Privacy Protection of Secure Sharing Electronic Health Records Based on Blockchain

Yuan Wang[1]*, Lin Sun[2]
School of Economics and Trade, Anhui Business and Technology College, Hefei, 231131, China[1]
Baidu Online Network Technology, Beijing, Beijing, 100086, China[2]

***Abstract*—The secure sharing and privacy protection of medical data have become pain points for medical data management platforms. Therefore, a secure sharing electronic health record privacy protection method based on blockchain is proposed in the study, aiming to improve data security privacy and ensure absolute ownership of patients' medical data. Attribute encryption and blockchain computing are utilized to construct a data secure sharing model, and zero-knowledge proof and ElGamal encryption algorithms are introduced to further improve the construction of data privacy protection methods. Experimental verification showed that the data secure sharing method proposed in the study has more advantages in terms of production key size and time cost. Compared with other public recognition mechanisms, zero-knowledge proof reduced the average time cost of generating keys by 54.36%. The proposed data privacy protection method had an average increase of 7.73% in protection effectiveness compared to other methods. The results indicate that the data secure sharing and privacy protection methods proposed in the study can improve the overall performance and security of the system while fully ensuring the absolute ownership of patients' data. This method has positive application value in the privacy protection of medical data.**

***Keywords*—Blockchain; secure sharing; electronic health records; privacy protection; zero-knowledge proof; attribute encryption**

I. Introduction

With the continuous improvement of the national economic level, the process of medical intelligence and wireless technology is gradually improving. However, limited medical resources, uneven distribution of medical levels, and heterogeneity of system data based on different medical systems have led to the phenomenon of isolated medical data [1-2]. Meanwhile, the optimization and advancement of Internet of Things technology have led to threats to the privacy and security of data information. Issues such as hacker attacks, data information leakage, and patient privacy protection urgently need to be addressed [3]. Blockchain technology has achieved decentralization through distributed consensus, data encryption, economic incentives, and other methods, improving data privacy protection. It has been widely applied in research on data privacy protection in the Internet of Things. However, the current privacy protection methods for electronic health records still need further development and optimization. Based on this, a secure shared electronic health record privacy protection method is proposed on the basis of blockchain, aiming to improve the protection performance of medical data and enhance patients' sharing rights over their medical data. By putting patients at the center, we ensure the privacy and security of user Electronic Health Record (EHR) data while safeguarding patients' absolute rights to their own medical data. At the same time, zero knowledge proof (ZKP) was introduced in combination with ElGamal encryption algorithm to explore EHR data privacy protection.

The overall structure of the research includes six sections: Section I summarizes the research achievements and shortcomings of blockchain and medical data privacy protection at home and abroad; Section II studies and designs a privacy protection method for secure shared electronic health records based on blockchain technology; Section III conducted experiments and analysis on the proposed privacy protection method for secure shared electronic health records; Section IV summarizes the experimental results. Discussion and conclusion are given in Section V and Section VI respectively.

II. Related Works

With the continuous application and development of big data technology, how to effectively share medical data information and protect privacy has become a new focus in the current research field. Ortega Calvo and others proposed an artificial intelligence modern data platform to address the limitation of healthcare data management systems being unable to utilize the generated data. Based on big data, artificial intelligence management, and efficient data processing, different components were utilized to regulate data collection and heterogeneous data was analyzed. By constructing a security and data governance layer, the privacy and integrity of the system database were maintained [4]. Kumar et al. raised a secure and efficient data sharing framework based on blockchain and deep learning to address the issues of unreliable connection security and privacy in real-time monitoring of patients in public networks. By utilizing consensus mechanisms based on smart contracts to register and verify communication entities, and using stacked sparse mutation autoencoders for key verification, privacy protection for real-time transmission of healthcare data was improved [5]. Shuaib et al. proposed a medical data sharing system based on licensed blockchain technology (BCT) to address the limitation of BCT relying on centralized databases. By integrating BCT, threshold signatures for decentralized file systems, and using the Istanbul Byzantine consensus algorithm as key verification, the performance and security of the system were improved [6]. To improve privacy protection during medical data sharing, Liu et al. proposed combining

*Corresponding Author

federated learning with neural architecture search and developed a multi-objective convolutional interval type 2 fuzzy rough model based on neural architecture search. By combining convolutional neural networks with fuzzy rough sets, the interpretability of deep neural networks was effectively improved [7].

The development of BCT provides security and privacy protection for data transmission in the era of intelligent informatization [8]. To improve the resistance of e-government systems to malicious attacks, Elisa et al. proposed a decentralized peer-to-peer e-government system framework using BCT. By utilizing BCT to verify and store existing and new data, the information security and privacy of the system were enhanced [9]. Sharma et al. raised a distributed application that protects privacy in response to various security attacks on traditional healthcare solutions. By utilizing BCT to create and maintain healthcare integers, the security, privacy, and transparency of healthcare platforms were improved [10]. Awotunde et al. raised a network architecture based on blockchain, which combines a hybrid convolutional neural network and kernel principal component analysis, to protect the system from potential threats and ensure network traffic security. By extracting features through kernel principal component analysis and then using convolutional neural networks for classification and detection, the security, privacy, and maintainability of IoT smart cities were improved [11]. To address the data security and management issues between IoT edge nodes and massive heterogeneous devices, Zhonghua et al. proposed an IoT access control model combining BCT. By proving the workload of traditional consensus algorithms, the Proof of Work (PoW) mechanism was optimized to provide decentralized, fine-grained, and dynamic access control management for IoT environments [12]. Due to the difficulty in ensuring security and privacy in data management, information verification, and dissemination, Patil established a medical record security system based on BCT. By utilizing BCT to improve the access of medical data management systems to monitoring drugs, hospital assets, etc., the service efficiency of medical service systems was improved [13].

Based on the above, relevant experts and scholars have explored various aspects of privacy protection of medical data and the application of BCT, and have achieved good results. However, current medical data systems still have a high dependence on third-party service providers, and patients cannot have absolute ownership of their own medical data. Therefore, the study innovatively proposes a secure sharing Electronic Health Record (EHR) privacy protection method based on patient-centered blockchain, aiming to ensure the privacy and security of user EHR data while safeguarding the absolute right of patients to their own medical data. In addition, to improve the privacy and security of the system, Zero Knowledge Proof (ZKP) is introduced. The combination of ZKP and ElGamal encryption algorithm has been explored for EHR data privacy protection.

III. Methods and Materials

The study first designed an EHR Security Sharing (EHRSS) based on BCT. On this basis, the study further introduced ZKP based on blockchain for EHR Privacy Protection (EHRPP) method design.

