2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT) | 978-1-7281-8575-0/20/$31.00 ©2020 IEEE | DOI: 10.1109/ISCTT51595.2020.00032

A New Blockchain-based Electronic Medical Record Transferring System with Data Privacy

Jian Li
Guangxi Normal University
College of Electronics and Engineering
Guilin, China
936875100@qq.com

***Abstract*—From the current national-level medical diagnosis and treatment perspective, the paper-version prescription is still the only slip issued by the doctor to allow for the patient to take the prescribed medicine from the pharmacy store or clinic. However, the paper prescription is not easy to store and its confidentiality is low, if it is lost, it will cause unnecessary trouble to the patient. Therefore, the use of digital prescriptions has been considered for future electronic medical record reform. The challenges of digital prescriptions are also upfront, for instance, it contains a large amount of patient's medical record information and private information, how to effectively maintain the level of security while the premise of privacy has caused a series of hot issues, which become a major concerns of people who promote the use of digital prescriptions. In recent years, blockchain has been widely regarded as a promising online business data and transaction security technology. Due to its decentralization, security and credibility, collective maintenance, and non-tampering, it is widely used in data storage and sharing. This paper proposes a blockchain-based medical prescription system. At the same time, in order to effectively protect the privacy of patients, during this period we adapted the k-anonymity algorithm and incorporated a differential privacy protection mechanism to enhance its secured performance.**

Keywords-component; Blockchain; prescription; K-anonymity; Differential privacy

I. INTRODUCTION

Nowadays, although online shopping has become popular all over the world, it is still very common in life to use paper-version prescriptions to get prescribed medicines from hospital pharmacies or outside pharmacies. Digital prescriptions formats such as jpg, pdf and html prescriptions in electronic file formats cannot be used directly. Some pharmacies overseas provide online purchase of prescription drugs, but in fact, customers still need to inform the pharmacy of their original prescription (paper prescription) by email after the transaction is completed. However, in some places, digital prescriptions can only be used for emergency supply of medicines, but under national-level practical regulations, a pharmacy can often only provide about three days of medicine. It is obvious that digital prescriptions can only be used in very limited circumstances. Therefore, it is very necessary to use digital prescriptions to replace traditional paper prescriptions, but the more difficult problem in this process is people's authentication of digital prescriptions and the hidden dangers of digital prescriptions, such as privacy issues and security issues.

Blockchain technology is the underlying structure of Bitcoin. It has been developed for more than ten years since the proposal of blockchain technology [1]. Blockchain has become one of the most popular technologies at the moment. It has the characteristics of decentralization, and the data in the blockchain has the characteristics of traceability and non-tampering. Blockchain has been adopted or considered by many industries. The main driving force for using blockchain in these applications is the introduction of digital identification, distributed security, smart contracts, and micrometrics through distributed blockchain ledger [2]. Due to the blessing of blockchain technology, data cannot be easily tampered with by attackers. Therefore, multiple service providers can jointly maintain the user's account information under this encryption function, and the user only needs to maintain the account information on the ledger for all identity verification, thereby improving its overall execution efficiency.

This article develops a medical prescription system that is based on blockchain and can provide online prescription processing with secure storage, identity verification and access rights. In addition, the system adopts the k-anonymity protection method based on differential privacy in the event of issuing electronic prescriptions, so that the security of the data is effectively improved while maintaining data privacy.

