Research on Medical Sensitive Data Protection Algorithm Based on Differential Privacy

Xiaofeng Li[1], Zhongwei Chen[1(✉)], and Zhichang Huang[2]

[1] College of Information Engineering, Guangxi University of Foreign Languages, Nanning 530222, China
top00112233@163.com

[2] Information Engineering College, Nanning University, Nanning 530200, China

Abstract. In order to avoid the wrong transmission behavior of medical data and realize the effective protection of sensitive information samples, the protection algorithm of medical sensitive data based on differential privacy is studied. According to the application principles of Laplace mechanism and index mechanism, a query control model is constructed, and then the analysis of differential privacy protection technology for medical sensitive data is realized by solving the security trust index. Based on the R-tree clustering structure, according to the sensitivity index calculation results, the search and processing of sensitive objects are completed, and the design of medical sensitive data protection algorithm based on differential privacy is completed. The experimental results show that under the effect of the principle of differential privacy, the error transmission probability of medical sensitive data will never be higher than 10%. It has strong application feasibility in solving the problem of medical data error transmission and effectively protecting sensitive information samples.

Keywords: Differential Privacy · Medical Sensitive Data · Data Protection · Query Control Model · Safety Trust Index · R-Tree Clustering Structure

1 Introduction

Since the birth and development of the Internet, it has greatly facilitated the lives of the masses. People can make mobile payments, e-shopping and surf the Internet through the Internet, but at the same time, more traces are left on the Internet. Under the circumstance of protection, relevant information will be collected and used by criminals, resulting in property damage and even life safety threats. With the development and progress of the times, various attack methods are emerging one after another. Although the traditional privacy protection technology has guaranteed sufficient protection to a certain level, it is unable to defend against background knowledge attacks. Differential privacy technology can not only analyze the data from big data to obtain effective information with unchanged statistical properties, but also protect the privacy information of users at the cost of certain loss of performance, regardless of the background attack.

S. Wang (Ed.): IoTCare 2022, LNICST 501, pp. 73–87, 2023.
https://doi.org/10.1007/978-3-031-33545-7_6

Through the technical disturbance and distortion of data, the balance between the effectiveness of data analysis and privacy protection is achieved [1]. With the rapid development of medical diagnostic technologies such as medical information systems and high-throughput sequencing, the medical and health field is gradually entering the "big data era". Medical big data includes basic data such as residents' behavioral health, electronic medical records, diagnosis and treatment data, detection reports, medical images, medical management, economic data, etc. It is characterized by large scale, rapid growth, diversified structure, and high application value. Therefore, medical sensitive data needs to be protected.

Reference [2] proposed a sensitive data protection algorithm based on sharding. This algorithm perturbs the data based on the idea of dynamic sharding, and completes the adjustment of data cluster size and sharded data. The isolated nodes are integrated into a set to avoid data interference. Reference [3] proposed a sensitive data protection algorithm based on Bayesian network. This algorithm uses clustering method to discretize data, and uses Bayesian model to protect data privacy. Although the above methods can complete the protection of sensitive data, there is a problem of low security factor, which makes the data easy to be stolen, intruded or tampered.

In order to improve the security of medical sensitive data, this research proposes a medical sensitive data protection algorithm based on differential privacy.

2 Differential Privacy Protection Technology for Medical Sensitive Data

The protection of medical sensitive data is based on differential privacy application technology. This chapter will study the practical methods of differential privacy protection technology from three aspects: Laplace mechanism and index mechanism, query control model, and security trust index.

2.1 Laplace Mechanism and Index Mechanism

The method of realizing the privacy protection of medical sensitive data in differential privacy protection is called "implementation mechanism". Laplace mechanism and index mechanism are the two most basic implementation mechanisms of differential privacy protection. Laplace mechanism is mainly used to protect numerical medical sensitive information results, and exponential mechanism is mainly used to protect non numerical data results, similar to geometric mechanism and Gaussian mechanism. Different noise mechanisms are suitable for different occasions, which are closely related to function sensitivity and privacy budget parameters, and provide differential privacy protection for different types of data.