A. Design of Secure Sharing Method Based on Blockchain

The processing and sharing of EHR data are mainly achieved by commonly used data sharing management platforms in medical systems, but during the platform sharing process, users need to upload the data to cloud storage themselves [14-15]. However, the security of this operation is extremely low, and it overly relies on third-party service providers, making it difficult for users to guarantee their absolute ownership of the uploaded data. Therefore, this study proposed an EHRSS method based on BCT. The specific architecture is shown in Fig. 1.

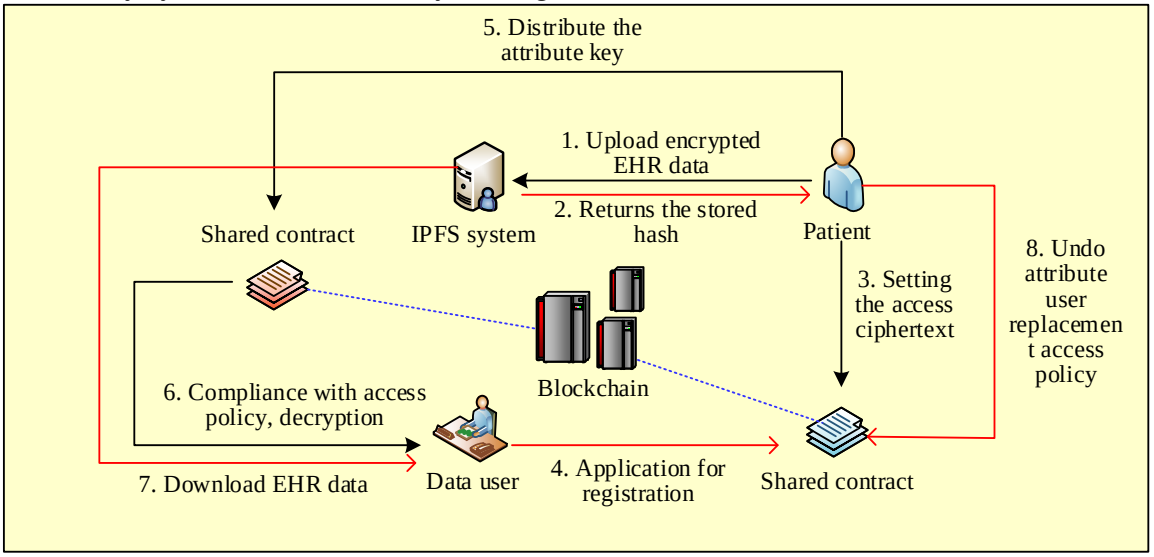


Fig. 1. Blockchain-based EHRSS model architecture.

In Fig. 1, the EHRSS model proposed in the study is mainly composed of the proprietary Inter Planetary File System (IPFS), blockchain, patients, and data users. Among them, the IPFS interstellar file system is responsible for storing the patient's EHR, and the blockchain is responsible for storing the public information and user operation records generated in the entire EHRSS model, while also considering the communication channel between patients and data users. The patient mainly refers to the owner of EHR, who creates and deploys smart contracts in the EHRSS model. The data users mainly refer to doctors, nurses, hospital administrators, and medical institutions. When the attributes of the data user comply with the strategy embedded in the ciphertext, the data sharing right is obtained based on the decryption address and key information. The execution process of the EHRSS model is denoted in Fig. 2.

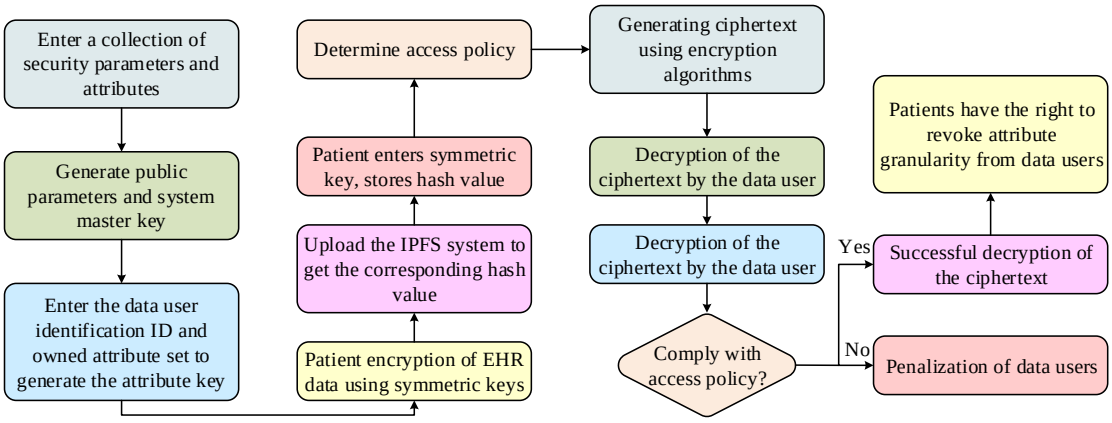


Fig. 2. Blockchain-based EHRSS model flowchart.

As shown in Fig. 2, The EHRSS model first determines the system's security parameters, attribute sets, random parameters, etc., and generates common parameters and the system's master key. The generation process is shown in Eq. (1).

$$\begin{cases} Pk = \left(N, e(g,g)^b, g, t = g^c, \{D_i = g^{d_i}, h_i = g^{\gamma_i}\}_{i \in I}\right) \\ Msk = (b, c, \{d_i\}_{i \in I}, Y) \end{cases} \quad (1)$$

In Eq. (1), Pk represents a common parameter. Msk represents the system master key. N represents the product of two prime numbers multiplied. g represents the generator. b and c represent random numbers. γ_i and h_i represent calculations related to attribute revocation. d_i and D_i represent attribute related calculations. t represents the calculation related to the identity of the data user. e represents bilinear mapping. I represents a set of attributes. Y represents the random private key of the data user. Based on the generated public parameters and system master key, the data user inputs their unique Identity Document (ID) and the corresponding attribute set, in order to obtain the exclusive attribute key for the data user. The specific calculation method is shown in Eq. (2).

$$\begin{cases} K_{i,1} = g^{b\gamma + d_i \gamma_i + c\gamma_i} Y_{i,1} \\ K_{i,2} = g^{\gamma_i} Y_{i,2} \\ K_{i,3} = (t^{ID} h_i)^{\gamma_i} Y_{i,3} \end{cases} \quad (2)$$

In Eq. (2), $K_{i,1}$, $K_{i,2}$, and $K_{i,3}$ represent the attribute private key information of the data user. t^{ID} represents the identity ID of the data user. On this basis, the patient encrypts the EHR using their symmetric key, uploads it to the IPFS system, obtains the corresponding storage hash value, and stores the ciphertext in a shared contract. Among them, the EHR expression is shown in Eq. (3).