II. RELATED WORK

With the improvement of the world's intelligence, people's understanding of the blockchain has slowly made great progress. After that, they began to consider applying the blockchain to various fields. Gupta et al. [3] discussed how the blockchain realizes the interoperability of medical data and the security of access. But Gupta's solution is only to store metadata about medical health and medical events in the blockchain, such as patient identity, access ID, payer ID and other data. MIT Azaria et al. developed a system called MedRec [4] to solve the problem of data interoperability and rights management for managing medical records. MedRec proposes three types of smart contracts, namely, registration contracts, doctor-patient relationship contracts and summary contracts to realize data authority management. The registration contract is used to manage the user information table, and the public key is used as its identity to realize the anonymous login of the patient; the doctor-patient relationship contract defines a series of pointers and related access rights, through which the data in the database is accessed, which is used for one doctor to one doctor. summary contract is used to manage all medical

978-1-7281-8575-0/20/$31.00 ©2020 IEEE
DOI 10.1109/ISCTT51595.2020.00032

information data collections generated by patients. We have adopted a similar smart contract design method in this article. The difference is that we have designed five types of smart contracts, including blockchain medical information system contracts, patient contracts, doctor contracts, researcher contracts, and case treatment contracts.

The current privacy protection methods can be divided into three types: anonymous privacy protection [5-6], the use of hidden identifier attributes [7] (identity attribute, ID number, name and other attributes that can represent individual information) and generalized quasi-identification Quasi-identifier attribute (age, gender, birthday, zip code, etc. can be deduced to represent the attributes of individual information) to protect sensitive attributes (sensitive attributes, diseases, salary and other attributes that users are not willing to disclose) from being leaked purpose; Agrawal et al. proposed for the first time an algorithm to construct a classification tree on the data after noise interference, which maintains the classification results to the greatest extent [8]. Sweeney et al. proposed a k-anonymity algorithm for anonymizing data, which ensures that any record is indistinguishable from other k-1 records, thereby protecting private data [9]. Differential privacy [10] is a new privacy definition proposed by Dwork in 2006 for the problem of privacy leakage in statistical databases. This concept, as the name suggests, is to prevent differential attacks [11] for privacy protection. When querying two adjacent data sets with statistical differences and only one record difference, the probability of obtaining the same value is extremely close; Differential privacy is essentially privacy protection based on cryptography, and the purpose of privacy protection is achieved through data encryption. But in essence, the computational resources consumed by such methods are too large, so they are rarely applied in data publishing and data mining. This paper introduces a differential privacy protection mechanism in the K-anonymity algorithm to prevent external users from obtaining medical data through exchange and sharing, and then obtaining sensitive information of patients through links and inferences, resulting in privacy leakage.

III. K-ANONYMITY AND DIFFERENTIAL PRIVACY

A. Differential Privacy Protection

Differential privacy protection is to add noise to the original data for random disturbance to cause its distortion, and finally make it impossible for an attacker to use known data information to infer the effect of more data content.

Definition 1: Two data sets given only one record apart D_1 and D_2 , Range(K) represents the range of values of the random algorithm K, if any result S ϵ Range(K) of the dataset in K satisfies equation 1, then say that Algorithm K satisfies ε-differential privacy [12].

$$\Pr\left[K(D_1) \in S\right] \leqslant \exp\left(\varepsilon\right) \times \Pr\left[K(D_2) \in S\right] \qquad (1)$$

Where $K(D_1)$ and $K(D_2)$ are the output results obtained from K with D_1 and D_2 as inputs. $\Pr\left[K(D_1) \in S\right]$ denotes the probability that the result is S, also known as the risk of privacy being compromised [13]. The smaller ε is, the greater the noise, the better the data distortion effect, and thus the higher the level of privacy protection, and vice versa.

Data privacy is guaranteed when the function K satisfies definition 1, and is independent of the degree of background knowledge the attacker has. The main way to achieve differential privacy protection is to add noise, and the two commonly used noise mechanisms are the Laplace mechanism and the exponential mechanism, so the mechanism used in this paper is the Laplace mechanism. The Laplace mechanism is suitable for numerical data protection and implements ε-differential privacy protection by adding random noise that obeys a Laplace distribution to the exact query result.