Laplace Mechanism

The main method of Laplace mechanism to achieve differential privacy protection is to add random noise obeying Laplace distribution to the returned result of medical sensitive data query, so that the returned result after adding noise satisfies the differential privacy

protection constraint in formula (1).

$$Q = \frac{1}{\chi}\begin{cases} \exp\left(-\dfrac{e_1 - e_2}{\dot{W}}\right), e_2 < e_1 \\ \exp\left(-\dfrac{e_2 - e_1}{\dot{W}}\right), e_2 \geq e_1 \end{cases} \tag{1}$$

where, e_1 represents the location parameter of medical sensitive data, e_2 represents the scale parameter, $\dot{W}$ represents the sensitive eigenvalue of medical data, and χ represents the negative feedback vector.

For a given dataset $\Im$, there is a query function R, and its mean sensitivity is $\Delta\alpha$. Simultaneous formula (1) can express the differential privacy query principle of medical sensitive data based on the Laplace mechanism as:

$$\begin{cases} Y_R = \Delta\alpha \cdot Q + \left[\beta\left(\frac{\overline{r}^2}{r_1 \times r_2}\right)\right] \\ r_1 \in \Im \\ r_2 \in \Im \end{cases} \tag{2}$$

where, r_1 represents the sensitivity marking coefficient, r_2 represents the privacy query marking coefficient, $\overline{r}$ represents the average value of coefficients r_1 and r_2, and β represents the Laplace definition index.

Regardless of whether a medically sensitive data is recorded in or not in a private dataset, adding Laplace-distributed noise to the real query results has little effect on the final query results, that is to say, the attacker's same query differs between two records that differ by only one record. The probability ratio of the same outcome on the dataset is close to 1.

Index Mechanism

The index mechanism is mainly applicable to the case where the output result is non numerical. The result calculated by directly adding noise may damage the numerical characteristics. In order to select the "best" response, the index mechanism responds to any utility query (as well as any non numerical query) by designing a scoring function, while maintaining differential privacy protection. Let δ represent privacy protection parameters, ΔI represent the unit accumulation of medical sensitive data to be queried, and $\overline{U}$ represent the mean value of query vector. With the support of the above physical quantities, the simultaneous formula (2) can define the exponential mechanism expression based on differential privacy as:

$$Q' = \exp\left(\frac{Y_R}{\delta^2}\right) \cdot \Delta I \cdot \overline{U} \tag{3}$$

An exponential mechanism often provides a strong utility guarantee, as its effect decreases rapidly exponentially as the score decreases. The probability of returning an option with good availability by the index mechanism is limited by the privacy protection parameter δ. When the value of δ is large, the difference between the options is large, and the option with a higher score is more likely to be output.

2.2 Query Control Model

Considering the characteristics of query trust from two aspects: the query authority and reputation value of medical sensitive data and the privacy attribute of query data, a data hierarchical query control strategy using query trust to measure the level of the inquirer is proposed. The hierarchical query control processing is carried out on the data information stored in the data collection server combined with differential privacy protection technology, so as to provide different query users with available data with different accuracy. Differential privacy protection mechanism is widely used in privacy protection data publishing, data mining and other fields, aiming to protect individual sensitive data in the database when publishing data sharing. Define the query control model as:

$$P = \frac{\left[\phi\left[\hat{q}\right]^2 - \phi'\left[q'\right]^2\right]}{\varepsilon \times Q'} \tag{4}$$

where, ϕ represents the probability of medical sensitive data privacy being disclosed, $\hat{q}$ represents the disclosed data information, ϕ' represents the probability of medical sensitive data privacy being differentially processed, q' represents the differentially processed data information, and ε represents the discrimination coefficient of data information.

In the differential privacy protection mechanism, the difference of the queryer's authority and the privacy of the query data are considered, which can effectively prevent the privacy leakage threat caused by users with different authority levels querying sensitive data. In order to achieve the purpose of publishing available data information with different data accuracy rates for different levels of query users, a data hierarchical query control model based on differential privacy protection is proposed as shown in Fig. 1. Partial composition. Among them, hierarchical query control is the most important part, mainly including data preprocessing module, hierarchical query control module and usability analysis module.