$$M = (key \| hash_{ipfs}) \quad (3)$$

In Eq. (3), M represents HER data. $hash_{ipfs}$ represents the hash value of the storage address in the IPFS system. key represents symmetric key information. The ciphertext expression is shown in Eq. (4).

$$CT = (C_0, C_1, \{C_{x,0}, C_{x,1}, \{C_{x,y,1}, C_{x,y,2}\}_{y=\{1,\ldots,l_{\rho(x)}\}}\}_{x \in \{1,\ldots,I\}}) \quad (4)$$

In Eq. (4), CT represents ciphertext information. x and y represent rows and columns. $\rho(x)$ represents attributes. l represents the length of the access address. C represents encryption. Meanwhile, the data user decrypts the ciphertext information based on their own attribute private key. When the data user meets the set orientation strategy and is not included in the attribute revocation list, they can obtain the storage information and decryption key of patient EHR data in IPFS. The identity discrimination calculation method for data users is shown in Eq. (5).

$$\begin{cases} F_{x,1} = \prod_{y=1}^{l_{\rho(x)}} \left(\frac{e(K_{\rho(x),2}, C_{x,y,2})}{e(K_{\rho(x),3}, C_{x,y,1})}\right)^{1/ID - ID_y} \\ F_{x,2} = \frac{e(K_{\rho(x),1}, C_{x,0})}{e(K_{\rho(x),2}, C_{x,1})} \end{cases} \quad (5)$$

In Eq. (5), $F_{x,1}$ and $F_{x,2}$ represent the conditions that satisfy the data user being accessed. The expression for EHR data obtained by data users is shown in Eq. (6).

$$M = (key \| hash_{ipfs}) = C_0 \cdot \frac{1}{e(K_0, C_1)} \prod_{x \in L} (F_{x,1}, F_{x,2})^{u_x} \quad (6)$$

In Eq. (6), u_x represents the recovery coefficient. In addition, in the EHRSS model, patients have the right to specify the users of their private data and perform fine-grained revocation of attribute sets, without updating the private key information of other data users associated with the ciphertext. Taking revoking a certain identification ID as an example, the patient needs to add the data user's ID to the revocation list corresponding to the attribute, and re encrypt and upload the electronic health record data to the IPFS system, replacing the access policy set in the shared contract.

B. Design of Privacy Protection Methods Based on Blockchain

In the process of EHR data sharing, relying solely on the security of the IPFS system and blockchain cannot fully guarantee the privacy of patient EHR data. Therefore, based on the proposed EHRSS model, further research was conducted on the privacy protection of EHR using ZKP and ElGamal encryption algorithms on the basis of BCT. ZKP can prove to other verifiers that a proposition is true without disclosing any actual information of the verifier [16-17]. Therefore, the proof process of ZKP in the proposed EHRPP method is shown in Fig. 3.

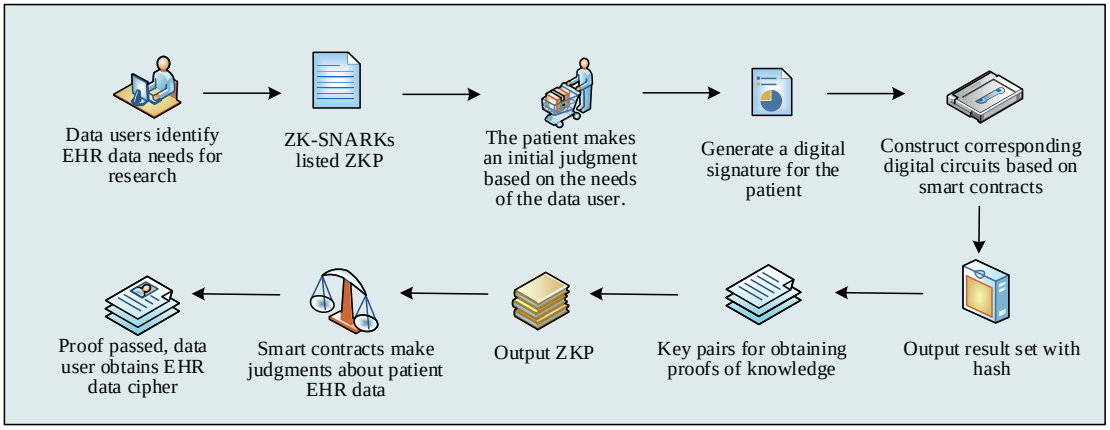


Fig. 3. ZKP's proof process in EHRPP.

In Fig. 3, data users use Zero-Knowledge Succinct Non-Interactive Argument of Knowledge (ZK-SNARKs) list the ZKP of patient EHR data required for their research, and publish the correlation results and hash values generated based on digital circuits into smart contracts. Secondly, patients make a preliminary judgment on whether they meet their expectations based on the data keywords disclosed by medical institutions. If so, continue to validate the patient's data to ensure that it meets the needs of the data user. The patient randomly selects a numerical value and combines it with information such as ID, timestamp, and EHR data to generate a digital signature. The specific expression is shown in Eq. (7).

$$q_z = AuthSign(x_p, H_1(ID_p, \tau, M, k)) \quad (7)$$

In Eq. (6), q_z represents digital signature. $AuthSign(*)$ represents authorized signature. x_p represents the patient's private key. ID_p represents the patient's identification information ID. τ represents timestamp. k represents a random number. H_1 represents a hash function that can resist collisions. At the same time, patients establish corresponding digital circuits based on smart contracts, and combine random values to obtain the EHR dataset, additional data, and common parameters of the system. The specific expression is shown in Eq. (8).

$$\begin{cases} M' = < m_1 ..., m_n, k > \\ < ID_p, \tau > \\ V(< m_1 ..., m_n, r >) \rightarrow (R, H) \end{cases} \quad (8)$$

In Eq. (8), M' represents the HER dataset obtained by selecting a random number k. m_n represents HER data. V represents the digital circuit constructed by the patient. R represents the result set. H represents the hash value. r represents the output result. Based on the above parameters, it inputs and calculates the result set and hash value of EHR to prove the authenticity and availability of the obtained EHR data. After inputting system security parameters and digital circuits, it can obtain the key pair information of ZKP. The specific expression is shown in Eq. (9).

$$ZKPkeygen(1^{\eta}, V) \rightarrow (Pk_V, Uk_V) \quad (9)$$

In Eq. (9), $ZKPkeygen(*)$ represents the ZK-SNARKs algorithm. Pk_V represents the key for listing ZKP. Uk_V represents the key for verifying ZKP. η represents system security parameters. Input the patient's EHR data, digital signature, and the key generated by ZKP, as well as the obtained result set and hash value, and then output ZKP. The specific expression formula is shown in Eq. (10).

