3$, where $\xi_1,\xi_2 \in_R Z_n^*$; select $(u_1, u_2, w) \in G^3$, a hash function $H\colon \{0, 1\}^* \rightarrow \{0, 1\}^n$, then$\{cpk, ik, ok\} : = n, G, G_T, e, h, g, g_1, g_2, \vec{g}_1, \vec{g}_2, \vec{g}_3, \Omega, u_1, u_2, w, H, \gamma, (\alpha_1, \alpha_2)$.

That is, $Setup\,(1^n) \rightarrow (cpk, ik, ok)$. Where gpk is the group signature public key, ik and $ok$are the issuer's and initiator's keys respectively, and gpk is the group signature public key, ik and ok are the issuer's and initiator's keys respectively.

Run $(gsk = x, gvk = g^x)$ to generate group M membership certificate K_1 and K_2. If $K_1, \Omega, K_2 = e\,(g, h)\, e\,(g, gpk)$, group signature private key $csk\colon = (gsk, gpk, K_1, K_2)$. Message m signature value $\theta_4 = g^{1/(x+H(m))}$, where x is its private key gsk;$r_i, s_i, t_i \in Z_p^*$ where $i = 1, \cdots, 4$. At the same time, calculate the commitment values $\vec{c}_i = \iota\,(\theta_i) \circ \vec{g}_1^{r_i} \circ \vec{g}_2^{s_i} \circ \vec{g}_3^{t_i}$ and $\vec{\pi}_1, \vec{\pi}_2$. Select $z_1, z_2 \in_R Z_p^*$, using CCApublic key encryption algorithm to encrypt θ_1. Each equation has $NIZK\ \phi$, expressed by $\vec{\phi}_1, \vec{\phi}_2$, $\vec{\phi}_3$ separately. At the same time, calculate the commitment value $\vec{d}_1, \vec{d}_2, \vec{d}_3$ of $\left(\bar{r}, \bar{s}, \bar{t}\right)$. Finally, the hash value $\tilde{H}$ is calculated and the encrypted value v_1, v_2 is generated. Thus, we can get the group signature value:

$$\sigma = \left(\vec{c}_1, \vec{c}_2, \vec{c}_3, \vec{c}_4, \vec{d}_1, \vec{d}_2, \vec{d}_3, e_1, e_2, e_3, \vec{\pi}_1, \vec{\pi}_2, \vec{\phi}_1, \vec{\phi}_2, \vec{\phi}_3, \vartheta, v_1, v_2\right) \tag{2}$$

That is $GSig\,(gpk, gsk\,[i]\,, m) \rightarrow \sigma$.

Therefore, the integer signature value $(\sigma_0, \sigma_1, \cdots, \sigma_{2^u-1})$ between 0 and $2^u - 1$ can be obtained in the above way and the following bilinear mapping:

$$T = (T_0, T_1, \cdots, T_{2^u-1}) = (e\,(\sigma_0, g)\,, e\,(\sigma_1, g)\,, \cdots, e\,(\sigma_{2^u-1}, g)) \tag{3}$$

After receiving the group signature, the verifier returns $gvk = GVf\,(cpk, m, \sigma)$. Finally, output the common parameter$PP = (p, G, G_T, e, g, h, g_T, gvk, \sigma, T)$.

- **PartyInitial.** For example, when the medical data t_A of a medical institution needs to be shared with the data of a research institution, the system is triggered. First, we use the key homomorphism technique described in Definition 1 to initialize the patient data information to be shared:
- Private key: $gsk_A = (x_{A_1}, x_{A_2}) \in Z_p^2$;
- Public key: $gpk_A = (X_{A_1}, X_{A_2}) \in G^2$, where $X_{A_1} = g^{x_{A1}}, X_{A2} = g^{x_{A2}}$;
- Encrypted data: $C_A = Enc_{gpk_A}\,(t_A; (y_1, y_2)) = \left(C_1 = X_{A_1}^{y1}, C_2 = X_{A_2}^{y2}, C_3 = g_1^{y_1+y_2} \cdot h^{t_A}\right)$

According to the rules of sharing smart contract of medical data security, in the data sharing scenario, we explain how to construct a data sharing sentence x in which medical institutions send t patients' data to research institutions. Firstly, the certifier obtains the ciphertext of t_A, $\tilde{C} = (\tilde{C}_1, \tilde{C}_2, \tilde{C}_3) = (X_{A_1}^{\tilde{y}1}, X_{A_2}^{\tilde{y}2}, g^{\tilde{y}_1+\tilde{y}_2} \cdot h^{t_A})$, from the medical CB, and decrypts it to obtain t_A. Using randomness $y_1, y_2 \leftarrow_\$ Z_p$, the prover encrypts the shared data t with its public key and the verifier:

$$C = (C_1, C_2, C_3) = \left(X_{A_1}^{y_1}, X_{A_2}^{y_2}, g^{y_1+y_2} \cdot h^t\right); \widehat{C} = (\widehat{C}_1, \widehat{C}_2, \widehat{C}_3 = C_3) = (X_{B_1}^{y_1}, X_{B_2}^{y_2}, g^{y_1+y_2} \cdot h^t) \tag{4}$$

5.2 NIZK Scheme

Assume that P and V are the prover and verifier, respectively. Fiat-Shamir heuristic [26] converts the Σ-protocol into $NIZK$ parameters: After calculating a from P, the challenge $c = H(a)$ is obtained. Among them, H is a random oracle. Finally, calculate z and send the proof (a, c, z) to V.