In the process of publishing and sharing medical sensitive data, different query users have different needs and purposes for the published data. If they ignore the differences between the inquirers and provide data results with the same degree of privacy protection, users with low query authority will obtain information containing more sensitive data, resulting in the chain leakage of user sensitive data; Secondly, the privacy attributes of different sensitive data information are also different, and the data owners have different requirements for the degree of privacy protection of data. Each type of data will have different privacy requirements. The existing differential privacy protection research does not fully consider the impact of query user permissions, reputation values and data privacy attributes on privacy protection, which is easy to cause query users with low permissions and reputation to obtain high availability data, and then cause the indirect disclosure of user sensitive data.

The hierarchical query control module combines differential privacy protection technology and is responsible for hierarchical control of queries. Query based on data research and use needs to consider the influence of factors such as query authority and the privacy properties of query data. The solution is to perform trust quantification analysis according to the authority, reputation value of the data queryer, and privacy

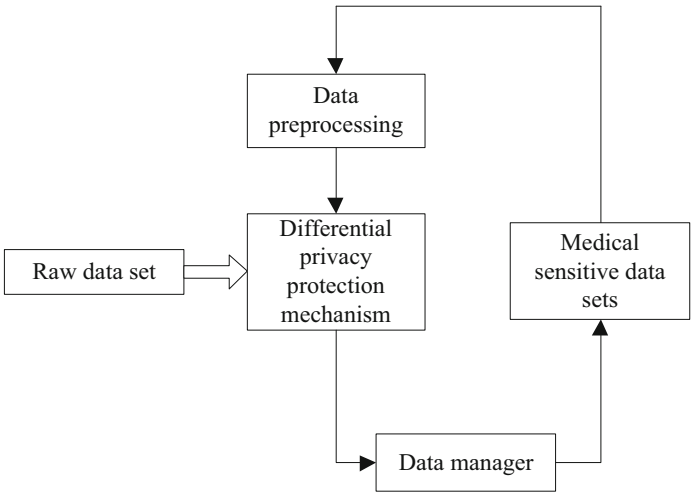

Fig. 1. Data Hierarchical Query Control Model

attributes of the data, and classify the corresponding levels to correspond to different privacy protection parameters, so as to achieve differentiated differential privacy protection effects [4]. The availability evaluation module measures the data availability after differential privacy protection processing, realizes controllable sensitive data privacy and availability protection, can effectively prevent malicious attackers from obtaining user privacy information through information query, and can greatly reduce privacy leakage. Increase the effective use of published data. Finally, the processed results are fed back to the query user.

Data asset management is to manage the data in the storage cluster according to certain rules, including data standard management, data quality detection, data operation monitoring, data resource directory, data security management and data value evaluation. Realize value realization through managed data. For example, the intelligent medical information management platform can be used for medical insurance quality control, semantic retrieval, etc., to improve medical quality. Manage the medical activities of the whole hospital through the president's cockpit, adjust the hospital resources in real time according to various reports, allocate medical resources reasonably, and solve the problem of waste of medical resources.

2.3 Calculation of Security Trust Degree

This section proposes the query security trust degree to quantify the security and trustworthiness of medical sensitive data. The query security trust degree is determined by the query subject attribute value and the query object attribute value. The larger the value, the more trustworthy. In this section, we mainly consider that the query security trust degree consists of two parts: the queryer's trust value and the data privacy attribute value. Its calculation publicity network:

$$O = \left(1 - \varphi^2\right)P + \gamma\vec{A} \tag{5}$$

where, φ represents the differential confidence index of medical sensitive data, $\vec{A}$ represents the privacy information storage vector, and γ represents the privacy query coefficient.

The weight distribution is determined by the data owner according to the data privacy protection requirements. For different privacy protection requirements, the weights of the queryer trust value and the data privacy attribute value can be adjusted to perform security trust calculation and data processing. For example, when the security trust degree does not consider the data privacy attribute, $\gamma = 0$. The queryer trust value reflects the subjective differences of data query users with different rights and reputation values, and the trust values corresponding to different query rights and reputation values are also different. The data privacy attribute value reflects the degree of privacy of different sensitive data and is usually determined by domain experts and data owners. To facilitate trust calculation, we quantify queryer authority and reputation values and data privacy attribute values as scalars between 0 and 1.

