Research on Secure Storage of Healthcare Data in the Environment of Internet of Things

Haipeng Ke[1], Juanfen Shi[2(✉)], and Tianlin Fu[3]

[1] Fujian Zhangzhou No. 1 Vocational Secondary School, 363000 Zhangzhou Fujian, China
[2] School of Electronic Engineering, Henan Information Engineering School, Henan 450008, China
hijuanfen@163.com
[3] College of Mathematics and Data Science, Minjiang University, Fuzhou Fujian 350000, China

Abstract. In order to improve the security and storage efficiency of medical care data, a safe storage method of medical care data in the Internet of Things environment is designed. The security analysis of medical care data resources in the Internet of Things environment, and the establishment of trusted identifiers for storage resource provision and use requests, to ensure data security to the greatest extent. Compliance verification of storage resource provision and use requests can solve the problem of data security degradation caused by simple data access methods. For data distance calculation and data division, establish a spatial database, and use distributed storage to realize the safe storage of medical and health care data, in order to solve the problem of low reliability of data storage. The experimental results show that the secure storage method of health care data in the Internet of Things environment studied in this paper not only improves the efficiency of data encryption and retrieval, but also reduces the number of times that data is attacked and stolen, and has a better effect of data secure storage.

Keywords: Internet of Things Environment · Health Care Data · Safe Storage · Constraint Parameters · Trusted Identity · Verification

1 Introduction

With the development of the Internet of Things and the continuous integration of the Internet of Things, the Internet and 3G mobile phones, the degree of intelligence of the Internet of Things will become higher and higher in the future. At the same time, the impact of Internet of Things on security cannot be ignored. The theft of Internet of Things signals will directly affect the information security of the entire Internet of Things. From the perspective of information security and privacy protection, the widespread introduction of Internet of Things terminals not only provides richer information, but also increases the risk of exposing such information [1]. The patient's electronic medical record contains a large amount of privacy information, such as the patient's condition information, consumption records, etc. these information will be centrally stored on the hospital's server, and most hospitals do not attach great importance to information

S. Wang (Ed.): IoTCare 2022, LNICST 501, pp. 271–288, 2023.
https://doi.org/10.1007/978-3-031-33545-7_19

security. There are countless vulnerabilities in the hospital website, so that attackers can easily get the hospital's database and patient information, resulting in the privacy disclosure of users, that is, there is a risk of data disclosure. At the same time, if the attacker deliberately destroys and tampers with the data, it will seriously hinder the availability of the medical system, that is, there is a risk of data tampering.

Some progress has been made in the research on data security storage methods. Reference [2] proposed a data security storage method based on blockchain. In this method, wireless sensor networks and blockchain technologies are used to construct the distribution framework of federated chain sensing nodes and the data storage federated chain model, and data blockchain is formed through the consensus of data acquisition base stations Specific types of hash chains are formed according to data attributes, hash values and other information to facilitate the safe storage of the same type of historical data. Reference [3] proposed a method for medical data security storage based on hybrid encryption. The classical data encryption standard (DES) and asymmetric encryption algorithm (Rivest Shamir Adleman, RSA) are analyzed, and a hybrid encryption method for medical data security storage enhancement is proposed An improved algorithm IBDES is proposed to enhance the security strength through double encryption, and an improved algorithm EPNRSA is proposed to reduce the time complexity of RSA encryption while ensuring the security quality of encryption. An enhanced hybrid encryption method for medical data based on IBDES and EPNRSA is formed to realize the safe storage of data. Reference [4] proposed a design method of hospital financial data security storage system based on homomorphic encryption algorithm. Under the 3-tier hardware structure, two IBMP5570 minicomputers are used to handle the same transaction in a dual cluster mode. Four IBM 3850 servers are used and RADWARE load balancers are configured to automatically distribute workload. Use dual storage disk array cabinets to back up data with each other. Configure SAN switch for data exchange between server and storage. All homomorphic encryption is used to encrypt, decrypt and retrieve the ciphertext files to complete the safe storage of data.

However, the application of the above methods to the safe storage of medical care data has the problems of the security of medical care data and low storage efficiency. In order to solve the problems of traditional methods, a safe storage method of medical care data in the Internet of Things environment is designed to improve the security of hospital data. The main structure of this paper is as follows:

(1) To analyze the security of medical care data resources, on the one hand, the existing asymmetric encryption algorithm is used to verify the identity authenticity of the system participants' nodes, so as to ensure the credibility of transactions on the chain. On the other hand, we design the data structure of the storage resources on the blockchain to provide information and use the request information and carry out trusted identification for them, and design compliance verification methods for the two kinds of information to enhance the credibility of the storage designed in this paper.
(2) Through the analysis of the trusted release of storage resources in the above process, the uniquely identified storage resources are disseminated to other nodes in the P2P network in the form of messages for a certain period of time.