The proof language L defining the P is as follows:

The data sharing sentence $x = \left(C, \widehat{C}, gpk_A, gpk_B, \tilde{C}\right) \in \mathfrak{L}$ and the corresponding verifier $w = (gsk_A = \left(x_{A_1}, x_{A_2}\right), y_1, y_2, t_A, t)$, and then

A) $C_i = X_i^{y_i}, for\ i = 1, 2;$
B) $\widehat{C}_i = X_{B_i}^{y_i}, for\ i = 1, 2;$
C) $C_3 = g^{y_1+y_2} \cdot h^t;$
D) $\frac{\tilde{C}_3}{C_3} = \tilde{C}_1^{\frac{1}{x_{A_1}}} \cdot \tilde{C}_2^{\frac{1}{x_{A_2}}} \cdot g^{-y_1-y_2} \cdot h^{t_A-t};$
E) $t \in \left[0, 2^{\mathfrak{L}}\right), t' = t_A - t \in \left[0, 2^{\mathfrak{L}}\right), where\ t = \sum_{j=0}^{l-1} t_j \cdot (2^u)^j, t' = \sum_{j=0}^{l-1} t'_j \cdot (2^u)^j, 0 \le t_j, t'_j < 2^u;$ *or* there exists $\omega \in Z_p$, and then
F) $h = h_1^w.$

Proof generation by P.

Certifier A takes PP as the public input and generates the $NIZK$ proof of the above statement using private input $(gsk_A, y_1, y_2, t_A, t)$, as follows:

Proof of formula (A-F) through Σ-protocols. Formula (E) can be proved by the value range in [27,28]. Randomly select samples $r_1, r_2, \mathfrak{l}, k \leftarrow_\$ Z_p$. Calculate $i \in \{1, 2\}$: $R_i = X_{A_i}^{r_i}; \widehat{R}_i = X_{B_i}^{r_i}$; For $j \in [0, l)$, randomly select samples $v_j, v'_j, s_j, w_j, q_j, m_j \leftarrow_\$ Z_p$, and then calculate: $V_j = \sigma_{t_j}^{v_j}, V'_j = \sigma_{t'_j}^{v'_j}; D_1 = \prod_{j=0}^{l-1}\left(h^{(2^u)^j \cdot s_j}\right) \cdot g_1^{r_1+r_2}; D_2 = \prod_{j=0}^{l-1}\left(h^{(2^u)^j \cdot w_j}\right) \cdot \tilde{C}_1^{\mathfrak{l}} \cdot \tilde{C}_2^k \cdot g_1^{-r_1-r_2}; a_j = T_{t_j}^{-s_j \cdot v_j} \cdot g_T^{q_j}, a'_j = T_{t'_j}^{-w_j \cdot v'_j} \cdot g_T^{m_j}$; Randomly select $\widehat{c} \leftarrow_\$ Z_p, \widehat{z} \leftarrow_\$ Z_p$, set $\alpha = g^{\widehat{z}}/h^{\widehat{c}}$, and use α to represent $(R_1, R_2, \widehat{R}_1, \widehat{R}_2, \left\{V_j, V'_j\right\}_{j=0}^{l-1}, D_1, D_2, \left\{a_j, a'_j\right\}_{j=0}^{l-1}, \alpha)$. The challenge is obtained through calculation: $\tilde{c} = H(a); c = \tilde{c} + \widehat{c}$; After that, the secure hash function instance H. Calculation (on modulus p):

$$z_1 = r_1 - c \cdot y_1; z_2 = r_2 - c \cdot y_2; \tag{5}$$

$$z_{v_j} = q_j - c \cdot v_j; z_{v'_j} = m_j - c \cdot v'_j; \tag{6}$$

$$z_{t_j} = s_j - c \cdot s_j; z_{t'_j} = w_j - c \cdot t'_j; \tag{7}$$

$$z_{\mathfrak{l}} = \mathfrak{l} - \frac{c}{x_{A_1}}; z_k = k - \frac{c}{x_{A_2}}; \tag{8}$$

Finally, A sends proof $\pi = (a, c, z)$ to B, where $z = (z_1, z_2, \left\{z_{v_j}, z_{v'_j}\right\}_{j=0}^{l-1}, \left\{z_{t_j}, z_{t'_j}\right\}_{j=0}^{l-1}, z_{\mathfrak{l}}, z_k, \widehat{z})$.

The verifier V calculates c after receiving the proof π. Using the public input $PP, \forall i = 1, 2; j \in [0, l)$, the verifier checks the following conditions:

$$R_i = C_i^c \cdot X_{A_i}^{z_i}; \tag{9}$$

$$\widehat{R}_i = \widehat{C}_i^c \cdot X_{B_i}^{z_i}; \tag{10}$$

$$D_1 = \prod_{j=0}^{l-1}(h^{(2^u)^j \cdot z_{t_j}}) \cdot C_3^c \cdot g^{z_1+z_2}; \tag{11}$$

$$D_2 = \prod_{j=0}^{l-1}(h^{(2^u)^{j \cdot z_{t'_j}}}) \cdot \left(\frac{\tilde{C}_3}{C_3}\right)^c \cdot g_1^{z_1+z_2}; \tag{12}$$

$$a_j = e\left(V_j, gvk\right)^c \cdot e\left(V_j, g\right)^{-z_{t'_j}} \cdot g_T^{z_{v'_j}}; \tag{13}$$

$$a'_j = e\left(V'_j, gvk\right)^c \cdot e\left(V'_j, g\right)^{-z_{t'_j}} \cdot g_T^{z_{v'_j}}; \tag{14}$$

$$g^{\widehat{z}} = \alpha \cdot h^{\widehat{c}}; \tag{15}$$

5.3 Security of NIZK Scheme

Theorem 1. Assuming *DLIN*, q-Strong Diffie-Hellman(*q-SDH*) assumptions, Σ-protocols is a *NIZK* demonstration with perfect completeness, perfect zero knowledge and computational reliability in oracle model. In addition, complete zero knowledge is established in the standard common random stringmodel.