Remembering that the Laplace distribution with position parameter 0 and scale parameter b is $\mathrm{Lap}(b)$, the probability density function is:

$$\mathrm{P(x)} = \frac{1}{2\mathrm{b}} exp\ (-\frac{|\mathrm{X}|}{\mathrm{b}}) \qquad (2)$$

Definition 2: Suppose there is a query function $f: D \to \mathcal{R}$, input to a dataset and output to a d-dimensional real vector, for any neighboring dataset D and D':

$$\Delta f = \max_{D,D'} \|f(D) - f(D')\|_1 \qquad (3)$$

where $\|f(D) - f(D')\|_1$ is the 1-order parametric distance between $f(D)$ and $f(D')$. Sensitivity Δf is only related to the type of query f. It considers the maximal difference between the query results on neighboring datasets.

Definition 3 (Laplace mechanism): For a function $f: D \to \mathcal{R}$ over a dataset D, the mechanism K in (4) provides the ε-differential privacy.

$$K(D) = f(D) + (\mathrm{Lap}\ (\Delta f/\varepsilon)) \qquad (4)$$

where $\mathrm{Lap}\ (\Delta f/\varepsilon)$ is random noise, obeying a Laplace distribution with a scale parameter of $\Delta f/\varepsilon$, and Δf is the sensitivity of the query function f. The sensitivity is only related to the query function itself and is independent of the size of the dataset. The greater the sensitivity of the query function, the more noise needs to be added.

B. K-anonymity

The two main data anonymization schemes currently available are generalized anonymization techniques [12] and microaggregation-based anonymization techniques. Among them, generalized anonymization techniques are widely used, and k-anonymization is one of the most representative methods among the generalized anonymization techniques.

Definition 4: Given a data table $R(Q_1, Q_2, \cdots, Q_n)$, QI is a quasi-identifier associating with R. R satisfies k-anonymous if and only if each sequence of values occurs at least k times in R[QI].

As shown in TABLEI, the quasi-identifier QI in this instance table is {Race, Birth, Sex, Zip, Marital status}, and any ordered tuple value appearing in R[QI] repeats at least twice in R[QI], r_1[QI]= r_2[QI]= r_3[QI], r_4[QI]= r_5[QI]. Then the instance table satisfies the k-anonymity protection of k=2, and the individual tuple data derived by the attacker using an external data source cannot point to any particular individual.

K-anonymity was mainly achieved by generalization techniques. TABLEII shows the generalization level of this data set for the disease skin allergy, age 29, zip code 212000. The conversion process is represented by a three-dimensional vector of (x, y, z), with generalization levels corresponding to x for disease, y for age, and z for zip code, respectively. The highest data generalization level for this example is (2,1,3). As the level of generalization increases, so does the amount of information lost.

TABLE I K-ANONYMITY INSTANCE TABLE

Num	Race	Birth	Sex	Zip	Marital status	Disease
r_1	Asian	94/0 */* *	F	941 * *	Been-married	Hypertension
r_2	Asian	94/0 */* *	F	941 * *	Been-married	Obesity
r_3	Asian	94/0 */* *	F	941 * *	Been-married	Chest pain
r_4	Asian	93/0 */* *	M	941 * *	Been-married	Obesity
r_5	Asian	93/0*/* *	M	941 * *	Been-married	Short breath

TABLE II GENERALIZATION LEVEL

Disease	Age	Zip	Level
		* * * * * *	3
*		21 * * * *	2
Sensibility	⩽50	2120 * *	1
Skin-sensibility	29	212000	0

C. *A Differential Privacy-based Approach to Data Anonymization protection Privacy Protection Methods*

In this paper, the process of partitioning attributes in a dataset is used as a method for anonymizing privacy protection. The Boolean values of the sensitive attributes are added with appropriate Laplace noise that meets the differential privacy conditions, thus improving the security of the sensitive information.

The privacy budget added in step 4, ε can be set independently according to the level of protection required, and the noise is added by integrating Eq. 2 to obtain the Laplace cumulative distribution function, and then the noise value $\mathrm{Lap}(1/\varepsilon)$ is obtained from the cumulative distribution function.

$$\mathrm{Lap}(1/\varepsilon) = -b * \mathrm{sgn}(p-0.5) * \mathrm{In}(1-2*|p-0.5|) \quad (5)$$

D. *Algorithom Design Process*

The steps of the privacy protection algorithm are as follows.