Calculation of querier trust value: the querier trust value is composed of query permission attribute and reputation value. When calculating the trust value of the inquirer, first analyze the authority value and reputation value of the inquirer according to the user attribute. The authority value and reputation value are random variables with normal distribution between 0 and 1. The specific calculation formula is as follows:

$$S = O^2 + (\lambda - 1)\frac{s_{\max}^2 - s_{\min}^2}{\dot{d}} \tag{6}$$

Among them, λ represents the query security item coefficient of medical sensitive data, $\dot{d}$ represents the trust measurement value, $s_{\min}$ represents the minimum value of the security confidence index of medical sensitive data, and $s_{\max}$ represents the maximum value of the security confidence index. As requests for access and public disclosure of health data continue to increase, and the line between privacy protection of personally identifiable information and aggregated data becomes increasingly blurred, laws and regulations alone cannot effectively constrain the privacy of personal information. Therefore, it is particularly important to use privacy-preserving technologies to improve information retrieval of medical big data. Privacy-preserving technologies can not only improve access management, monitoring, and control of data, but also improve the identification of personal health information, helping to assess and reduce the risk of re-identification.

Smart hospitals use intelligent technology, Internet technology and some AI technologies in the field of medical services, making it a new trend of modern medical development in China. Smart hospitals help hospitals integrate resources, optimize processes, reduce hospital operating costs, and improve service quality, management level, and work efficiency. Patients use the hospital introduction in the palm hospital to understand the hospital before seeing a doctor, and then directly locate the hospital location according to the navigation [5]. Learn about doctors through specialized consultation and introduction of famous doctors in palm hospital, and select appropriate doctors for registration and appointment according to their actual situation. After making an appointment, you can use real-time query to understand the treatment situation of experts, and arrange an appropriate time for medical treatment according to the current treatment situation, so as to save the waiting time of patients.

The security trust degree is a key indicator to judge the differential privacy performance of medical sensitive data, and the interaction is based on whether it meets a certain trust level. The division of trust levels mainly reflects the relationship between the level of trust in the inquirer, and it is not necessary to be too precise in specific applications. By analyzing the difference of the queryer's authority attribute and the dynamic change of the data privacy attribute, the query security trust degree is calculated and corresponding to different trust levels and different privacy protection parameters, so that when the sensitive data privacy attribute changes, the query users of different levels can obtain different results. Availability of privacy-preserving data to achieve hierarchical control of the accuracy of sensitive data information in query responses under the premise of protecting privacy.

3 Medical Sensitive Data Protection Methods

Based on the application principle of differential privacy, according to the processing flow of R-tree clustering structure construction, sensitivity calculation and sensitive point search, the design of medical sensitive data protection algorithm based on differential privacy is completed.

3.1 R-tree Clustering Structure

In order to achieve effective access to medical sensitive data, a clustering model based on R-tree index is used for data retrieval of intelligent medical information system. When constructing the R-tree clustering model, if the distribution law of data is unknown, setting the clustering center in advance will make the final clustering result deviate from the reality, thus affecting the efficiency of the constructed R-tree model index [6]. In order to effectively determine the clustering center, DCC algorithm is introduced to construct R-tree model. The R-tree model can accurately calculate the distance between medical sensitive data, improve the clustering accuracy of medical sensitive data, and help improve the integrity and security of medical sensitive data protection.

Set the distance index of measuring adjacent sensitive data as l, which is expressed as:

$$l = \frac{S}{\sqrt{\frac{f}{G}}} \tag{7}$$

In the formula, f is the quantity of medically sensitive data, and G is the range of a given spatial area. Using the dynamic R-tree generation algorithm, reasonable leaf nodes are inserted into the target object, and the above-mentioned dynamic determination of the cluster center algorithm is used to construct the large-scale dynamic optimization of the R-tree. The R-tree generation for either spatial dataset is shown in Fig. 2. The main process of its establishment is as follows: First, establish the minimum bounding rectangle for all spatial objects. Then the base rectangles are grouped according to the algorithm of dynamically determining the cluster center. For example, in Fig. 2, the R12 closest to the mean point is first selected as the initial cluster center. Then select R19

farthest from R12 and R8 farthest from R19 as the cluster center and start clustering. Among them, R13, R14, R17 and R18 are divided into R19, R9, R10, R11, R12, R15 and R16 are divided into R8. At this time, two clusters are formed, and the cluster center and cluster measurement function are calculated. Select the cluster with the largest radius and its cluster center R12 from the two clusters, select R15 with the farthest distance from R12 and R11 and R18 with the farthest distance from R15 as the cluster center to re-cluster, and then calculate its cluster center and the clustering measure function. The loop repeats until the clustering function is known to converge.