Proof. The following describes our proof conclusion.

Perfect Completeness. This property can be directly verified.

Soundness. It is assumed that the *CCA* anonymous dynamic group signature based on *q-SDH* assumptions is unforgeable, and its reliability is proved under the random oracle model. If the PUSH-PULL traffic(*PPT*) certifierP^* is an invalid sentence, generate the acceptance parameter $\pi = (a, c, z)$, where $a = (R_1, R_2, \widehat{R}_1, \widehat{R}_2, \left\{V_j, V'_j\right\}_{j=0}^{l-1}, D_1, D_2, \left\{a_j, a'_j\right\}_{j=0}^{l-1}, \alpha)$ and $z = (z_1, z_2, \left\{z_{v_j}, z_{v'_j}\right\}_{j=0}^{l-1}, \left\{z_{t_j}, z_{t'_j}\right\}_{j=0}^{l-1}, z_{\mathfrak{l}}, z_k, \widehat{z})$.

Then, construct the extractor Ext: P^* back to return $c = H(a)$. Modify the random oracle so that $c' = H(a)$ and $c \neq c'$. Use $c' = H(a)$ to execute P^*. Finally, another valid parameter can appear:

$$\pi' = \left(a, c', z' = \left(z'_1, z'_2, \left\{z'_{v_j}, z'_{v'_j}\right\}_{j=0}^{l-1}, \left\{z'_{t_j}, z'_{t'_j}\right\}_{j=0}^{2}, z'_{\mathfrak{l}}, z'_k, \widehat{z\prime}\right)\right). \tag{16}$$

Verifier information can be extracted by calculation (for $i = 0, 1; j \in [0, l)$):

$$y_i = \frac{z_i - z'_i}{c' - c}, t_j = \frac{z_{t_j} - z'_{t_j}}{c' - c}, t'_j = \frac{z_{t'_j} - z'_{t'_j}}{c' - c}, x_{A_1} = \frac{c' - c}{z_{\mathfrak{l}} - z'_{\mathfrak{l}}}, x_{A_2} = \frac{c' - c}{Z_k - z'_k}. \tag{17}$$

Under the condition of extracting the prover, if $t \notin [0, 2^{\mathfrak{L}})$ or $t' \notin [0, 2^{\mathfrak{L}})$, you can crack the *CCA* anonymous dynamic group signature weak selection message attack model with P^* subroutine.

Perfect zero knowledge. Construct a simulated *Sim* to prove the proposition $h = g^w$, thereby proving perfect zero-knowledge. See Fig. 3.

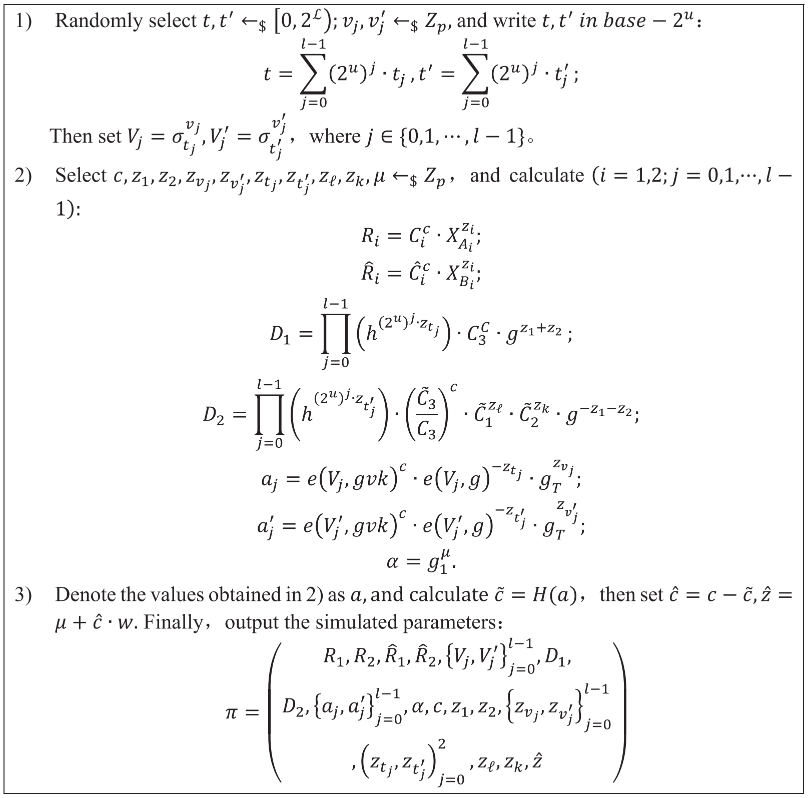
The parameters are divided into 3 parts:

$$\pi = \left(a = \left(R_1, R_2, \widehat{R}_1, \widehat{R}_2, \left\{V_j, V'_j\right\}_{j=0}^{l-1}, D_1, D_2, \left\{a_j, a'_j\right\}_{j=0}^{l-1}, \alpha\right),\right.$$

$$\left. c, z = \left(z_1, z_2, \left\{z_{v_j}, z_{v'_j}\right\}_{j=0}^{l-1}, \left\{z_{t_j}, z_{t'_j}\right\}_{j=0}^{l-1}, z_{\mathfrak{l}}, z_k, \widehat{z}\right)\right). \tag{18}$$