Input: raw data set $R(Q_1, Q_2, \cdots, Q_n) = R[Q]$;

Output: data set $R[Q']$ under privacy protection.

Step 1: Divide the input raw dataset $R[Q]$ into two small datasets, including the dataset $R(I_1,..., I_m) = R[I]$ and its complement set $R(J_1,..., J_{n-m}) = R[J]$. The data set obtained by merging $R[I]$ and $R[J]$ is recorded as $R[I, J]$, and therefore has $R[Q] = R[I, J]$. Remember $R[H]$ to represent the sensitive attributes of the original dataset $R[I]$.

Step 2: The method dataset $R[J]$ is obtained for $R[J']$ using a low-level generalization hierarchy for processing sensitive attributes so that it satisfies the k-anonymity protection requirement.

Step 3: Convert $R[K]$ to a data set $R[K] = R[I, J']$, where $R[J']$ is a k-anonymity protected data set, compress $R[J']$, save only one tuple that is the same in $R[J']$, and a Number attribute is added to the table to record the number of times the same tuple exists. Also, convert the expression of the attribute in $R[I]$ yields $R[H']$. For example, in the case of the HIV property, the value of the property is both Y and N. Converting the property to HIV(Y) yields $R[K']$.

Step 4: Add Laplacian noise that meets the differential privacy protection requirements to the data set $R[H']$ in $R[K']$ to obtain the data set $R[Q']$.

Step 5: Return the data set $R[Q']$.

IV. MODEL OVERVIEW

The electronic medical record storage model is to store the patient's personal base data, the patient's examination data, the patient's treatment data, and the patient's final examination and treatment cost data in a block, as shown in Figure 1. The patient's examination data can be encrypted by the patient's own public key. In this way, the patient has autonomy over his or her own examination data and can give others the right to view his or her examination data according to his or her own wishes. For the patient's treatment plan, since it is given by the patient's doctor, the treatment data is encrypted with the hospital's public key, so that the patient's doctor can study it with other medical experts in the hospital. The patient's personal data, as well as the final cost of the treatment, are stored directly in the block, which is not key protected and can be accessed at any time by both the hospital and the patient.