$$Prove(M, q_z, Pk_V, R, H) \rightarrow \pi \quad (10)$$

In Eq. (10), $Prove(*)$ represents the output of patient

related information. π represents ZKP. After the patient submits the ZKP, the EHR data of the patient is determined based on the smart contract to determine whether it meets the needs of the data user. The specific expression is shown in Eq. (11).

$$Verify(Uk_V, \pi, q_z, y_p, R, H) \rightarrow (true / false) \quad (11)$$

In Eq. (11), y_p represents the patient's public key. ZKP verifies the digital signature of EHR data using the patient's public key, and determines the ZKP, result set, hash value, ZKP generated by the data user, result set, and hash value. When all the above results meet the verification requirements, ZKP will output "true" to EHRPP, otherwise it will output "false". Therefore, the EHRPP model architecture based on the proposed ZKP is shown in Fig. 4.

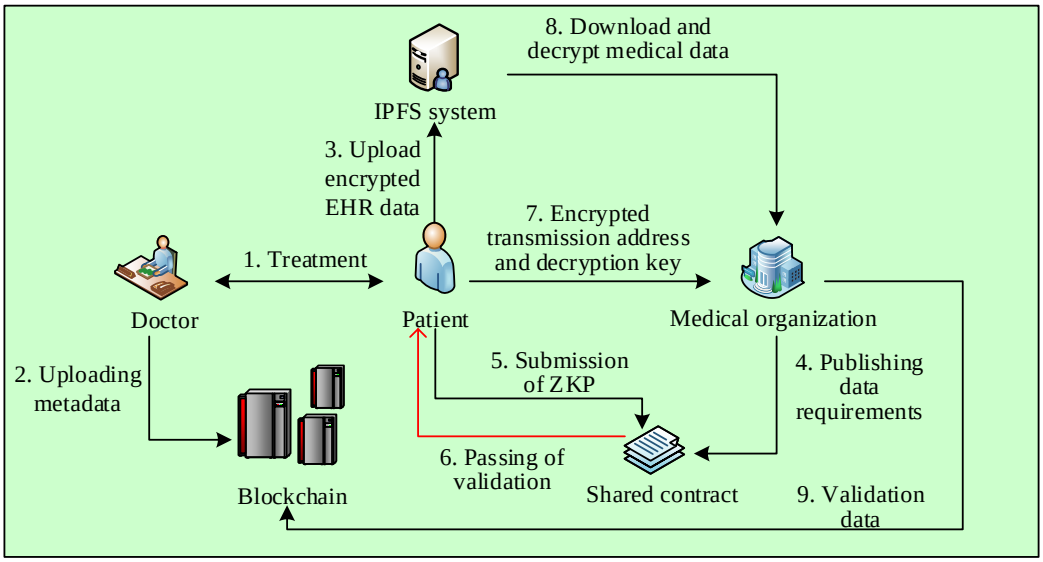


Fig. 4. Blockchain-based EHRPP model architecture.

From Fig. 4, the EHRPP model proposed in the study differs from the EHRSS model in that it divides data users into doctors and medical institutions. This is because the study considers that on the basis of secure sharing of patient EHR data, medical institutions need to use EHR data for research or analysis to promote the recovery of medical diseases. Therefore, the study split the data users in the EHRPP model. Among them, doctors mainly generate EHR data for patients and are responsible for uploading patient metadata to blockchain for recording. Before using EHR data, medical institutions need to prove and define ZKP, and write the required keywords into smart contracts. Therefore, the process of the EHRPP model proposed in the study is shown in Fig. 5.

From Fig. 5, the EHRPP model first sets security parameters and parameters such as large prime numbers, meta groups, cyclic groups, priority over representation, hash functions, etc., to generate common parameters and system master keys. The specific expression is denoted in Eq. (12).

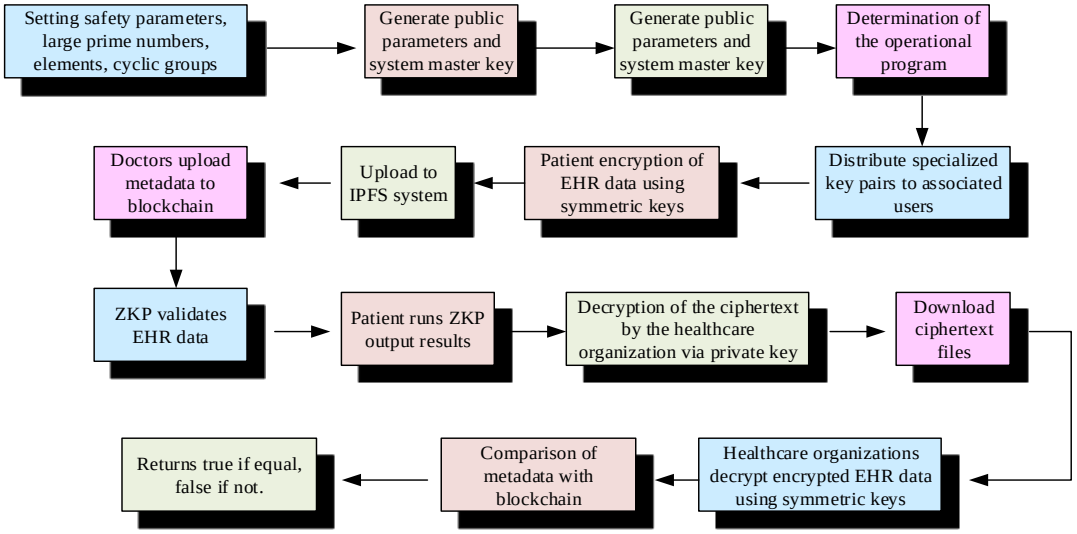


Fig. 5. Blockchain-based EHRPP model flowchart.

$$Pk = (p, g, G_1, Z_p, H_1, H_2) \tag{12}$$

In Eq. (12), p represents large prime numbers. G_1 represents a cyclic group of order p. Z_p represents a finite field. H_2 represents an reversible hash function. The system conducts qualification review for patients, doctors, and medical institutions with requirements, and creates corresponding key pairs for them. The three obtain their respective public key calculation formulas as shown in Eq. (13).

$$\begin{cases} y_p = g^{x_p} modp \\ y_d = g^{x_d} modp \\ y_r = g^{x_r} modp \end{cases} \tag{13}$$