For clarity and convenience, we express the simulation parameters as

$$\pi = \begin{pmatrix} \mathfrak{a} = \left(\mathfrak{R}_1, \mathfrak{R}_2, \widehat{\mathfrak{R}}_1, \widehat{\mathfrak{R}}_2, \left\{\mathfrak{V}_j, \mathfrak{V}'_j\right\}_{j=0}^{l-1}, \mathfrak{D}_1, \mathfrak{D}_2, \left\{\mathfrak{a}_j, \mathfrak{a}'_j\right\}_{j=0}^{l-1}, \alpha\right), \mathfrak{c}, \\ \mathfrak{z} = \left(\mathfrak{z}_1, \mathfrak{z}_2, \left\{\mathfrak{z}_{v_j}, \mathfrak{z}_{v'_j}\right\}_{j=0}^{l-1}, \left\{\mathfrak{z}_{t_j}, \mathfrak{z}_{t'_j}\right\}_{j=0}^{l-1}, \mathfrak{z}_{\mathfrak{l}}, \mathfrak{z}_k, \widehat{\mathfrak{z}}\right) \end{pmatrix}. \tag{19}$$



1) Randomly select $t, t' \leftarrow_{\$} [0, 2^{\mathcal{L}}); v_j, v_j' \leftarrow_{\$} Z_p$, and write $t, t'\ in\ base - 2^u$:

$$t = \sum_{j=0}^{l-1} (2^u)^j \cdot t_j\ , t' = \sum_{j=0}^{l-1} (2^u)^j \cdot t_j'\ ;$$

Then set $V_j = \sigma_{t_j}^{v_j}, V_j' = \sigma_{t_j'}^{v_j'}$, where $j \in \{0,1, \cdots, l-1\}$。

2) Select $c, z_1, z_2, z_{v_j}, z_{v_j'}, z_{t_j}, z_{t_j'}, z_\ell, z_k, \mu \leftarrow_{\$} Z_p$, and calculate $(i = 1,2; j = 0,1,\cdots, l-1)$:

$$R_i = C_i^c \cdot X_{A_i}^{z_i};$$

$$\hat{R}_i = \hat{C}_i^c \cdot X_{B_i}^{z_i};$$

$$D_1 = \prod_{j=0}^{l-1} \left(h^{(2^u)^j \cdot z_{t_j}}\right) \cdot C_3^C \cdot g^{z_1 + z_2}\ ;$$

$$D_2 = \prod_{j=0}^{l-1} \left(h^{(2^u)^j \cdot z_{t_j'}}\right) \cdot \left(\frac{\tilde{C}_3}{C_3}\right)^c \cdot \tilde{C}_1^{z_\ell} \cdot \tilde{C}_2^{z_k} \cdot g^{-z_1 - z_2};$$

$$a_j = e(V_j, gvk)^c \cdot e(V_j, g)^{-z_{t_j}} \cdot g_T^{z_{v_j}};$$

$$a_j' = e(V_j', gvk)^c \cdot e(V_j', g)^{-z_{t_j'}} \cdot g_T^{z_{v_j'}};$$

$$\alpha = g_1^{\mu}.$$

3) Denote the values obtained in 2) as a, and calculate $\tilde{c} = H(a)$, then set $\hat{c} = c - \tilde{c}, \hat{z} = \mu + \hat{c} \cdot w$. Finally, output the simulated parameters:

$$\pi = \begin{pmatrix} R_1, R_2, \hat{R}_1, \hat{R}_2, \{V_j, V_j'\}_{j=0}^{l-1}, D_1, \\ D_2, \{a_j, a_j'\}_{j=0}^{l-1}, \alpha, c, z_1, z_2, \left\{z_{v_j}, z_{v_j'}\right\}_{j=0}^{l-1} \\ , \left(z_{t_j}, z_{t_j'}\right)_{j=0}^{2}, z_\ell, z_k, \hat{z} \end{pmatrix}$$

Figure 3: Simulator for new *NIZK* parameters

$\widehat{c} \leftarrow_{\$} Z_p$ is observed to be independent of a, $c = H(a) + \widehat{c}$ is uniformly distributed in Z_p, and c is also randomly selected from Z_p in the simulation. Therefore, the distribution {c} is the same as $\{\mathfrak{c}\}$:$\{c\} \equiv \{\mathfrak{c}\}$.

Set $\mathfrak{f} = \{c\} = \{\mathfrak{c}\}$. Under the condition of: $\{c\} \equiv \{\mathfrak{c}\}$, $\overline{c} \in \mathfrak{f}$ is given for each $\rho \in Z_p$ because $\widehat{z}, r_1, r_2, \mathfrak{l}, k, q_j, m_j, s_j, w_j, v_j, v_j' \leftarrow_{\$} Z_p$, where $j \in [0, l)$, and for any element $\tilde{z}$ in z, we have

$$\Pr[\tilde{z} = \rho | \mathfrak{c} = \overline{c}] = \frac{1}{p} \tag{20}$$

In the argument of simulation, under the same conditions, the given value $\mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_{\mathfrak{l}}, \mathfrak{z}_k, \mathfrak{z}_{v_j}, \mathfrak{z}_{v_j'}, \mathfrak{z}_{t_j}, \mathfrak{z}_{t_j'}$, $\mu \leftarrow_{\$} Z_p$. For all elements $\tilde{z}$ in z, we have

$$Pr[\tilde{\mathfrak{z}} = \rho | \mathfrak{c} = \overline{c}] = \frac{1}{p}. \tag{21}$$

The Fiat-Shamir heuristic algorithm is applied to the Σ-protocols to build *NIZK*, so as to achieve perfect zero-knowledge. At the same time, the reliability of the construction is proven in the model.