(3) The distance function is established to reasonably divide the spatial data in consideration of the proximity and topological relationship between the spatial element objects, so as to establish a distributed spatial database. The distributed storage method is used to realize the safe storage of medical and health data.
(4) The effectiveness of this method is verified by experiments.
(5) Summarize the full text and draw conclusions.

2 Trusted Identification and Verification of Storage Resource Provision and Use Requests

The research work on the trusted identification and compliance verification of the information provided by the storage resources and the use of the requested information in the network is mainly carried out from two aspects: first, this paper will use the existing asymmetric encryption algorithm to verify the identity authenticity of the system participant nodes, so as to ensure the credibility of the transactions on the chain; The second is to design the data structure of the storage resources on the blockchain to provide information and use the requested information, and carry out trusted identification on it, and design compliance verification methods for the two kinds of information respectively to enhance the credibility of the storage designed in this paper.

2.1 Data resource security analysis.

Before extracting healthcare data, the security of data resources needs to be analyzed, and the main process is as follows:

Represent a hypersphere containing normal sample points as:

$$\begin{gathered} \min R^2 \\ s.t. x_i - c^2 \leq R^2, i = 1, \ldots, N \end{gathered} \tag{1}$$

In the above formula, R represents the radius of the hypersphere, c represents the spherical center vector, and x_i represents the distance to the spherical center vector.

Since the distance from the data to the spherical center vector is not necessarily less than or equal to the radius, it needs to be corrected. The formula is as follows:

$$\begin{gathered} \min R^2 + C\sum_{i=1}^{m}\xi_i \\ s.t. x_i - c^2 \leq R^2 + \xi_i, \xi_i \geq 0, i = 1, \ldots, N \end{gathered} \tag{2}$$

In the above formula, C represents the regularization coefficient, and $\xi_i$represents the slack variable.

After the correction, set the constraints as follows:

$$\begin{aligned} f_1(k) &= \begin{cases} 1, k = 1, \ldots, N_A \\ 0, k = N_A + 1, \ldots, N_A + M^A \end{cases} \\ f_2(k) &= \begin{cases} 0, k = 1, \ldots, N_A \\ 1, k = N_A + 1, \ldots, N_A + M^A \end{cases} \end{aligned} \tag{3}$$

where, N_A represents the constraint parameter of the A data, and M^A represents the constraint parameter of the A data.

After the above process, the principal components of the information are obtained. During retrieval, it is necessary to establish an inverted index of key features, according to which the location of the information can be quickly detected [5]. In the establishment, all parameter values are searched through the storage address in advance, and then the keyword search is carried out. When a search is completed, the final search results are transmitted to the interactive channel. Retrieval edge weight is an important physical index to measure the retrieval ability of key features of information. The size of edge weight is directly related to the strength of node interaction. The retrieval edge weight value is expressed as:

$$g = 1 - \frac{\left| \dot{k} - \sqrt{(I - U')^x} \right|}{\left| f' \cdot h^2 / l \cdot \bar{\lambda} \right|} \tag{4}$$

In the above formula, l represents the upper limit matching condition parameter of the information in the wireless network, U' is the lower limit matching condition of the information, x represents the statistical coefficient of the power term, h represents the programmed retrieval vector, and l and λ represent the average interaction constant of information nodes respectively. And information priority.

The above process retrieves the key features of the data, and on this basis, processes the data to analyze the security of the data resources. The specific process is as follows:

Step1: Using unsupervised clustering algorithm, allocate nodes, assume that the collected sample set is $M = (m_1, m_2, m_3, \ldots, m_n)$, define set as $A(I)$ to accumulate the sum of sample vectors belonging to various types, and define one of the counters $B(I)$ to register the number of corresponding normal sample categories;

Step2: Divide the cost function in the unsupervised clustering algorithm and calculate the density index of the collected data sample points. The calculation formula is as follows:

$$D_u = \sum_{i=q} d \exp\left(\frac{\left\| x_i - x_j \right\|}{d/2} \right) \tag{5}$$

In formula (5), $\sum_{i=q} d$ is the density index of node i, $\left\| x_i - x_j \right\|$ is the calculation factor of the density index size of the sample point, and exp is the neighborhood radius of the collected data.

Step3: Calculate each remaining sample point above and determine the Euclidean distance of the unsupervised clustering center. If $r \leq d_2$, this sample point will be classified into the corresponding class of the unsupervised clustering center. If $r \geq d_2$, it will not be classified temporarily, where d_2 is the preset threshold. The larger the value, the more the number of clusters, and vice versa;

Step4: Take the unclassified samples in Step3 as a new sample set, re execute the above processes of step1, Step2 and Step3, and repeat this cycle to make a safety judgment on all data [6];

Step5: Calculate the security situation category of each data, and the calculation formula is as follows:

$$c_i = \frac{E(i)}{B(i)}, i = 1, 2, \ldots, n \tag{6}$$

In formula (6), $E(i)$ represents the initial data center in unsupervised clustering, and $B(i)$ is the maximum distance between data nodes in the network.