In Eq. (13), y_d and y_r respectively represent the public keys of doctors and medical institutions. x_d and x_r respectively represent the private keys of doctors and medical institutions. The patient encrypts EHR data using a symmetric key and uploads the ciphertext to the IPFS system to obtain the corresponding storage hash value. Meanwhile, doctors upload patient metadata to blockchain for recording and storage. Medical institutions provide ZKP certification based on the patient EHR data they need. After the ZKP verification is passed, the medical institution sends an application to the patient to obtain EHR data information. The patient randomly outputs the shared data and synchronously stores it in the system. The specific expression is shown in Eq. (14).

$$\begin{cases} key' = H_2(key \| hash_{ipfs}) \\ s_1 = g^{\gamma_1} \\ s_2 = y_r^{\gamma_1} key' \\ s_3 = H_2(key \| hash_{ipfs}) \cdot y_r \end{cases} \tag{14}$$

In Eq. (14), s_1, s_2, and s_3 represent the results obtained by calculating the application information of medical institutions. key' represents the symmetric key of the medical institution. Medical institutions obtain IPFS information and symmetric keys for stored HER data based on the key. At this point, the system checks the medical institution based on the hash value and identification ID, as shown in Eq. (15).

$$\begin{cases} k_s = s_1 s_2^{-x_r} \\ k_s' = H_2^{-1}(k_s) = key \| hash_{ipfs} \\ check = H_1(key \| hash_{ipfs}) \cdot y_r \end{cases} \tag{15}$$

In Eq. (15), k_s and k_s' represent the application information calculated by the patient and medical institution, respectively. The system compares the examination values of medical institutions with the medical institution information stored by patients in the system. When the two are equal, it indicates that the transaction is legal. At this point, medical institutions can obtain encrypted EHR data by downloading based on hash values. Conversely, the system determines that the medical institution is a malicious user and punishes them. After downloading EHR data, medical institutions can use symmetric keys to decrypt the data and obtain the original EHR data. At the same time, it compares the hash value of EHR data with the metadata of blockchain records to determine whether the data is EHR data required by medical institutions.

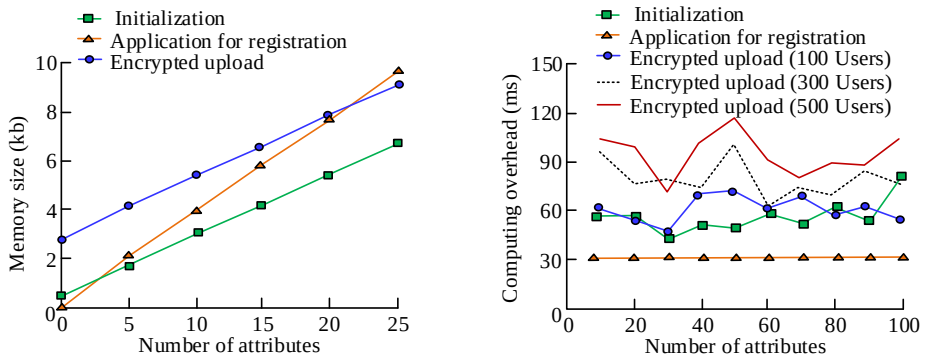
IV. Results

To verify the effectiveness of the EHR data security sharing and privacy protection methods proposed on the basis of blockchain, the study first analyzed the properties and encryption efficiency of the EHRSS method during the encryption and upload stages. Secondly, performance validation and analysis were conducted on the proposed EHRPP method.

A. Verification and Analysis of Security Sharing Methods Based on Blockchain

To effectively validate the effectiveness of the EHRSS method, simulation experiments were conducted on the Java Pairing Based Cryptography (JPBC) library in the Java language. It assumed that the cyclic group and generator are both 1024 bits, the ID length is 64 bits, the account length is 160 bits, and the IPFS address length is 256 bits. In EHRSS, it interacted with blockchain during initialization, registration application, encryption, and upload stages. Therefore, the study first analyzed the changes in storage size, computational cost, and number of attributes in three stages, as shown in Fig. 6.

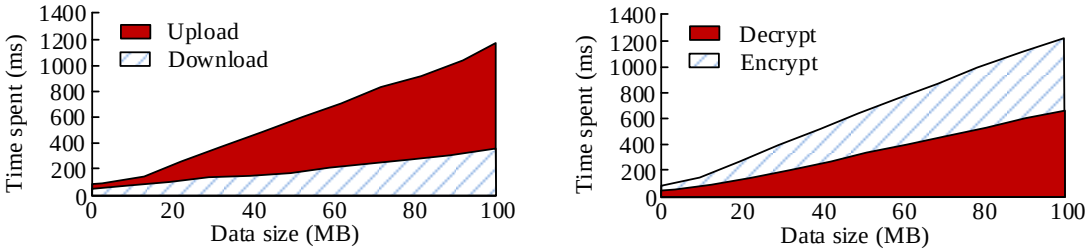
Fig. 6(a) showcases the relationship between the storage phases of the EHRSS model in three stages and the amount of attributes when the revocation list has 10 data users. As the amount of attributes increases, the storage overhead for the three stages of model initialization, application for registration, and encryption upload all increased. Based on the calculation cost of the three stages in Fig. 6(b), as the amount of attributes changes, the calculation cost of the initialization stage first increased and then decreased with the increase of the number of attributes, but the overall change is relatively small. The computational cost during the registration application stage remained generally stable as the amount of attributes increased. However, the computational cost of EHRSS encryption and uploading was not affected by the amount of attributes for different user numbers. This indicated that users can expand the attributes in the EHR data sharing project as needed, and the computational efficiency will not be reduced by the increase in the number of attributes. On this basis, the study further analyzed the impact of different sizes of EHR data on IPFS system upload and download, encryption and decryption, as shown in Fig. 7.



(a) Variation of storage size with the number of attributes for the three phases

(b) Variation of computational overhead with the number of attributes for the three phases

Fig. 6. Relationship between the three phases of EHRSS and changes in the number of attributes.



(a) Time overhead of uploading/downloading EHR data of different sizes to/from IPFS

(b) Time overhead of symmetric encryption and decryption of EHR data of different sizes

Fig. 7. Impact of different sizes of EHR data on uploading and downloading, encryption and decryption in IPFS systems.

From Fig. 7 (a), as the EHR data increased, the upload and download time overhead of the IPFS system also increased. When the EHR data size was 100MB, the upload time cost in the IPFS system was 1.17s, and the download time cost for h was 0.36s. Based on Fig. 7 (b), the proposed security sharing method had lower encryption and decryption time costs for EHRSS under different EHR data sizes, and had ideal efficiency in processing large-scale EHR data. Therefore, the study further compared the performance of medical data security sharing methods proposed by other scholars with EHRSS. The EHR data size was set to 2GB, and the specific comparison results are denoted in Table I.