Set the set of $\mathbb{J} = z^1, z^2, \left\{z_j^3, z_j^4\right\}_{j=0}^{l-1}, \left\{z_j^5, z_j^6\right\}_{j=0}^{l-1}, z^7, z^8, z^9$: $z_i \leftarrow_{\$} Z_p, i \in [10]$. Given $\overline{c} \leftarrow_{\$} \mathfrak{f}$, for each $\overline{z} \in \mathbb{J}$,

$$\Pr[z = \overline{z} | c = \overline{c}] = \Pr[\mathfrak{z} = \overline{z} | \mathfrak{c} = \overline{c}]. \tag{22}$$

Under the condition of $Pr\left[z=\overline{z}|c=\overline{c}\right]=Pr\left[\mathfrak{z}=\overline{z}|\mathfrak{c}=\overline{c}\right]$, given $\overline{c}\in\mathfrak{C},\overline{z}\in\mathbb{J}$, from the verification strategy, message $R_1, R_2, D_1, D_2, a_j, a'_j, \alpha\ in\ \pi$ is deterministic, where $j\in[0,l)$. For $\left\{V_j, V'_j\right\}$, we have

$$Pr\left[V_j=\mathfrak{g}|c=\overline{c},z=\overline{z}\right]=Pr\left[\sigma_{t_j}^{v_j}=\mathfrak{g}|c=\overline{c},z=\overline{z}\right]=\frac{1}{p}. \tag{23}$$

$$Pr\left[V'_j=\mathfrak{g}|c=\overline{c},z=\overline{z}\right]=Pr\left[\sigma_{t'_j}^{v'_j}=\mathfrak{g}|c=\overline{c},z=\overline{z}\right]=\frac{1}{p}. \tag{24}$$

where $\mathfrak{g}\leftarrow_{\$} G$, start from $v_j, v'_j\leftarrow_{\$} Z_p$.

Set $A = \left\{a^1,a^2,\left\{a_j^3,a_j^4\right\}_{j=0}^{l-1},a^5,a^6,\left\{a_j^7,a_j^8\right\}_{j=0}^{l-1},a^9: a^1,a^2,a^3,a^4,a^5,a_j^6,a_j^7,a^9\leftarrow_{\$} G,a_j^7,a_j^8\leftarrow_{\$} G_T\right\}$. Therefore, given $\overline{c}\in\mathfrak{C},\overline{z}\in\mathbb{J}$, for any $\overline{a}\in A$,

$$Pr\left[a=\overline{a}|c=\overline{c},z=\overline{z}\right]=Pr\left[\mathfrak{a}=\overline{a}|\mathfrak{c}=\overline{c},\mathfrak{z}=\overline{z}\right]. \tag{25}$$

Combining (22) and (25), We conclude that for any inconsistent PPT adversary $A=(A_1,A_2)$,

$$Pr\left[\begin{array}{c}(x,w)\leftarrow A_1\left(1^n\right)(a,c,z)\leftarrow P\left(x,w,PP\right)\\ :\ (x,w)\in RA_2\left(a,c,z\right)=1\end{array}\right]=Pr\left[\begin{array}{c}(x,w)\leftarrow A_1\left(1^n\right)(x,c,z)\leftarrow Sim\left(x\right):\\ (x,w)\in RA_2\left(a,c,z\right)=1\end{array}\right] \tag{26}$$

(perfect) get zero knowledge property. The next step is to verify the correctness of $NIZK$ scheme. Instead of verifying

$$a_j=e\left(V_j,gvk\right)^c\cdot e\left(V_j,g\right)^{-z_{t'_j}}\cdot g_T^{z_{v'_j}}; \tag{27}$$

$$a'_j=e\left(V'_j,gvk\right)^c\cdot e\left(V'_j,g\right)^{-z_{t'_j}}\cdot g_T^{z_{v'_j}}; \tag{28}$$

by calculating $4l$ pairing calculation. V from Z_p, randomly select $2l$ elements $\left\{d_j\right\}_{j=0}^{l-1},\left\{d'_j\right\}_{j=0}^{l-1}$, and check whether the following formula is valid:

$$\prod_{j=0}^{l-1}a_j^{d_j}\cdot\prod_{j=0}^{l-1}(a'_j)^{d'_j}=e\left(\prod_{j=0}^{l-1}V_j^{cd_j}\cdot\prod_{j=0}^{l-1}\left(V'_j\right)^{cd'_j},gvk\right)\cdot e\left(\prod_{j=0}^{l-1}V_j^{-z_{t_j}d_j}\cdot\prod_{j=0}^{l-1}\left(V'_j\right)^{-z_{t'_j}d'_j},g_2\right)\cdot g_T^{\sum_{j=0}^{l-1}z_{v_j}d_j+\sum_{j=0}^{l-1}z_{v'_j}d'_j}. \tag{29}$$

In the formula above, only two pairing calculations are calculated, which is more efficient than ((1),(2)). Meanwhile, ((1),(2))⇒ (29): after replacing all values $\left\{a_j\right\}_{j=0}^{l-1},\left\{a'_j\right\}_{j=0}^{l-1}$ in ((1),(2)) , Eq. (29) is obtained.