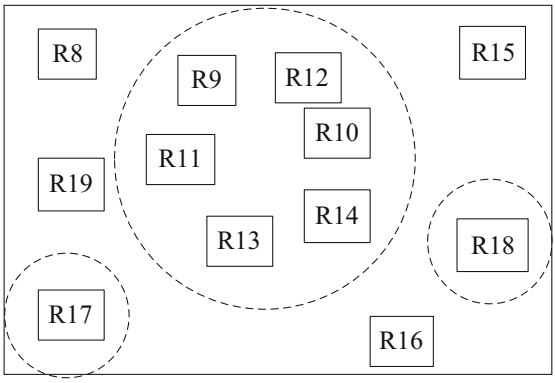

Fig. 2. R-tree structure clustering center

Although using big data can improve medical quality and solve current problems, its unique characteristics also bring challenges to technology and management. For example, big data has many sources and formats, from early images and text to video, audio, graphics and even GPS data from mobile phones. Complex data forms greatly increase the difficulty of storing, mining and analyzing data. Because many hospitals have digitized their administrative and treatment procedures, the speed and quantity of data generation exceed the limitations of traditional data processing software. In the database system, when the stored objects are very large, the efficiency of data query and retrieval is an important bottleneck that restricts the application of medical information. Therefore, improving the information retrieval ability in the context of medical big data is of great significance to improve the level of medical services and promote the construction of medical informatization.

With the development of medical informatization, the traditional paper medical records are gradually replaced by electronic medical records. However, for special medical records, paper medical records still have an irreplaceable position. Users can access real-time through personal computers or mobile phones, providing users with convenient and fast information query services. When the medical record room manages these paper medical records, they can quickly query relevant information through the smart hospital system to complete the management of paper medical records [7]. Medical institutions are responsible for the management and protection of personal information, and bear the corresponding legal responsibilities and risks when processing these information. In

the face of increasingly complex data processing needs, it is very important to establish and improve relevant laws to meet the reasonable requirements for data processing.

3.2 Sensitivity

Differential privacy protects the user's private information by adding noise to the query results. The amount of noise added is the key. It is required that the added noise can not only protect the user's privacy, but also prevent the data from being added too much. The noise makes the data unavailable. Function sensitivity is an important parameter to control noise. By controlling the amount of noise generated by global sensitivity, a privacy protection mechanism that meets the requirements of differential privacy can be implemented.

The global sensitivity for a query data h is defined as follows:

$$K(h) = \max_{\iota=1} l \times [F(h) - X(h)] \tag{8}$$

where, ι represents the tuple definition coefficient, $F(h)$ represents the privacy query function based on coefficient h, and $X(h)$ represents the differential definition function based on coefficient h. Global sensitivity reflects the upper limit of the output that can be disturbed when protecting privacy. It is independent of the data set and is determined by the query function itself.

When solving the sensitivity index of medical data information, the cooperation of the following equipment structures is also required.

(1) Information perception layer: It is a physical layer that includes sensor networks, sensors, data collectors, etc., which are used to store parameters or identifiers in object events. Its task is to complete the acquisition, communication and coordination of data.
(2) Network Cognitive Layer: Provides connection and transmission between objects, network devices, wireless or wired and cloud systems, and sends and processes data obtained locally; it also includes data received from the perception layer. Gateway component [8]. Starting from the information perception layer, processing hosts can obtain a large amount of heterogeneous perception information, and use the interconnection mechanism between networks to distribute and share these perception information.
(3) Intelligent application layer: responsible for providing applications and services to human or non-human users. It can specify and locate various application processes, programs and application software in the Internet of things. For example, intelligent medical record management system, medical image management system, etc.