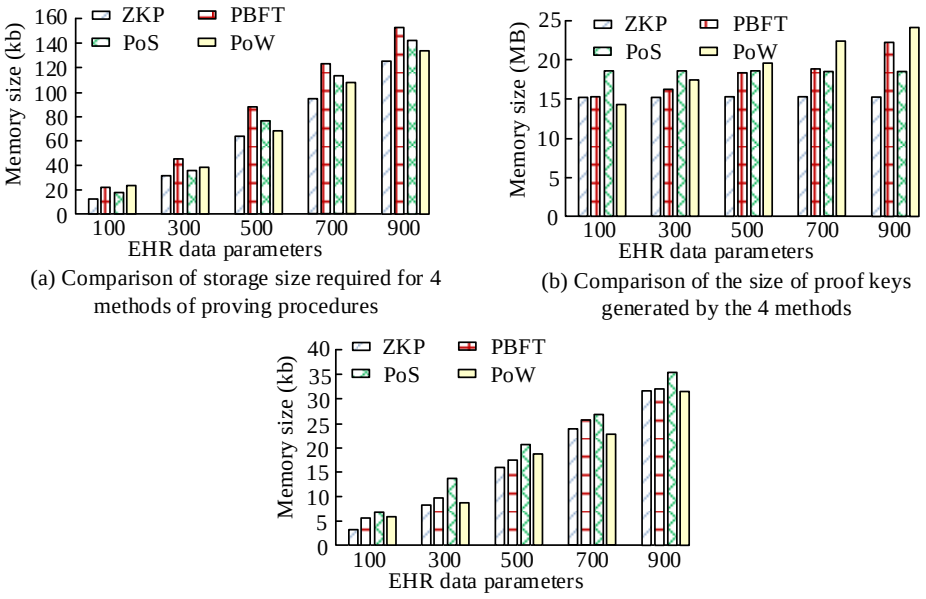
TABLE I. PERFORMANCE COMPARISON OF DIFFERENT SECURITY-SHARING METHODS

Methods of secure data sharing	Encryption overhead (s)	Generated key size (kb)	Decryption overhead (s)
Reference [6]	45.23	124.24	40.35
Reference [7]	42.54	150.00	34.62
Reference [18]	35.46	128.00	29.88
Reference [19]	22.43	89.75	15.87
EHRSS	19.23	54.32	6.63

From Table I, the EHRSS method proposed in the study required significantly less encryption and decryption time compared to other methods. The encryption time required for EHRSS was reduced by an average of 47.20% compared to other methods, while the decryption time was reduced by an average of 78.03%. This indicated that the introduction of attribute revocation lists on the basis of blockchain has improved the encryption and decryption efficiency of data security sharing. By comparing the key sizes generated by different algorithms, the proposed method reduced them by 56.28%, 63.79%, 57.56%, and 39.48%, respectively, compared to other methods. This indicated that the algorithm raised in the study not only improves the granularity of attribute revocation, but also enhances the convenience of ciphertext applications. Compared to other methods, EHRSS has superior computational efficiency and practicality.

B. Verification and Analysis of Privacy Protection Methods based on Blockchain

To further demonstrate the privacy and security of EHRPP in protecting patient EHR data, this study verified and analyzed the performance of the ZK-SNARKs algorithm in the EHRPP method, the required time for verifying keys, generating proofs, and the time cost for verifying proofs. The research set the security parameter to 128 bits, ZKP was defined by the libSNARK code library, and each experiment was repeated 10 times. The average of each indicator was taken for the experimental results. Meanwhile, the Practical Byzantine Fault Tolerance (PBFT) mechanism, Proof of Stake (PoS) mechanism, and PoW mechanism were introduced and compared with ZKP. The required storage sizes for the four mechanisms under different EHR data scales are shown in Fig. 8.



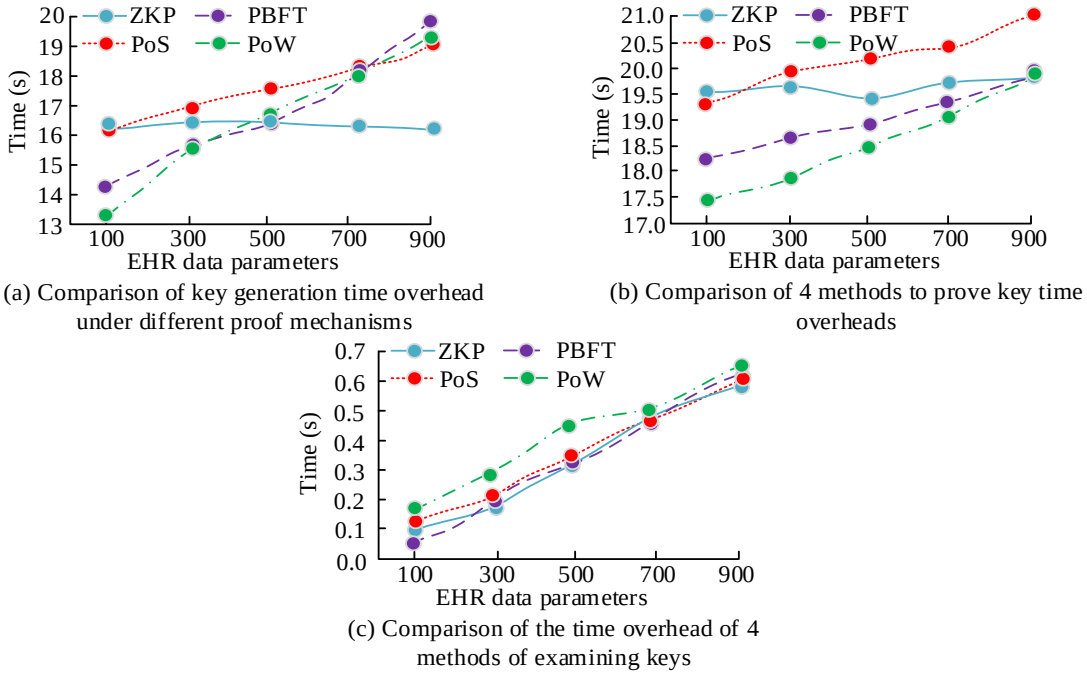
(a) Comparison of storage size required for 4 methods of proving procedures

(b) Comparison of the size of proof keys generated by the 4 methods

(c) Comparison of storage size of authentication keys generated by 4 methods

Fig. 8. Comparison of key generation and verification storage size under different proof mechanisms.