(29)⇒ ((1),(2)): Consider formula (29):

$$Right\_Side=\prod_{j=0}^{l-1}\left(\left(e\left(V_j,gvk\right)^c\cdot e\left(V_j,g\right)^{-z_{t_j}}\cdot g_T^{z_{v_j}}\right)^{d_j}\cdot\left(e\left(V'_j,gvk\right)^c\cdot e\left(V'_j,g\right)^{-z_{t'_j}}\cdot g_T^{z_{v'_j}}\right)^{d'_j}\right); \tag{30}$$

$$Lift_{Side}=\prod_{j=0}^{l-1}\left(\left(a_j\right)^{d_j}\left(a'_j\right)^{d'_j}\right). \tag{31}$$

If $Left\_Side=Right\_Side$, there are two situations:

1) $\forall j\in[0,l)\,,a_j=e\left(V_j,gvk\right)^c\cdot e\left(V_j,g\right)^{-z_{t_j}}\cdot g_T^{z_{v_j}},a'_j=e\left(V'_j,gvk\right)^c\cdot e(V'_j,g_2)^{-z_{t'_j}}\cdot g_T^{z_{v'_j}}$. This implies the correctness of equation ((1),(2)).

2) There are some d_j *or* $d'_j = 0$, which can lead to $a_j \neq e\left(V_j, gvk\right)^c \cdot e(V_j, g)^{-z_{t'_j}} \cdot g_T^{z_{v_j}}$ or $a'_j \neq e\left(V'_j, gvk\right)^c \cdot e(V'_J, g)^{-z_{t'_j}} \cdot g_T^{z_{v'_j}}$ for a certain $j \in [0, l)$. This is likely to happen.

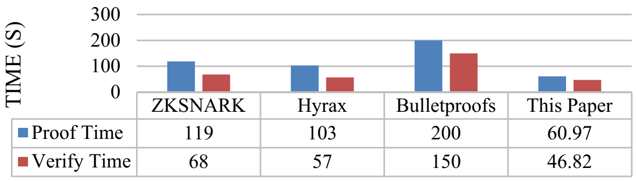
$$\sum_{i=1}^{2l}(C_{2l}^i \frac{1}{p^i}\left(1-\frac{1}{p}\right)^{2l-i}) = 1-\left(1-\frac{1}{p}\right)^{2l} < \frac{2l}{p} < \frac{1}{2^{n-1}}; \tag{32}$$

That is, the probability is negative because p is a prime number of n bits.

Overall, the overwhelming probability is $1-\frac{1}{2^{n-1}}$, equation ((1),(2)) ⇔ (29).

5.4 *Evaluation of NIZK Scheme*

In this section, the *NIZK* scheme is compared with the existing zero-knowledge succinct non-interactive argument of knowledge(*ZKSNARK*), Hyrax and Bulletproofs. The specific comparison results are shown in Fig. 4. As shown in Fig. 4, compared with other work, our *NIZK* scheme will significantly improve the running time of the prover. In the actual comparison, we can see that our scheme is more than twice as fast as the Bulletproofs scheme in proving its running time. Compared with Hyrax and *ZKSNARK*, our prover also has advantages in running time. In the running time of the verifier, our scheme also optimizes the verification time efficiency of the verifier, which is significantly different from other schemes and reduces the data communication cost.



	ZKSNARK	Hyrax	Bulletproofs	This Paper
🟦 Proof Time	119	103	200	60.97
🟥 Verify Time	68	57	150	46.82

Figure 4: Comparison between *NIZK* scheme and existing zero knowledge proof(*ZKP*) system

Subsequently, the computational complexity of *NIZK* scheme is compared with the existing *ZKSNARK*, Hyrax and Bulletproofs schemes in theory and practice. This paper uses Python programming language to run experiments on Windows (Window 7, 64-bit), and uses Inter (R) Core (TM) i7-3687CPU with 2.10 and 8 GB RAM. However, the verifiers of *ZKSNARK*, Hyrax and Bulletproofs are memory intensive, so they are evaluated on a server with 100 GB memory and 322.80 GHz memory. The specific comparison results are shown in Tab. 1. Where C is the size of the circuit with depth d and *inp* and is the size of its input.

Table 1: Complexity comparison between *NIZK* scheme and *ZKP* system

Proof size	Hyrax [29]	Bulletproofs [30]	ZKSNARK [31]	This paper
Theoretical	$O\left(dlogC + \sqrt{inp}\right)$	$O(\log C)$	$O\left(log^2 C\right)$	$O(l)$
Practical	3.2 MB	2115 B	735 KB	3260 B

5.5 *Experiments and Results*

In order to verify the effectiveness of the overall scheme, this paper cites the medical dataset of the National Institutes of Health in the United States to evaluate the scheme. Firstly, we trained 60 epochs in the ensemble environment and used the *ANN* model to evaluate the results of the ensemble model training each epoch, as shown in Fig. 5. The experimental results show the fact that the original data has an impact on the accuracy of the model and the fact that it is beneficial to the training dataset under different circumstances.