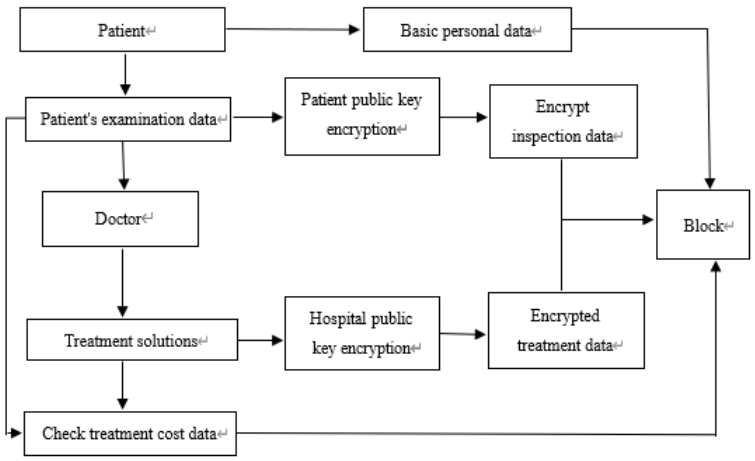

Fig. 1. Electronic medical record storage mode

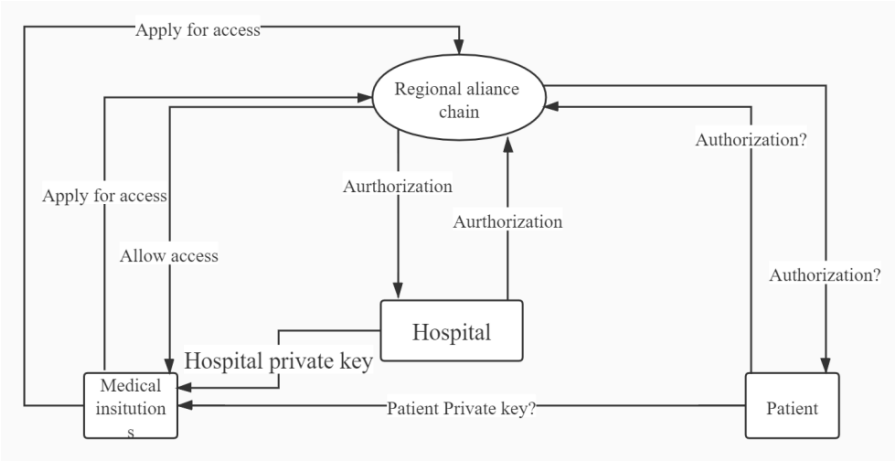

Fig. 2. Quartet transaction subject model

The four-party transaction entity model based on the regional medical alliance of patients, hospitals, and medical institutions is shown in Figure 2. The specific permissions of these four transaction subjects are shown in TABLEIII, and the internal model of the regional alliance chain is shown in Figure 3.

As we all know, the premise of out-of-hospital circulation of electronic prescriptions is that the rationality and safety of the prescriptions must be guaranteed, so as shown in Figure 4 in the flow chart to complete the prescription review process through the tertiary hospital prescription platform or regional electronic prescription service center. The hospital prescription platform, e-prescription service center, and pharmacy are arranged as nodes of blockchain to realize the uploading of diagnosis and treatment information, prescription examination information, drug purchase decision, and payment and distribution information, so as to ensure the compliance of the identity of the prescribing doctor, the confidentiality of patients' private information, the absence of tampering of prescription contents. and the traceability of the circulation process during the out-of-hospital circulation of e-prescriptions.

TABLE III COMPETENCE OF THE FOUR TYPES OF TRADING ENTITIES

	patient	Medical institutions	Hospital	Third party
Read and write access to own medical data	Have Permission	Have Permission	Have Permission	Have Permission
Read access to other medical data permissions	There is no permission by default, and you can get permission with the consent of the account owner	In special circumstances such as emergency situations, medical data can be read without authorization Generally, access to the default settings is only allowed	Have Permission	There is no permission by default, you can get permission with the consent of the account owner.

		with the consent of the account owner.		
White access to other medical data	There is no permission by default, and you can get permission with the consent of the account owner	There is no permission by default, and you can get permission with the consent of the account owner	Have Permission	There is no permission by default, and you can get permission with the consent of the account owner