According to the above process, complete the data clustering [7], according to the clustering results, the data security analysis, the calculation formula is as follows:

$$[D' = RT\frac{a_i}{f * |d|} \tag{7}$$

In formula (7), RT represents data topology, f represents abnormal parameters, $|d|$ represents protection strategy set, and a_i represents collected operation data.

Predict and collect operation data, and then normalize the data to provide the basis for subsequent data storage.

2.2 Trusted Identities for Storage Resource Provision and Use Requests

How to complete the trusted identification of storage resource information and storage resource usage requests in a decentralized cloud storage system needs to be studied and designed from two aspects. On the one hand, it is determined that the node identity verification method of the Ethereum system is adopted in this solution to realize the authenticity verification of the node identity in the decentralized system. Participating nodes in the system can join the blockchain network through an automatically generated key, and obtain an account address that can uniquely identify the node. Based on the above, the trusted identification of storage resource provision and storage resource usage request in the decentralized cloud storage system, as well as the subsequent double verification of authenticity and compliance are realized.

On the other hand, it is also the research focus of this section. Based on the node identity authenticity verification method, the data structure design of the storage resource provision information and storage resource usage request on the blockchain is completed. In this paper, the design of the demand side requesting the use of storage resources is directly completed by the system participants in the form of issuing storage resource use requests for specific storage resources. Therefore, for the trusted identification method of storage resource provision information and storage resource usage request, the focus is on the allocation process of storage resource and the definition of the behavior of issuing storage resource usage request. For the definition of "electronic money" transaction behavior in Bitcoin, the purpose is to bind node behavior and node identity, and nodes cannot deny that they have done their own behavior, and other nodes cannot pretend to be "behavior initiators". Refer to the design of the Uspent Transaction Output (UTXO) model in the Bitcoin system.

The basic definition of an "electronic currency" is a chain of linked digital signatures. When the owner of the "electronic currency" wants to transfer the "electronic currency" to others, its operation process is to connect a new digital signature at the end of the

digital signature chain. The specific content of the signature includes obtaining the hash of the last transaction of the "electronic currency" and the public key of the transfer target account. In the process of value transfer, all current accounts can prove their ownership of "e-money" by digitally signing the wallet address of the account in the previous transaction, which can be verified by others to match their public key, as shown in the following figure (Fig. 1):

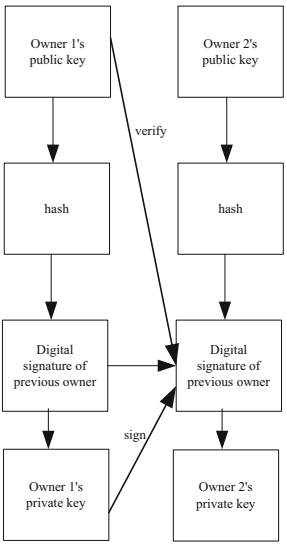

Fig. 1. Data storage structure used in this article

Through the above model analysis, the following will design the data structure of storage resource information and storage resource usage requests. In the solution of decentralized cloud storage resource trust management, storage resource providers hold storage resources and have the qualification of the final storage resource usage authority determination is analogous to the UTXO model. In the data structure design of the storage resource provision information [8], the available attribute information Q of the storage resource and the public key W of the storage resource provider can be used as input, and the output can be provided. The party's open policy for the storage resources held by itself, for example, it can be a set of fixed node public keys, or it can be empty, indicating that the storage resources are only open to a few nodes or completely open. Finally, perform $HASH$ on the above data structure, and then use the provider's private key to digitally sign. Express the formula as:

$$R = T(HASH) * Q * W \tag{8}$$

Through the analysis of the trusted release of storage resources through the above process, the uniquely identified storage resources are spread to other nodes in the P2P network in the form of messages over a certain period of time.

2.3 Compliance Verification of Storage Resource Provision and Usage Requests

Further research and design for its compliance verification method. First of all, the key to verify whether the information provided by the storage resource is legal is to ensure the authenticity of the information provided by the storage resource and the identity of the storage resource provider node. This is also the key link of decentralized trust management in this paper, which is specifically manifested in two points: first, if the provider wants to contribute storage resources in the decentralized system to obtain benefits, the premise is to provide legitimate and trusted storage resources. Before the data is linked, it needs to go through the authenticity verification of node identity and review whether the storage resources it provides are legitimate according to the information provided by the specific storage resources. This is the cornerstone of decentralized cloud storage resource trusted management. Secondly, the data structure that the demand direction sends the use request to meet the required storage resources contains the information provided by the specified storage resources, so it is also necessary to verify the authenticity of the information provided by the storage resources in the compliance verification process of the storage resource use request. The formula is as follows:

$$E=\frac{t}{y}*\sum_{i=1}^{n}h \tag{9}$$

In the above formula, y represents the resource verification node, and h represents the digital signature information.