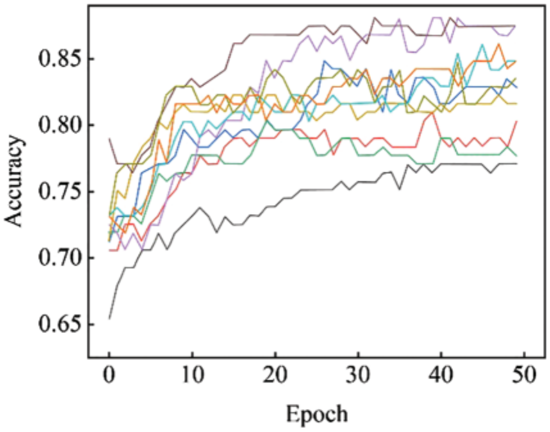


Figure 5: Experiments 10 times to verify the accuracy of medical data

Secondly, in order to test the accuracy of the medical dataset in the *CB* and *FL* environment. The 10 participating institutions were simulated to assign the dataset in a random fashion. In this environment, 3 experimental results were randomly selected. The results show that the *FL* environment (black line) is significantly lower than the accuracy of the medical dataset in the consortium chain and *FL* environment (red line). Finally, using the *ANN* model can protect the medical data of the participating parties from attacks in the context of this paper. as shown in Fig. 6.

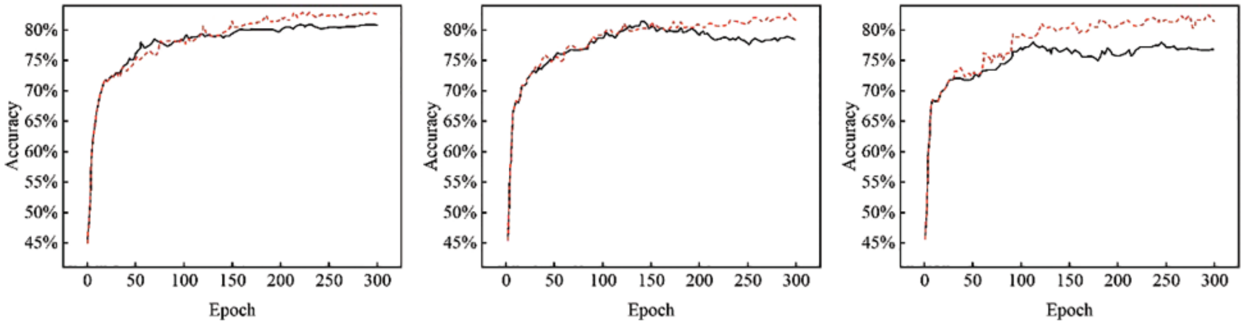


Figure 6: Results accuracy of the three participating institutions in the consortium chain and Federated learning environment

6 Conclusions

This paper presents a data management framework of medical network community based on *CB* and *FL*. An intelligent contract system based on *KH-PKE* scheme is designed to ensure the security of important data in medical network community. This paper also designs a *NIZK* scheme suitable for lightweight devices, and introduces *CCA* anonymous dynamic group signature, which greatly improves the time efficiency of generating proof and realizes the privacy protection of important data.

Funding Statement: This work is supported by the NSFC (No. 62072249). Yongjun Ren received the grant and the URLs to sponsors' websites is https://www.nsfc.gov.cn/.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] X. R. Zhang, X. Sun, X. M. Sun, W. Sun and S. K. Jha, "Robust reversible audio watermarking scheme for telemedicine and privacy protection," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3035–3050, 2022.

[2] S. Balasubramanian, V. Shukla, J. S. Sethi, N. Islam and R. Saloum, "A readiness assessment framework for blockchain adoption: A healthcare case study," *Technological Forecasting and Social Change*, vol. 165, no. 1, pp. 120536, 2021.

[3] Y. J. Ren, Y. Leng, Y. P. Cheng and J. Wang, "Secure data storage based on blockchain and coding in edge computing," *Mathematical Biosciences and Engineering*, vol. 16, no. 4, pp. 1874–1892, 2019.

[4] C. P. Ge, Z. Liu, J. Y. Xia and L. M. Fang, "Revocable identity-based broadcast proxy re-encryption for data sharing in clouds," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 3, pp. 1214–1226, 2021.

[5] T. Li, Y. Ren and J. Xia, "Blockchain queuing model with non-preemptive limited-priority," *Intelligent Automation & Soft Computing*, vol. 26, no. 5, pp. 1111–1122, 2020.

[6] X. R. Zhang, W. F. Zhang, W. Sun, X. M. Sun and S. K. Jha, "A robust 3-D medical watermarking based on wavelet transform for data protection," *Computer Systems Science & Engineering*, vol. 41, no. 3, pp. 1043–1056, 2022.

[7] C. P. Ge, W. Susilo, J. Baek, Z. Liu, J. Y. Xia *et al.,* "A verifiable and fair attribute-based proxy re-encryption scheme for data sharing in clouds," *IEEE Transactions on Dependable and Secure Computing*, vol. 21, no. 7, pp. 1–12, 2021.

[8] Y. Ren, K. Zhu, Y. Q. Gao, J. Y. Xia, S. Zhou *et al.,* "Long-term preservation of electronic record based on digital continuity in smart cities," *Computers, Materials & Continua*, vol. 66, no. 3, pp. 3271–3287, 2021.

[9] J. Wang, H. Han, H. Li, S. He, P. K. Sharma *et al.,* "Multiple strategies differential privacy on sparse tensor factorization for network traffic analysis in 5G," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1939–1948, 2022.