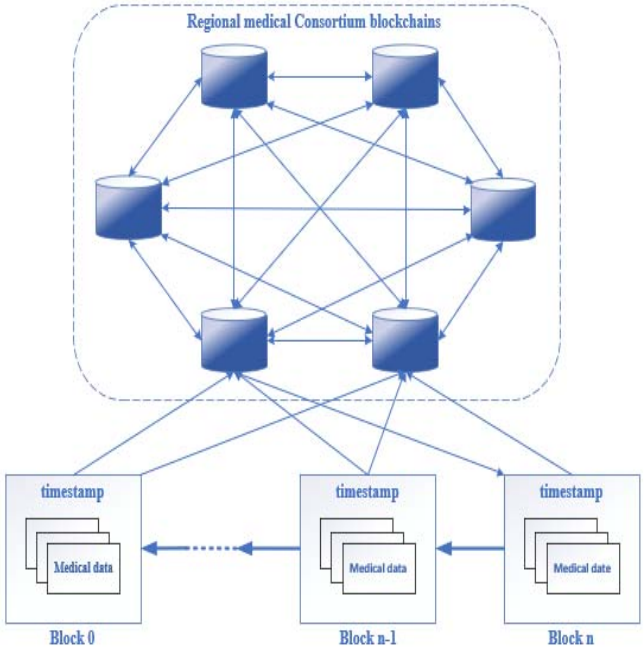

Fig. 3. Regional medical alliance chain mode

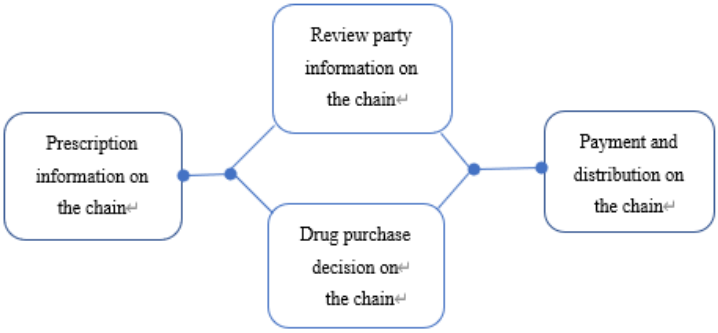

Fig. 4. Flow chart of prescription circulation

V. Design Of Smart Contract In Blockchain System

In terms of the combination of blockchain storage technology and medical information systems with their own characteristics, we design smart contracts to implement blockchain medical information systems, including blockchain medical information system contracts, patient contracts, physician contracts, researcher contracts, and medical record treatment contracts. We use the relationship between contracts to describe the logical structure of the blockchain health care system, representing the data to be stored and the services that can be provided in the whole system.

Figure 5 shows the overall relationship structure of the contracts in the system, including the blockchain medical information system contract (MISBC), the patient contract (Patient), the physician contract (Physician), the researcher contract (Researcher), and the case treatment contract (Case), as well as the various links between them.

Blockchain medical information system contract. The medical information system based on block chain (MISBC) contract is used to represent the entire blockchain medical system, storing information about all patient cases, as well as storing basic information about all doctors registered to the system, and providing access to this information for other users. The contract also stores the basic information of all researchers enrolled in the system to support the management of researcher information and access rights, as well as to provide information services to researchers. In addition, it provides the necessary access mechanisms for accessing publicly available case information. Anyone using this contract should first access this contract.

Patient contract. The patient contract is used to store the basic information of the patient user, and the patient can store and manage its medical record information through this contract.

Case Treatment Contracts. The case treatment contract can be used to store the relevant medical record information that the patient needs to store, as well as the diagnosis and treatment information of the patient by the doctor, which are generally stored in the block by the patient himself. Such information is generally stored in the block by the patient himself. The medical record information is divided into public medical record information and non-public medical record information, and the non-public medical record information is encrypted by private key. Usually, only the patient himself has the permission to use his medical record information, but if necessary, authorized doctors in the hospital can view the information stored on the chain. For the medical record information in which the information is publicly available, researchers in the system can have permission to view and use such information. We have deployed privacy protection algorithms to protect the patient's private information to a certain extent when the data is transmitted in pharmacies and healthcare facilities in the future.

Doctor contract. The doctor contract is used to store the basic information of the doctor user in the system, and can query the relevant information of the doctor member list stored in MISBC.

Researcher contract. Researcher contracts are used to store the basic personal information of related researchers. This contract is used to store and manage the information of researchers who need to investigate related public cases in the system. It can also be used to query related information through the list of researchers stored in MISBC.