After the above processing, for the conversion of plaintext information, the *Hash* function, as the core function in cryptography, can also be called a dispersion function. Through this function, information of varying degrees of complexity can be converted into a fixed language sequence. If the function is represented by H, then the process function of information conversion is $h : h = H(M)$. When it is applied to the actual signature, the information will automatically form a summary of the language. This method can not only reduce the time for handwritten signatures, but also prevent others from counterfeiting signatures and using them to commit crimes [9]. The *Hash* function is used as a password. The basic encryption function of learning ensures the security and reliability of information. Generally, the commonly used *Hash* functions are divided into the following two types: single function and polynomial function. Assuming that the single function $f : X \to Y$, X and Y are two sets, and any element $x \in X$, you can easily get $Y = f(x) \in Y$, but if the condition is $y \in Y$, you want to require Obtaining $x \in X$ makes $f(x) = Y$ more difficult, then in this case it is a single function, otherwise it is a polynomial function.

Suppose there is a secret S. in order to prevent theft, it is divided into multiple pieces of information, each of which is called a sub secret and owned by a user. It contains the following properties:

(1) Multiple pieces of information can still be reorganized into secrets;
(2) If a fragment is missing, the security barrier of the whole secret has been destroyed and cannot be reconstructed;
(3) Some information fragments cannot predict the main content of the secret;

(4) A user *Alice* takes any point $Ep(a, b)$ on the curve and the other point P as the standard point;
(5) Sets a private password lock $k(0 < k \leq n)$, n is a positive integer greater than 0, and the password Q can be known from $Q{=}kp$;
(6) Use *Alice* to transport the password Q and the point $Ep(a, b)$ in the curve to another user *Bob* through the standard point;
(7) Acts as the information transmission hub [10], transmits the password and plaintext information to other points M in the curve, and combines with the previous parameters to generate an arbitrary integer set r and $r < n$;
(8) User *Bob* uses the binary algorithm to calculate the point $C1 = M + rQ, c2 = rP$;
(9) Uses its own function to return all $C1$ and $C2$ to user *Alice*;
(10) After *Alice* successfully obtains the data information, after obtaining the result of $C1 - kC_2$, all the information of the point M is obtained.
(11) Perform byte transposition operation according to the following formula, and the calculation expression is:

$$b = c' - a'''(C_N + 1) \tag{10}$$

In formula (10), a''' represents plaintext multiple parameters, c' represents fixed transformation parameters, and C_N represents the column element of the N data.

According to the above calculation results, establish the calculation matrix [11]:

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{bmatrix} = \begin{bmatrix} 02\ 03\ 01\ 01 \\ 01\ 02\ 03\ 01 \\ 01\ 01\ 03\ 02 \\ 03\ 01\ 01\ 02 \end{bmatrix} \tag{11}$$

After the above process, the plaintext content of the data is converted. After the conversion, the plaintext is obfuscated. The calculation process is as follows:

First, assume that a certain data is x;

Second, operate $x * 01$, and the result is x itself;

Third, the binary operation of $x * 02$ and x moves to the left, and the right side is filled with 0. If the highest binary bit of x is 1, the next step is calculated;

Fourth, the obfuscation calculation is completed in this way, and the result of $x * x$ is obtained.

After the above calculation, the accurate plaintext information is obtained.

3 Implementation of Secure Storage of Healthcare Data

3.1 Data Distance Calculation

Through the above process of data clustering and key distribution, on this basis, the next prediction is made through the probability left by each data, and the distance function is determined based on the gradual reduction of its probability to find the dynamic changes of the data. The premise of establishing the function is to ensure the heredity between the data. The distance function can be defined as:

$$f = \sum_{k}^{i=1} \sum_{n}^{j=1} w d_{ij}^2 \tag{12}$$

Among them, w represents the distance factor, d_{ij}^2 represents the squared distance between adjacent data, and $d_{ij} = x_j - y_i$, y_i represent the average distance between matrix parameters, C_i represents the parameter type, then there is a formula:

$$w_{ij} = \begin{cases} 1, \text{The n object belongs to the } C_i \text{ class} \\ 0, \text{The n object does not belong to the } C_i \text{ class} \end{cases} \tag{13}$$

Combined with the irreversibility of clustering data, the formula can be transformed into:

$$f = \sum_{j=1}^{k} \sum_{x \in c\mathrm{i}} d_{ij}^2 \tag{14}$$

Because the above formula is carried out on the basis of constant distance, it can also be called distance difference criterion function. Due to the diversity of data mining, the most common type is the mixed type [12], including numbers and images, so the constraint formula restricting the mixed type is:

When $x_i = x_j$, there is $d_{ij} \geq 0$; when $i, j, k \geq 0$, there is $d_{ij} \leq d_{ik} + d_{kj}$, then the distance function formula for clustering is:

$$d_{ij} = \left(\sum_{k=1}^{m} \left| x_{ik} - x_{jk} \right|^m \right)^{\frac{1}{p}}, p > 0 \tag{15}$$