[10] Y. J. Ren, Y. Leng, J. Qi, K. S. Pradip, J. Wang *et al.,* "Multiple cloud storage mechanism based on blockchain in smart homes," *Future Generation Computer Systems*, vol. 115, no. 3, pp. 304–313, 2021.

[11] C. P. Ge, W. Susilo, J. Baek, Z. Liu, J. Y. Xia *et al.,* "Revocable attribute-based encryption with data integrity in clouds," *IEEE Transactions on Dependable and Secure Computing*, vol. 21, no. 8, pp. 1–12, 2021.

[12] Y. Ren, F. J. Zhu, S. P. Kumar, T. Wang, J. Wang *et al.,* "Data query mechanism based on hash computing power of blockchain in internet of things," *Sensors*, vol. 20, no. 1, pp. 1–22, 2020.

[13] C. P. Ge, W. Susilo, Z. Liu, J. Y. Xia, L. M. Fang *et al.,* "Secure keyword search and data sharing mechanism for cloud computing," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 6, pp. 2787–2800, 2021.

[14] X. F. Li, Y. R. Mei, J. Gong, F. Xiang and Z. X. Sun, "A blockchain privacy protection scheme based on ring signature," *IEEE Access*, vol. 8, pp. 76765–76772, 2020.

[15] A. A. Omar, M. S. Rahman, A. Basu and S. Kiyomoto, "Medibchain: A blockchain based privacy preserving platform for healthcare data," in *Int. Conf. on Security, Privacy and Anonymity in Computation, Communication and Storage*, Canton, CAN, CHN, pp. 534–543, 2017.

[16] J. W. Liu, X. L. Li, L. Ye, H. L. Zhang, X. J. Du *et al.,* "BPDS: A blockchain based privacy-preserving data sharing for electronic medical records," in *2018 IEEE Global Communications Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, pp. 1–6, 2018.

[17] Y. Ren, J. Qi, Y. P. Liu, J. Wang and G. Kim, "Integrity verification mechanism of sensor data based on bilinear map accumulator," *ACM Transactions on Internet Technology*, vol. 21, no. 1, pp. 1–20, 2021.

[18] L. M. Fang, M. H. Li, Z. Liu, C. T. Lin, S. L. Ji *et al.,* "A secure and authenticated mobile payment protocol against off-site attack strategy," *IEEE Transactions on Dependable and Secure Computing*, vol. 21, no. 2, pp. 1–12, 2021.

[19] J. Wang, C. Y. Jin, Q. Tang, N. X. Xiong and G. Srivastava, "Intelligent ubiquitous network accessibility for wireless-powered MEC in UAV-assisted B5G," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 4, pp. 2801–2813, 2021.

[20] X. H. Yue, M. J. Sun, X. B. Wang, H. Shao and Y. He, "An efficient dynamic group signatures scheme with CCA-anonymity in standard model," in *Int. Symp. on Cyberspace Safety and Security*, Canton, CAN, CHN, pp. 205–219, 2019.

[21] Y. J. Ren, F. Zhu, J. Wang, P. Sharma and U. Ghosh, "Novel vote scheme for decision-making feedback based on blockchain in internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 1639–1648, 2022.

[22] J. S. Almeida, "Predictive non-linear modeling of complex data by artificial neural networks," *Current Opinion in Biotechnology*, vol. 13, no. 1, pp. 72–76, 2002.

[23] S. Ma, Y. Deng, D. He, J. Zhang and X. Xie, "An efficient NIZK scheme for privacy-preserving transactions over account-model blockchain," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 2, pp. 641–651, 2021.

[24] L. Ren, J. Hu, M. Li, L. Zhang and J. Xia, "Structured graded lung rehabilitation for children with mechanical ventilation," *Computer Systems Science & Engineering*, vol. 40, no. 1, pp. 139–150, 2022.

[25] D. Boneh, B. Lynn and H. Shacham, "Short signatures from the Weil pairing," *Journal of Cryptology*, vol. 17, no. 4, pp. 297–319, 2004.

[26] A. Fiat and A. Shamir, "How to prove yourself: Practical solutions to identification and signature problems," in *Conf. on the Theory and Application of Cryptographic Techniques*, Berlin, Germany, pp. 186–194, 1986.

[27] J. Camenisch, R. Chaabouni and A. Shelat, "Efficient protocols for set membership and range proofs," in *Int. Conf. on the Theory and Application of Cryptology and Information Security*, Berlin, Germany, pp. 234–252, 2008.

[28] T. Li, W. D. Xu, L. N. Wang, N. P. Li, Y. J. Ren *et al.,* "An integrated artificial neural network-based precipitation revision model," *KSII Transactions on Internet and Information Systems*, vol. 15, no. 5, pp. 1690–1707, 2021.

[29] R. S. Wahby, I. Tzialla, A. Shelat, J. Thaler, M. Walfish *et al.,* "Doubly-efficient zksnarks without trusted setup," in *2018 IEEE Symp. on Security and Privacy (SP)*, London, LON, United Kingdom, pp. 926–943, 2018.

[30] B. Bünz, J. Bootle, D. Boneh, A. Poelstra, P. Wuille *et al.,* "Bulletproofs: Short proofs for confidential transactions and more," in *2018 IEEE Symp. on Security and Privacy (SP)*, London, LON, United Kingdom, pp. 315–334, 2018.

[31] E. Ben-Sasson, I. Bentov and Y. Horesh, "Scalable, transparent, and post-quantum secure computational integrity," *IACR Cryptol. ePrint Archive*, vol. 2018, no. 46, 2018.