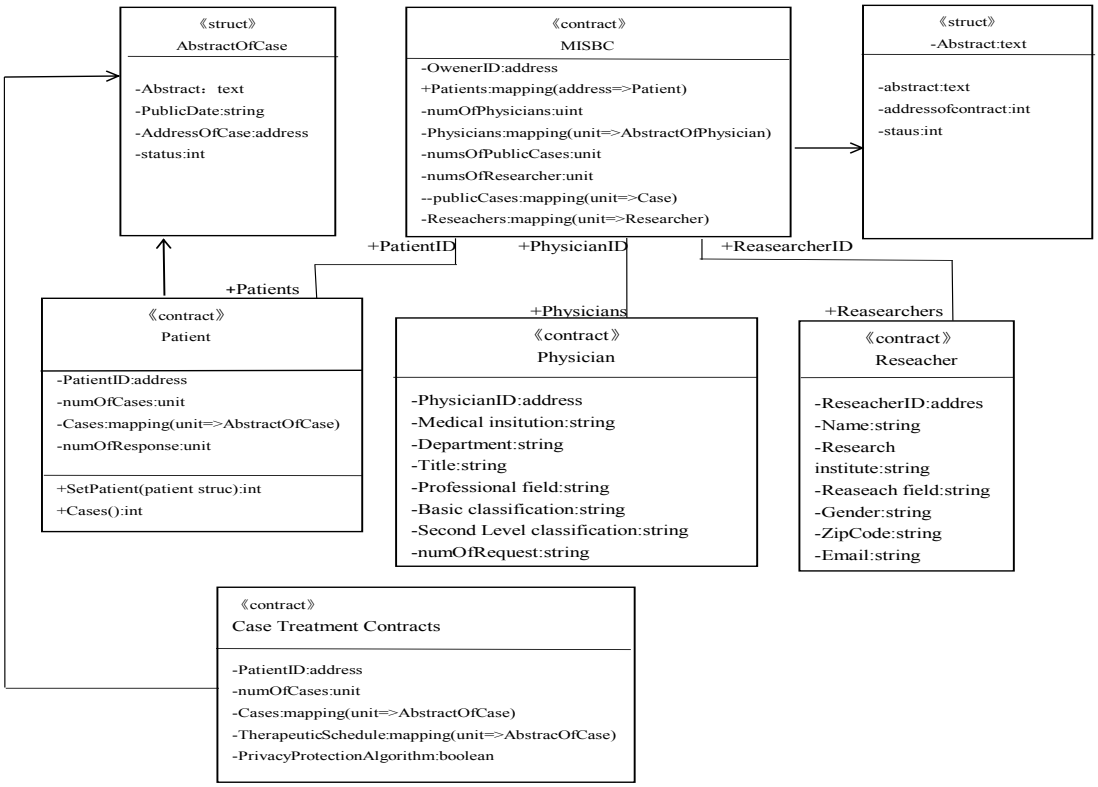

Fig.5.Blockchain medical system structure diagram

VI. Experiments And Analysis

The data set used in this experiment is "ADULT", which is derived from the US Adult Population Census. The data set has 30,162 instances and 9 attribute values. The experiment uses Java language programming. In order to verify the advantages of our proposed method in terms of data availability and security, the following will analyze the experimental results of histogram release and data security risk assessment.

A. Histogram Release

Histogram is one of the main techniques for approximate estimation data release. Taking the Boolean-type sensitive attribute Marital-status (value: Spouse-present or Spouse-not-present) as the object, and publishing statistics on the number of people whose marital status is spouse (Spouse-present) at different ages, we analyze the publication accuracy of the privacy protection methods in the article. Figure 6 shows the comparison results of the histogram release of the original data and the output protection data.

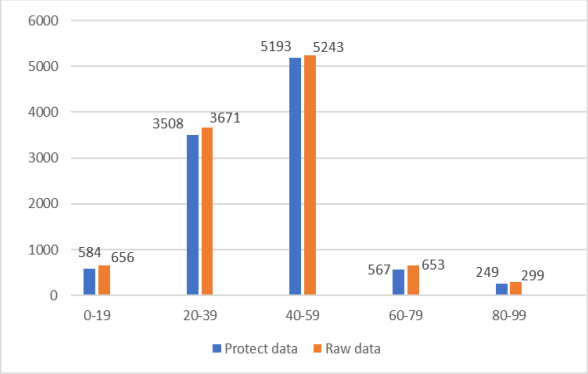

Fig.6 Comparison chart of statistics on the number of spouses in each age group

B. Security Risk Analysis

In the process of security risk analysis, in order to verify the advantages of the method in the article in terms of security performance, we assume two attack models. Model 1 assumes that the quasi-identifier attribute data in the data set has been known by the attacker; Model 2 assumes that the attacker has no background knowledge. TABLEIV shows the success rate of data sets being attacked under the k-anonymity, differential privacy protection, and data anonymity methods based on differential privacy protection under the above two attack models.