Then the geometric data table and related attribute data table are designed for the hierarchical geometric object set. The geometry table consists of the unique identifier of the geometry object and two binary coordinate fields (strings). Accordingly, the unique identifier of the object and several attribute fields are designed in the attribute data table, and the connection between the geometric data table and the attribute data table is realized through the unique identifier.

Although privacy-preserving schemes emerge in an endless stream, they all have a common disadvantage, all relying on the background knowledge of the attacker and not making reasonable assumptions about the attack model. The emergence of the differential privacy model effectively solves this problem. The concept of differential privacy comes from cryptography and aims to maximize the accuracy of data queries and minimize the chances of records being identified when querying from statistical databases. The differential privacy model ensures that the public results will not change significantly because of whether an individual is in the data set by adding random noise to the data set, and provides a quantitative model for the degree of privacy leakage [9]. Because the change of an individual will not have a significant impact on the data query results, the attacker cannot infer the private information of individual samples from the public results based on their own background knowledge. Differential privacy model does not depend on the background knowledge possessed by attackers and provides a higher level of semantic security for private information, so it is widely used as a new type of privacy protection model.

3.3 Sensitive Point Search

Because the data collected by the medical Bracelet changes in real time, its data usually remains almost constant or increases (or decreases). For example, when collecting and monitoring the vital signs of patients, their ECG and blood pressure are almost unchanged in general, but they will increase or decrease in unstable periods. Therefore, the goal of the first stage of collecting such health data is to search for sensitive points when the trend begins to change.

In order to better protect the privacy information in medical and health data, after analyzing the data, we further develop an anonymous privacy publishing model suitable for medical and health data. Due to the correlation between the attributes of medical and health data, publishing all attributes in the same data table and then surfacing this data to all users will increase the risk of privacy disclosure [10]. One of the principles of privacy protection is to reduce the number of people who have access to privacy. According to the data characteristics of medical and health data, two different publishing methods are proposed:

1. For some data users with lower authority, only data tables consisting of quasi-identifiers and main sensitive attributes, such as gender, age, and disease, are released, thereby reducing the probability of information leakage. Only disease is a sensitive attribute in the medical and health data table released at this time, so it is necessary to formulate a single-sensitive attribute privacy protection scheme suitable for medical and health data. The released single-sensitive attribute data sheet can not only reduce the probability of link attack and homogenization attack, but also have good performance against background knowledge attack and similarity attack.
2. For another part of data users with high authority, the data table includes not only multiple quasi identifiers and primary sensitive attributes, but also some secondary sensitive attributes, such as the attending doctor and the treatment method adopted. At the same time, we also need to consider the sensitivity of the main sensitive attributes of disease and the correlation between the attributes. For this kind of data, it is

necessary to formulate an appropriate privacy protection scheme for multi sensitive attribute medical and health data, so that the published data can not only deal with the semantic attacks that single sensitive attribute data may face, but also have the ability to resist sensitivity and association attacks.

Let J represent the assignment coefficient of medical sensitive data, ΔZ represent the unit accumulation of the data text to be processed, and ϖ represent the sensitivity feature measurement index. With the support of the above physical quantities, formula (8) can be combined to express the sensitive point screening index M as:

$$M = K(h) + J\left(\frac{\Delta Z}{\varpi^2}\right) \tag{9}$$

The similarity or dissimilarity between the values of each sensitive attribute can be calculated according to the position of the disease in the semantic hierarchy tree. For a given disease value, if the distance between other diseases and it is calculated, first find the specific positions of the two disease values in the semantic hierarchy tree, and traverse the hierarchy tree upwards according to their positions until the minimum of the two disease values is found. Common parent node, and then calculate the distance from the leaf node where the disease value is located to its smallest common parent node. On the basis of formula (9), let c_1, c_2, $\cdots$, and c_n E represent the definition coefficients of n different medical sensitive data cotyledon nodes, ϑ represents the sensitivity measurement index, $\tilde{b}$ represents the fluctuating transmission characteristics of medical sensitive data, and σ represents data processing vector.