In the formula, p represents displacement. When the value of p continues to increase, causing the distance between cluster centers to be farther, the formula becomes:

$$d_{ij} = \left(\sum_{k=1}^{n} \left| x_{ik} - x_{jk} \right|^4 \right)^{\frac{1}{2}} \tag{16}$$

When specifying the criteria for the same clustering center, it is necessary to ensure that the same species are clustered with each other to further improve the quality of data mining. If there is an error, iterative calculation can be performed continuously. Then the functional equation that limits the occurrence of the error is:

$$Z_C = \sum_{j=1}^{c} \sum_{k=1}^{n_j} x_k^{(j)} - m d_j^2 \tag{17}$$

where Z represents the error and d_j represents the displacement out of range. The dynamic data has the same characteristics as the samples. When the function is determined, the more complex the feature vector is, the greater the error will be, and the result will be greatly different from the prediction. The density of data is related to the number of samples collected. Only by ensuring that the dynamic data is within a certain range can the error be minimized. Therefore, the final clustering function is:

$$W = \sum_{i=1}^{k} \sum_{x_{j<c}} x_j - c_i^2 \tag{18}$$

Based on the above process, a dynamic distance function is established to analyze the dynamic changes of data and provide a basis for subsequent data storage.

3.2 Spatial Data Division

The data distributed storage mode in the Internet of Things environment should be analyzed according to the specific form of data, and then the spatial data should be divided. The quality of spatial data division directly affects the load of each storage node and the overall performance of spatial data management system. If the stored data to be divided is simple data, divide the data into several parts, and each node in the computer can store certain data. If the object to be divided is complex, the data to be allocated is divided according to the critical matrix. Considering the diversity of spatial data, a jdio curve division method is proposed to ensure the storage balance between each storage node, estimate the number of nodes required according to the amount of data and the calculated workload, and migrate dynamic data between nodes to ensure load balance. Using the distributed storage method of data layer segmentation, the whole spatial area is divided into initial grids. Each grid must contain multiple spatial element objects. Each coordinate in the spatial object is searched through the layer segmentation algorithm. If the amount of spatial object data is large, the data is encoded, and the code of the grid through which the jdio curve passes is taken as the corresponding element object code in the grid, form one or one to many relationships with spatial objects [13]. On this basis, the spatial element objects are sorted according to the initial curve, the data volume of the spatial element objects in the corresponding grid is accumulated from the sub-grid of the initial coding, and each element is divided according to the storage node information of the computer., if the accumulated amount at this time exceeds the corresponding storage node, the grid will be decomposed multiple times until the corresponding data is divided into the specified nodes, and each data in the space is allocated to the space element according to the above process. Subset of objects. The data of spatial division is allocated to storage nodes. At this time, the initial grid is decomposed hierarchically, and the number of objects of spatial elements is greater than the number of storage nodes, so as to improve the division efficiency of spatial element objects. For hierarchical decomposition, the termination order is set. When the division reaches this order, the division is stopped, which can improve the efficiency without affecting the balance of the divided spatial data. The division algorithm is as follows:

$$Q_x = \frac{\omega(W_i)}{F \cdot v(j_0 + 1)} \tag{19}$$

In formula (1), Q_x represents the size of spatial elements, j_0 represents the total number of spatial element objects to be divided, W_i represents the set termination order, and ω represents the curve division order corresponding to the current sub grid. No directional analysis will be done in this calculation.

Through the above formula, the division of spatial data is completed, and the efficiency of distributed storage can be improved through the division of spatial data, which provides a basis for establishing a spatial database in the next step.

3.3 Establish a Spatial Database

Based on the above spatial data division, a distributed spatial database is established. Considering the proximity and topological relationship between spatial element objects, the

spatial data is divided reasonably [14]. The database structure is shown in the following figure (Fig. 2):