TABLE IV Risk Assessment Form

Methods	Model1	Model2
K-anonymity	0.194 05	0.047 69
Differential privacy	0.092 05	0.039 43
Method of this article	0.026 87	0.024 83

VII. Summary And Outlook

This paper designs a new data anonymization privacy protection method based on the existing privacy protection technology, and successfully deploys it in a blockchain-based medical information system. This method combines the generalized anonymity mechanism and the Laplace noise mechanism. The loss rate of data information is reduced, and the method can effectively prevent the attack of background knowledge. The experimental results show that the data anonymization method based on differential privacy not only improves data security, but also effectively guarantees data availability. In order to further improve the accuracy of the experiment, we can experiment by changing the model of the algorithm or using a larger data set.

References

[1] Singh . R, Singh. J, Singh. R "TBSD: A Defend Against Sybil Attack in Wireless Sensor Networks",International Journal of Computer Science and Network Security, vol. 16, no.11, pp. **90-99, 2016**

[2] Casadei. R, Fortino. G, pianini .D, Russo . W, Savalio . C, and Viroli . M,"A Development Approach For Collective Opportunistic Edge-Of-Things Services", Information Sciences, vol.498, pp. **154-169,2019**

[3] Randall, D., Goel, P. , & Abujamra, R. . (2017). Blockchain applications and use cases in health information technology. *Journal of Health & Medical Informatics, 08*(03)**.**

[4] AZARIA A,EKBLAW A,VIEIRA T,et al.MedRec:Using blockchain for medical data access and pemission management[J]. Repot on Health Information Blocking, 2016,11(5):25-30.

[5] TRIPATHY B K,MITRA A. An algorithm to achieve k-anonymity and l-diversity anonymisation in social networks[C]//International conference on computational aspects of social networks.Sao Carlos,Brazil:IEEE,2013:126-131

[6] Lefvre K ,DeWitt D ,Ramrkrishnan R. Incognito: efficient fulldomain k-anonymity[C],In Proc. Of the International Conference on Management of Data, NY,USA:ACM Press,2005:49-60

[7] Machanavajjhala A, Gehrke J,Kifer D. L-diversity:Privacy beyond k-anonymity[C] //Proc of the 22nd Int Conf on Data Engineering.Piscataway, NJ: IEEE,2006.24-36

[8] Agrawal R, Strikant R. Privacy-preserving data mining[C]// Proceeding of the 2000 ACM SIGMOD International Conference on Management of Data.Dallas,Texas,May **2000;439-450**

[9] Sweeney L K-anonymity:A Model for Protecting Privacy[J]. International Joumal on Uncertainty[J]. Fuzziness and Knowledge-based Systems,2002,10(5);557-57

[10] Dwork C,Lei J.Differential privacy and robust statistics[C]// ACM Symposium on Theory of Computing.**ACM,2009:371-380**

[11] FENTON N E,NEIL M. Software metrics:roadmap[C]// Conference on the futrue of software engineering.[s.1.]:IEEE,2000 ; 357-370.

[12] MENZIES T,GREENWALD J,FRANK A. Data mining static code attributes to learn defect predictors[J]. IEEE Transactions on Software Engineering,2006,33(1):2-13.

[13] PAN S J,YANG Qiang. A survey on transfer learning[J]. IEEE Transactions on Knowledge &Data Engineering,2010, 22(10):1345-1359.