The search results of medical data sensitive points based on differential privacy are:

$$V = (c_1 + c_2 + \cdots + c_n)\frac{\vartheta \cdot M}{1 + \left|\tilde{b}\right|^{-\sigma}} \tag{10}$$

Combining the concept of differential privacy to develop a reasonable and available medical sensitive data protection algorithm is the research focus. In the process of processing, the research on the privacy protection scheme of single sensitive attribute and multi sensitive attribute medical health data will be carried out respectively, in order to reduce the attack vulnerability of the published medical health data table. The effectiveness of the scheme is proved by experiments. In addition, on the basis of privacy protection, how to minimize the loss of information in the process of anonymity is also one of the issues that need to be considered in the formulation of privacy protection schemes.

4 Case Analysis

In order to verify the practicability of the differential privacy-based medical sensitive data protection algorithm, the following comparative experiments are designed.

Step 1: Implant a sensitive data protection algorithm based on differential privacy in a medical operation system, use the system to normally store daily medical information, analyze the changes of relevant index parameters, and record the obtained data as the variable index of the experimental group;

Step 2: Implant the conventional data protection algorithm in the medical operation system, use the system to normally store daily medical information, analyze the changes of relevant index parameters, and record the obtained data as the variable index of the control group;

Step 3: Compare the variable indicators of the experimental group and the control group, and summarize the experimental rules.

The layout of the complete medical operation system is shown in Fig. 3.

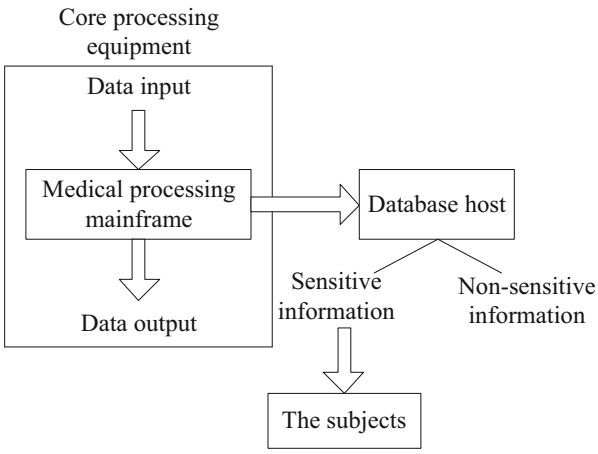

Fig. 3. Layout of medical operation system

In the medical operation system, the transmission effectiveness of sensitive information samples can be used to describe the processing ability of the medical host for data objects. Without considering other interference conditions, the stronger the transmission effectiveness of sensitive information samples, the stronger the processing ability of the medical host for data objects. During this experiment, if the sensitive information sample can be successfully transmitted to the target location, the coding result will be recorded as "1"; If the sensitive information sample cannot be successfully transmitted to the target location, the coding result will be recorded as "0". The specific experimental results are shown in Table 1.

Analysis of Table 1 shows that the algorithm in the experimental group has relatively weak coding ability for equipment usage indicators. Among the 10 variable indicators, only the coding result of this type of medical sensitive data is "0"; The algorithm in the control group has department information, drug information, the coding ability of the three types of indicators of treatment duration is weak. Among the 10 variable indicators, the encoding result of the above three types of medically sensitive data is "0".

Table 1. Encoding results of medically sensitive data

Data classification	The experimental group	The control group
The patient information	1	1
Department of information	1	0
Medical information	1	1
The fee information	1	1
The doctor's advice information	1	1
Return information	1	1
Drug information	1	0
Equipment usage	0	1
Drug dosage	1	1
The treatment time	1	0

The ability of the selected protection algorithm to transmit medical sensitive data can be expressed as:

$$\begin{cases} \eta_1 = \dfrac{M_1}{M_0} \times 100\% \\ \eta_2 = \dfrac{M_2}{M_0} \times 100\% \end{cases} \tag{11}$$

where, η_1 represents the effectiveness index of the experimental group, M_1 represents the total amount of sensitive data with the coding result of "1" in the experimental group, η_2 represents the effectiveness index of the control group, M_2 represents the total amount of sensitive data with the coding result of "1" in the control group, and M_0 represents the total amount of medical sensitive data samples.

Combine Table 1 and formula (11) to calculate the η_1 and η_2 indicators. Without considering other interference conditions, it can be seen that $\eta_1 = 90\%$ and $\eta_2 = 70\%$, the effectiveness of the experimental group is significantly greater than that of the control group.