From Fig. 8(a), as the EHR data parameters increased, the storage size required for all four proof methods also increased. Compared to other methods, ZKP required smaller storage. Combining the four proof methods in Fig. 8(b), ZKP reduced the storage requirements for keys by an average of 5.18%, 13.14%, 19.45%, 24.00%, and 30.05% compared to the other three methods when the medical data parameters were 100, 300, 500, 700, and 900, respectively. The size of ZKP and PoS proof keys remained basically unchanged, while PoW's proof key size increased as the number of parameters increased, although the proof size was less than ZKP when the parameter was 100. This indicated that the proof process of ZKP is more stable. By comparing the verification key sizes of the four methods in Fig. 8(c), when the parameter was 900, the verification key size of ZKP was 31.80kb, which was 4.70% less than the other three methods. This indicated that the EHRPP based on the ZKP proposed in the study has superior performance in protecting patient privacy, balancing the security sharing and privacy protection issues of EHR data. Meanwhile, the study further compared the time overhead for generating keys, proving keys, and verifying keys using four methods, as shown in Fig. 9.



(a) Comparison of key generation time overhead under different proof mechanisms

(b) Comparison of 4 methods to prove key time overheads

(c) Comparison of the time overhead of 4 methods of examining keys

Fig. 9. Comparison of the time overhead required for key generation and verification under different authentication mechanisms.

As shown in Fig. 9(a), the time overhead of generating keys for ZKP under different EHR data parameters showed no significant change, with an average time cost of 16.29 seconds. Compared with the other three methods, the average time overhead decreased by 3.93%, 7.39%, and 1.51%, respectively. From the comparison of the time cost required to prove the key using the four methods in Fig. 9(b), ZKP was least affected by the size of data parameters. The other three algorithms showed an upward trend with the increase of parameter size. This may be because during the key proof process, the storage capacity of the three algorithms for proving keys is relatively large, which requires more time for proof. Fig. 9(c) shows the time overhead for four algorithms to verify whether the key information is the data required by medical institutions. When the parameter quantity of EHR data was 900, the time overhead of ZKP was reduced by an average of 4.84% compared to the other three algorithms. The above verification indicates that the EHRPP proposed based on ZKP has superiority in overall performance. In addition, the study further compared the privacy protection effects of Study [18], study [19], EHRSS and EHRPP 4 methods on EHR data security sharing are shown in Fig. 10.

Fig. 10 (a), (b), (c), and (d) show the EHR data sharing protection effect of study [18], study [19], EHRSS, and EHRPP four schemes, respectively. The protection effect of study [18], study [19], EHRSS, and EHRPP was about 91%, 84%, 84%, and 93%, respectively. This may be because the method proposed in study [18] achieved data protection through secret sharing algorithms, which has less dependence on the IPFS system, while study [19], although storing data in an off chain database based on IPFS, still relied on the authorization verification of the Ethereum blockchain. However, overall, the EHRPP method raised in the study has better security than the other two methods. Compared with EHRSS, after introducing ZKP, its privacy protection effect on EHR data was significantly improved. The performance comparison results of EHRPP with study [18] and study [19] in EHR data with 500 input parameters are denoted in Table II.

From Table II, the size of the proof key generated by EHRPP was only 15.2MB, and the time overhead for generating the proof key was16.3s. Compared with the key sizes generated in studies [18] and study [19], EHRPP had an average reduction of 84.52%. This indicated that the EHRPP method proposed in the study had a faster speed in generating proof key pairs, while comparing the time overhead for verifying keys with the three methods, the time overhead required in study [18] was lower. This may be because both EHRPP and study [19] were IPFS systems, while study [18] defined a group secret sharing algorithm architecture. However, overall comparison shows that EHRPP still has significant advantages in overall performance and security privacy.

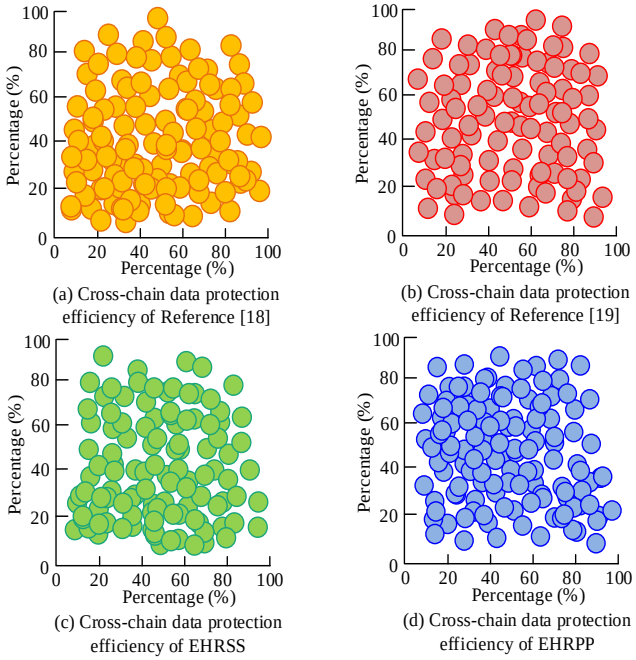


Fig. 10. Comparison of the effectiveness of cross-chain data protection.

TABLE II. PERFORMANCE COMPARISON OF DIFFERENT METHODS

Method	Key generation process (s)	Proof key (s)	Authentication key (s)	Proof key size (MB)	Authentication key size (KB)
EHRPP	16.3	19.5	0.32	15.2	16.2
Reference [18]	15.9	21.2	0.04	165.4	16.2
Reference [19]	21.5	24.6	0.45	31.0	16.2

V. DISCUSSION

The study proposes a blockchain based EHR secure sharing and privacy protection method aimed at improving the security and privacy protection of medical data, ensuring that patients have absolute ownership of their medical data. Through experimental verification, the proposed method has significant advantages in generating key size and time cost. Compared with existing recognized mechanisms, ZKP reduces the average key generation time cost by 54.36%. In addition, this method has an average improvement of 7.73% in data protection effectiveness compared to other methods. This is consistent with the results obtained by Konkin A et al. in their study of ZKP [20]. By combining attribute encryption and blockchain computing to construct a data security sharing model, as well as introducing zero knowledge proof and ElGamal encryption algorithm, the research has successfully improved the construction of data privacy protection. The proposed method shows high efficiency in generating key size and time cost. Especially, compared with studies [18] and [19], the reduction in key generation time of ZKP indicates its potential advantages in handling large-scale data. Through smart contracts and attribute based encryption, patients can have more precise control over access and sharing of their EHR data, ensuring their absolute rights to their data. It can be considered that the introduction of ZKP and ElGamal algorithms on the basis of existing blockchain technology is an innovative attempt to improve the security and privacy of data sharing. Compared with other proposed data sharing

frameworks, the research method shows lower time overhead and smaller key size in key generation, proof generation, and verification.