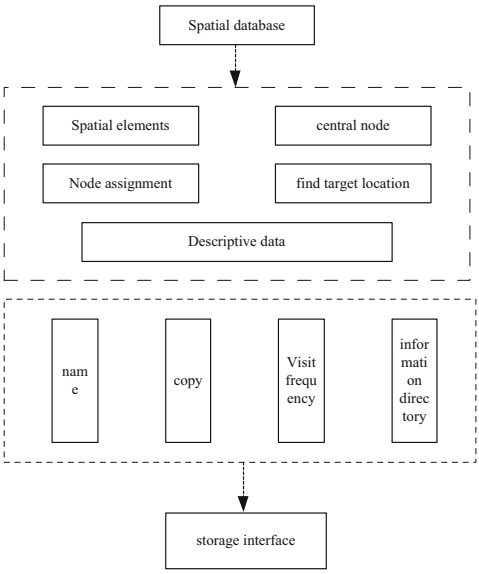

Fig. 2. Process diagram of establishing spatial database

The above figure is the process diagram of spatial database, which is designed according to the above key points when designing the database, so as to meet the characteristics of independence between storage nodes, reduce data transmission between nodes, and improve efficiency. In the distributed database, a node is designed as the central node to maintain the metadata information of the whole network. When a node in the database makes a request, the data node should be processed by the central node. After the task is assigned to other nodes, after the processing is completed, it returns to the target location, takes the database as local data, constructs a global schema, and describes the data from local data, query and process the distributed data, and provide users with a unified data storage interface for the distributed database, so as to realize the transparency of distributed data storage. In addition, the name node server is designed to store and manage the namespace of the file system and the access requests between users, and regularly store the sent data. In order to ensure the reliability of the data, the data should be copied before storage. In order to ensure that the database is damaged or lost when it is stored. On this basis, each node and user storage information directory are managed, and the user access frequency is optimized to optimize the optimal storage of data, so as to complete the design of distributed storage management database. In order to improve the capacity of distributed storage, the database is optimized in the next step.

3.4 Realize Distributed Storage of Cyberspace Data

On the basis of the completion of the database design, in order to increase the database capacity, the database capacity is designed, and the random mechanism is used to distribute the distributed massive data source data packets in cloud computing to all nodes in the distributed system according to a certain reception probability. The storage data packets are formed in the nodes, and the data packets are classified according to the repeatability and access rules of the distributed data, which are divided into hot, cold and repeated storage data packet areas, and partition storage according to the characteristics of different types of data activity factors. The distributed network in cloud computing consists of n nodes, and v nodes are data nodes. Different nodes generate data packets with their own characteristics, and the remaining data nodes are used to store and distribute data. The distributed network is regarded as a random image, the data nodes are described by the fixed points of the graph, and the process of transmitting each data packet to another node according to the transmission mechanism is described by the random walk on the random graph. In order to ensure the random graph They are connected with the maximum probability, and the communication radius of the data node satisfies the following conditions:

$$r \geq \left(sdgt / \hat{\lambda} o\right)^{1/2} \tag{20}$$

In formula (20), g represents the interference factor during data node communication, and o represents the random image, so the coverage of the random walk to the random image is:

$$DO_y = \sum_r t\sqrt{1 - e^i} \tag{21}$$

In formula (21), e^i represents the efficiency of different vertex numbers of random walks, and O_y represents the coverage time of random walks. No orientation analysis is performed in this calculation.

The coverage rate of the random graph is calculated through the above formula. From the coverage rate, we can know the coverage rate of the data nodes in the distributed network. When the distributed processing platform processes the data, according to the above calculation method, different data can be stored, and the data of different nodes can be called through the corresponding functions, so as to increase the storage capacity of spatial data.

Finally, according to the relationship between the information, the files in the system are clustered. The specific process is as follows.

Suppose, $F = \{f_1, f_2, \ldots, f_n\}$ represents a total of n merged high-density information sets, and the set F is divided by the agent hierarchy to obtain l information clusters $N = \{n_1, n_2, \ldots, n_3\}$. Measure the semantic approximation between high-density information data in ships, measure information by the approximation between entities, and finally take the reciprocal of high-density information to obtain the approximation between information and information, namely:

$$Sim\left(q_i, q_j\right) = 1 + \left\{1 + dist\left(q_i, q_j\right)\right\} \tag{22}$$

In formula (22), q_i and q_j respectively represent the high-density information in the hospital, *dist* represents the probability measure of being visited in the system, and *Sim* represents the approximation of the visit form.

Cluster high-density information according to the above process. On this basis, for high-density information storage, store clustered file nodes according to storage space and energy allocation, and set node threshold to prevent a node in the system from bearing too much file storage data. Use the following formula to calculate the grid number, namely:

$$\begin{cases} X = (x_1 - x_2)/L \\ Y = (y_1 - y_2)/L \end{cases} \tag{23}$$

In formula (23), x_1, x_2, y_1, and y_2 respectively represent the geographic coordinates describing the cluster file, X and Y respectively represent the origin of the coordinate system, and L represent the side length of the grid.

On this basis, a multi-threshold prediction mechanism is adopted to ensure the load balance of the high-density information security storage system. The calculation formula is as follows:

$$fit = \frac{sn}{SU(P_i) * bck} \tag{24}$$

In formula (24), *fit* represents the number of storage information nodes in the system, $SU(P_i)$ represents grid expansion information, *bck* represents the measurement index of system information, and *sn* represents storage information.

When the new high-density information is stored, the file is sent to the original saturated grid for storage, so as to realize the safe storage of data.

4 Experimental Analysis

In order to verify the effectiveness of the proposed data storage method, experiments are carried out to verify the proposed method through various aspects. This experiment mainly includes two aspects, one is to analyze the efficiency of the proposed method, that is, the efficiency of data encryption, etc., and the second is to analyze the effect of data security storage, that is, the number of attacks.