After the above experiments are completed, in order to further verify the performance of the algorithm in this paper, the security factor is taken as the indicator to test the protection security of the algorithm in this paper. Set the maximum safety factor to 1.0. The higher the safety factor, the stronger the protection performance of the algorithm. The algorithm in this paper is compared with the algorithm in reference [2] and the algorithm in reference [3]. The comparison results of the safety factors of the three algorithms are shown in Table 2.

It can be seen from the comparison results of the safety factors shown in Table 2 that, with the increase of the number of experiments, the safety factor of the algorithm in this paper is significantly higher than that of the two literature comparison algorithms. The maximum safety factor of the algorithm in this paper reaches 0.98, while the maximum

Table 2. Safety factors

Number of experiments	Safety factor		
	Algorithm in this paper	Reference [2] algorithm	Reference [3] algorithm
10	0.95	0.72	0.83
20	0.98	0.56	0.84
30	0.96	0.78	0.80
40	0.94	0.69	0.74
50	0.96	0.71	0.71
60	0.97	0.83	0.64
70	0.98	0.73	0.62
80	0.95	0.76	0.68
90	0.96	0.68	0.64
100	0.98	0.65	0.69

safety zone coefficient of the algorithm in reference [2] and the algorithm in reference [3] reaches 0.9. Therefore, the data protection security of this algorithm is high.

5 Conclusion

In order to improve the security of medical sensitive data, a medical sensitive data protection algorithm based on differential privacy is proposed. The performance of the algorithm is verified from both theoretical and experimental aspects. The algorithm has high coding and transmission performance when protecting medical sensitive data, and improves the security of medical sensitive data. The experimental results show that, compared with the protection algorithm based on fragmentation and Bayesian network, the algorithm in this paper can improve the security factor of medical sensitive data and improve the security of medical sensitive data.

Aknowledgement. 2020 Guangxi College and University Young and Middle aged Teachers' Basic Scientific Research Ability Improvement Project "Research and Development of Panoramic Campus Roaming System Cross platform" (2020KY63022).

References

1. Kou, L., Liu, N., Huang, H., et al.: Data aggregation privacy protection algorithm based on distributed compressive sensing and hash function. Appl. Res. Comput. **37**(01), 239–244 (2020)
2. Jun, W., Xu, Y., Li, L.: Lightweight data fusion privacy protection algorithm based on sharding. Comput. Eng. Des. **43**(05), 1207–1213 (2022)

3. Xiao, B., Yan, H., Luo, H., et al.: Research on improvement of Bayesian network privacy protection algorithm based on differential privacy. Netinfo Security **20**(11), 75–86 (2020)
4. Kshetri, N., Voas, J.: Thoughts on general data protection regulation and online human surveillance. Computer **53**(1), 86–90 (2020)
5. Qaisar, S.M., Alsharif, F., Subasi, A., et al.: Appliance identification based on smart meter data and event-driven processing in the 5G framework. Procedia Comput. Sci. **182**(3), 103–108 (2021)
6. March, R.D., Leuzzi, C., Deffacis, M., et al.: Innovative approach for PMM data processing and analytics. IEEE Trans. Big Data **6**(3), 452–459 (2020)
7. Silva, M., Kaesler, J.M., Reemtsma, T., et al.: Absorption mode spectral processing improves data quality of natural organic matter analysis by Fourier-transform ion cyclotron resonance mass spectrometry. J. Am. Soc. Mass Spectrom. **31**(7), 1615–1618 (2020)
8. Ttsch, K., Fjeldsted, J.C., Stow, S.M., et al.: Effect of sampling rate and data pretreatment for targeted and nontargeted analysis by means of liquid chromatography coupled to drift time ion mobility quadruple time-of-flight mass spectrometry. J. Am. Soc. Mass Spectrom. **32**(10), 2592–2603 (2021)
9. Wielgat, R., Jdryka, R., Lorenc, A., et al.: POLEMAD–A database for the multimodal analysis of Polish pronunciation. Speech Commun. **2020**(127), 29–42 (2021)
10. Zhou, A.I.: Research on weighted social network deep differential privacy data protection algorithm. Comput. Simul. **37**(10), 282–285+373 (2020)