VI. Conclusion

To improve the security sharing and privacy protection performance of medical data systems, research explored security sharing EHR data privacy protection methods based on blockchain. Firstly, an EHRSS method based on BCT was proposed to improve the security of EHR data through attribute encryption algorithms. Secondly, the EHRPP model was constructed by introducing ZKP and ElGamal encryption algorithms. Experimental verification showed that compared to the other four methods, the key sizes generated by EHRSS decreased by 56.28%, 63.79%, 57.56%, and 39.48%, respectively. When the parameter was 900, the verification key size of ZKP was 31.80kb, which is 4.70% less than the other three methods. The data protection effect of EHRPP obtained by introducing ZKP on the basis of EHRSS increased by 10.71% compared to EHRSS. Compared to other methods, the key generated by EHRPP was only 15.2MB, and the time overhead for generating the proof key was 16.3s, resulting in an average reduction of 84.52% in key size. The outcomes indicated that the EHR data security sharing and privacy protection method proposed in the study can improve the overall performance and security of the system, and has positive application significance in medical data security and privacy protection. However, the study only conducted theoretical exploration and experimental analysis of security sharing and privacy protection methods. In the future, it will consider further optimizing ZKP technology, compressing its scale and generation time, and improving the security of data privacy protection.

Fundings

The research is supported by Social Science Foundation of Anhui Province: Research on blockchain enabling healthcare data flow and governance transformation (SK2020A0863) and Top Talent Foundation of Anhui Province: Exploration of blockchain enabling medical insurance innovation and development (No. gxgnfx2021057); Social Sciences key foundation of Anhui business and technology college: Research on the optimization of individual pension system of China under the background of deep aging (SK2024A001).

References

[1] Reilly J B, Kim J G, Cooney R, DeWaters A L, Holmboe E S, Mazotti L, Gonzalo J D. Breaking down silos between medical education and health systems: creating an integrated multilevel data model to advance the systems-based practice competency. Academic Medicine, 2024, 99(2), 146-152.

[2] Riedel P, von Schwerin R, Schaudt D, Hafner A, Späte C. ResNetFed: federated deep learning architecture for privacy-preserving pneumonia detection from COVID-19 chest radiographs. Journal of Healthcare Informatics Research, 2023, 7(2): 203-224.

[3] Groumpos P P. A Critical Historic Overview of Artificial Intelligence: Issues, Challenges, Opportunities, and Threats. Artificial Intelligence and Applications. 2023, 1(4): 197-213.

[4] Ortega-Calvo A S, Morcillo-Jimenez R, Fernandez-Basso C, Gutiérrez-Batista K, Vila M A, Martin-Bautista M J. Aimdp: An artificial intelligence modern data platform. use case for Spanish national health service data silo. Future Generation Computer Systems, 2023, 143(1): 248-264.

[5] Kumar R, Kumar P, Tripathi R, Gupta G P, Islam A N, Shorfuzzaman M. Permissioned blockchain and deep learning for secure and efficient data sharing in industrial healthcare systems. IEEE Transactions on Industrial Informatics, 2022, 18(11): 8065-8073.

[6] Shuaib K, Abdella J, Sallabi F, Serhani M A. Secure decentralized electronic health records sharing system based on blockchains. Journal of King Saud University-Computer and Information Sciences, 2022, 34(8): 5045-5058.

[7] Liu X, Zhao J, Li J, Cao B, Lv Z. Federated neural architecture search for medical data security. IEEE transactions on industrial informatics, 2022, 18(8): 5628-5636.

[8] Narayanan U, Paul V, Joseph S. A novel system architecture for secure authentication and data sharing in cloud enabled Big Data Environment. Journal of King Saud University-Computer and Information Sciences, 2022, 34(6): 3121-3135.

[9] Elisa N, Yang L, Chao F, Cao Y. A framework of blockchain-based secure and privacy-preserving E-government system. Wireless networks, 2023, 29(3): 1005-1015.

[10] Sharma P, Namasudra S, Chilamkurti N, Kim B G, Gonzalez Crespo R. Blockchain-based privacy preservation for IoT-enabled healthcare system. ACM Transactions on Sensor Networks, 2023, 19(3): 1-17.

[11] Awotunde J B, Gaber T, Prasad L N, Folorunso S O, Lalitha V L. Privacy and security enhancement of smart cities using hybrid deep learning-enabled blockchain. Scalable Computing: Practice and Experience, 2023, 24(3): 561-584.

[12] Zhonghua C, Goyal S B, Rajawat A S. Smart contracts attribute-based access control model for security & privacy of IoT system using blockchain and edge computing. The Journal of Supercomputing, 2024, 80(2): 1396-1425.

[13] Patil S D, Kathole A B, Kumbhare S, Vhatkar K. A Blockchain-Based Approach to Ensuring the Security of Electronic Data. International Journal of Intelligent Systems and Applications in Engineering, 2024, 12(11): 649-655.

[14] Gousteris S, Stamatiou Y C, Halkiopoulos C, Antonopoulou H, Kostopoulos N. Secure distributed cloud storage based on the blockchain technology and smart contracts. Emerging Science Journal, 2023, 7(2): 469-79.

[15] Sharma P, Jindal R, Borah M D. Blockchain-based cloud storage system with CP-ABE-based access control and revocation process. the Journal of Supercomputing, 2022, 78(6): 7700-7728.

[16] Wan Z, Zhou Y, Ren K. zk-AuthFeed: Protecting data feed to smart contracts with authenticated zero knowledge proof. IEEE Transactions on Dependable and Secure Computing, 2022, 20(2): 1335-1347.

[17] Feneuil T, Joux A, Rivain M. Shared permutation for syndrome decoding: new zero-knowledge protocol and code-based signature. Designs, Codes and Cryptography, 2023, 91(2): 563-608.

[18] Shree S, Zhou C, Barati M. Data protection in internet of medical things using blockchain and secret sharing method. The Journal of Supercomputing, 2024, 80(4): 5108-5135.

[19] Azbeg K, Ouchetto O, Andaloussi S J. BlockMedCare: A healthcare system based on IoT, Blockchain and IPFS for data management security. Egyptian informatics journal, 2022, 23(2): 329-343.

[20] Konkin A, Zapechnikov S. Systematization of knowledge: privacy methods and zero knowledge proofs in corporate blockchains. Journal of Computer Virology and Hacking Techniques, 2024, 20(2): 219-224.