The experiment is carried out in the ubuntu10.10 environment of the Vmware Workstation virtual machine, and the cloudsim platform is used to conduct the simulation experiment, and use it to create a data center to provide the basis for the experiment. The traditional data security storage method based on blockchain is used as an experimental comparison method.

4.1 Initialization Time Overhead

When encrypting information, multiple attributes need to be authorized, which will increase the time of data initialization, thus affecting the performance of the entire encryption technology. Therefore, the time cost of information initialization processing

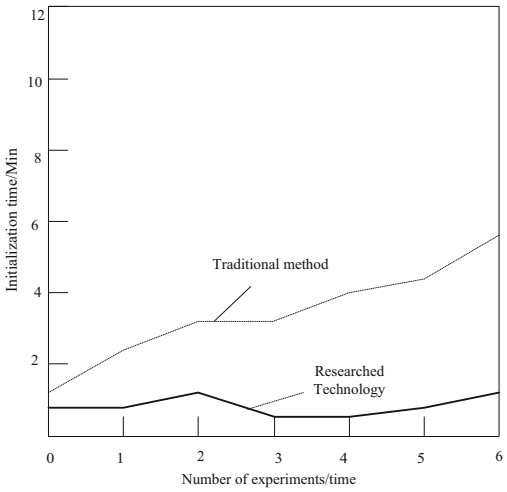

Fig. 3. Initialization time overhead

is taken as the comparison object. The initialization time cost of the proposed method and the traditional method in information encryption is shown in the following figure (Fig. 3):

It can be seen from the above figure that the studied secure storage method takes less time because the studied technology processes the data in advance, reducing the number of attribute authorization institutions, thereby reducing the time cost of initialization processing.

4.2 Key Distribution Time Overhead

The key distribution times for the three methods are as follows (Fig. 4):

Compared with the above figure, it can be found that the proposed data secure storage method spends less time on key distribution. The reason for the less time spent on key distribution of the studied method is that the encryption operation is performed before the file is uploaded and does not involve the generation of attribute keys, so the time for key distribution is less.

4.3 Encryption Time Overhead

The comparison results of the encryption time overhead between the proposed method and the traditional method are shown in the following figure (Fig. 5):

It can be seen from the above figure that the encryption time of the proposed method is less because the proposed method measures the distance between the data, reduces the occurrence of data encryption duplication, and thus reduces the encryption time of the data.

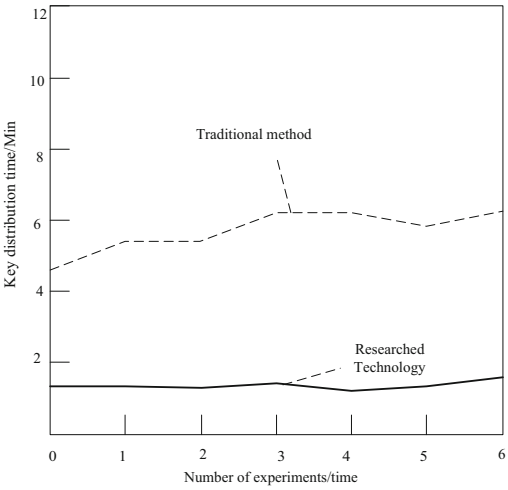

Fig. 4. Key distribution time overhead

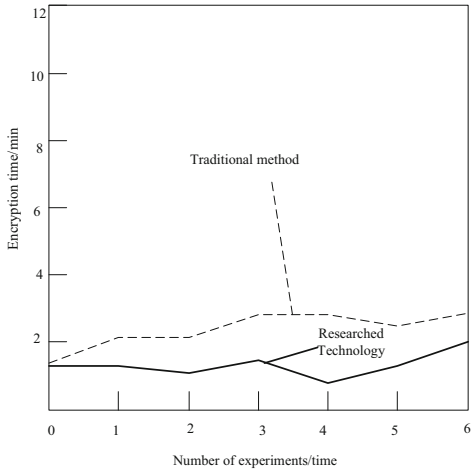

Fig. 5. Encryption time overhead

4.4 Retrieval Time Overhead

The comparison results of the time cost of the proposed method and the traditional method in data retrieval are shown in the figure below (Fig. 6):

Analyzing the above figure, it can be found that the information retrieval time of the two methods is quite different, because the encryption technology studied is not complicated, which reduces the information retrieval time.

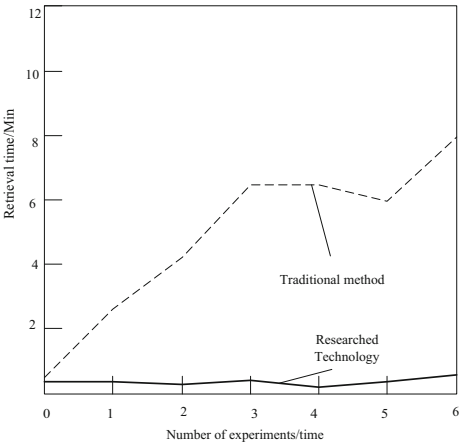

Fig. 6. Retrieval time overhead

4.5 Comparison of Encrypted Information Theft

Comparing the security of the information after the two methods of encryption, the information is stolen as shown in the following figure (Table 1):

Table 1. Stolen situation of encrypted information

Number of experiments/time	The number of times that the researched technical information has been stolen	The number of times the traditional method has been stolen/time
1	0	2
2	1	3
3	0	3
4	0	3
5	0	2
6	0	2
7	0	3
8	1	1
9	0	2
10	0	3

By analyzing the above table, we can find that traditional storage methods have information stolen, and the amount of stolen information is large. After the storage method studied, the information is stolen less, only twice, which can prove that the storage technology studied can improve the security of information storage.

4.6 Comparison of Information Changed

By analyzing the table below, we can find that the information is changed in both methods. The security storage method studied is changed once, and the traditional method is changed many times. The encryption effect is poor, and no research has been done. The storage method is safe (Table 2).

Table 2. Comparison of changes in information

Number of experiments/time	Number of times the technical information studied has been changed / time	The traditional method has been changed times / time
1	0	2
2	1	5
3	0	6
4	0	6
5	0	6
6	0	6
7	0	5
8	0	8
9	0	6
10	0	5

In conclusion, the researched data security storage method can not only reduce the information encryption time, but also improve the information security.

5 Conclusion

The medical data platform designed in this paper can not only realize the distributed storage of medical data, but also ensure the rights control of medical data by patients, and also realize the sharing application of medical data by third-party users. The data allocation strategy is a centralized strategy, which puts forward higher requirements on the computing power, storage capacity and security of the network center node, and further analysis is needed in the follow-up research.

References

1. Huang, J.C., Shu, M.H., Hsu, B.M., et al.: Service architecture of IoT terminal connection based on blockchain identity authentication system. Comput. Commun. **160**(8), 411–422 (2020)
2. Zhang, L.H., Jiang, T.F., Jiang, P.P., et al.: Secure storage scheme for high speed railway monitoring data based on blockchain. Computer Eng. Design **41**(4), 933–938 (2020)

3. Kang, H.Y., Deng, J.: Hybrid encryption method for secure storage of medical data. Trans. Beijing Instit. Technol. **41**(10), 1058–1068 (2021)
4. Deng, Y.X.: Hospital financial data secure storage system based on full homomorphic encryption algorithm. Tech. Autom. Appli. **41**(7), 44–47 (2022)
5. Reppucci, M.L., Acker, S.N., Emily, C., et al.: Improved identification of severely injured pediatric trauma patients using reverse shock index multiplied by Glasgow Coma Scale. J. Trauma Acute Care Surgery **92**(1), 69–73 (2022)
6. Dupuy, B., Romdhane, A., Eliasson, P., Yan, H.: Combined geophysical and rock physics workflow for quantitative co2 monitoring. Int. J. Greenhouse Gas Control **106**(3), 103217–103229 (2021)
7. Li, J., Yan, H., Zhang, Y.: Certificateless public integrity checking of group shared data on cloud storage. IEEE Trans. Serv. Comput. **14**(1), 71–81 (2021)
8. Gokulraj, J., Senthilkumar, J., Suresh, Y., et al.: Data consistency matrix based data processing model for efficient data storage in wireless sensor networks. Comput. Commun. **151**(1), 172–182 (2020)
9. Vuppala, A., Roshan, R.S., Nawaz, S., Ravindra, J.: An efficient optimization and secured triple data encryption standard using enhanced key scheduling algorithm. Procedia Comput. Sci. **1712**, 1054–1063 (2020)
10. Jeong, B.G., Youn, T.Y., Jho, N.S., Sang, U.S.: Blockchain-based data sharing and trading model for the connected car. Sensors **20**(11), 3141–3153 (2020)
11. Hua, D.A., Zheng, Q.A., Qw, B., Zg, B., Yz, C.: Flexible attribute-based proxy re-encryption for efficient data sharing. Inf. Sci. **511**(3), 94–113 (2020)
12. Manikandan, V.M., Bini, A.A.: An improved reversible data hiding through encryption scheme with block prechecking. Proc. Comput. Sci. **171**(1), 951–958 (2020)
13. Geetha, R., Padmavathy, T., Thilagam, T., Lallithasree, A.: Tamilian cryptography: an efficient hybrid symmetric key encryption algorithm. Wireless Pers. Commun. **112**(1), 21–36 (2020)
14. Bao, K.J., Zhang, X.J.: Simulation of remote sharing of database information based on particle swarm optimization. Comput. Simul. **39**(2), 487–49,495 (2022)