Eng Int Syst (2023) 2: 127–139

Engineering
Intelligent Systems

Personal Information Privacy Protection Methods Based on Relational Encryption for the Internet of Things

Li Feng*

School of Law, Dongbei University of Finance and Economics, Dalian 116025, Liaoning, China

With the gradual infiltration of Internet technology into people's lives, people are using the related technologies with increasing frequency. While technologies such as the IoT (IoT) bring convenience to people, they also expose users' information on the Internet, attracting more attention to the issue of the protection of personal information privacy. This article aims to study personal information privacy protection methods in relation to the IoT. For this reason, this paper proposes a privacy protection recommendation algorithm. By designing and improving the algorithm, combined with the understanding of related IoT privacy infringement issues, the improved algorithm can more specifically protect the privacy of personal information. In addition, this article designs related experiments and analyses the results in order to compare algorithms in terms of their performance and the extent to which they protect users' information. The experimental results show that the improved algorithm reduces the encryption and decryption time by 33%.. At the same time, the ability to protect personal information privacy has been explored, and its ability has increased by 41.8%. It has greatly improved the protection level of personal privacy and other related information, and prevented the infringement of users' personal privacy on the IoT.

Keywords: Internet of Things (IoT), Relational Encryption, Privacy Protection, Recommended Algorithm Design

1. INTRODUCTION

The concept of the Internet of Things (IoT) was first proposed by the Massachusetts Institute of Technology in 1999. With the development of the IoT technology and applications, the meaning of the IoT has continued to expand, and the definition of the IoT will also change.Generally speaking, the Internet of Things can be regarded as a network medium for network transmission information detection equipment, which can achieve a comprehensive interconnection of all objects anytime and anywhere according to an established agreement. Its main function is to obtain various information about the physical world through radio frequency identification, sensors, and global positioning systems, and to send and exchange information through a combination of the Internet, mobile communication networks, and other networks. Intelligent computing technology is used to analyze and process the collected information. The IoT improves the perception of the physical world and improves intelligent decision-making and control.

In the information age, big data is a valuable resource for economic development and theoretical research. The mining and analysis of big data has commercial value and can indicate market trends. Taking retail, e-commerce, communications, financial services and other industries as examples, decision makers can use data analysis to determine consumer interests, needs, purchase motivation, and brand emotion and loyalty

*Address for correspondence: Li Feng, School of Law, Dongbei University of Finance and Economics, Dalian 116025, Liaoning, China, Email: fengli@dufe.edu.cn

to make intelligent decisions about services and marketing. However, data is a double-edged sword, as it raises concerns about security and privacy issues. In order to analyze data, it is often necessary to publish the necessary data, which is very likely to give malicious users (attackers) an opportunity to access users' private information. However, the privacy of users' personal information can be safeguarded by means of innovative technologies.

With the rapid development of the Internet era, more and more Internet-related technologies are being continuously developed, and the IoT is one of them, and this has also made more and more people begin to invest in related research on the IoT among. Razzaque MA believes that the IoT (IoT) envisions a future in which digital and physical things or objects (such as smart phones, TVs, cars) can be connected through appropriate information and communication technologies to realize a series of applications and services. The characteristics of the IoT, including the ultra-large-scale IoT, device and network-level heterogeneity, and a large number of events spontaneously generated by these things, will make the development of diversified applications and services a very challenging task [1]. Stojkoska and Trivoldaliev (2017) reviewed the most advanced IoT applications in smart grids and homes. From a review of the literature, they derived a definition of the overall framework of the smart home and its key features. However, the general description of the smart home management model based on the overall framework needs to discuss the current and a future challenge based on IoT solutions [2]. Mishra et al. (2017) used rigorous bibliometrics and network analysis tools to review studies on the IoT conducted a 16-year period16 years, and at the same time provides future directions for the IoT research community and its impact on managers and decision makers. They applied bibliometrics and network analysis techniques, and reviewed the key research topics in articles on the IoT published from 2000 to 2015 [3]. Mostafa et al. (2017) reviewed both scientific papers and commercial results on wearable healthcare devices. They demonstrated that by means of a specifically-designed architecture that includes both hardware and software, data can be collected from wearable devices, sensors, smartphones, medical applications, and medical centers, and subsequently analyzed and stored [4]. Lin et al. (2017) stated that with the advantages of distributed architecture and close proximity to end users, fog/edge computing can provide faster response and higher quality of service for IoT applications. Therefore, the IoT based on fog/edge computing has become the infrastructure for the future development of the IoT. To develop this type of IoT infrastructure, firstly, the architecture, enabling technologies and issues related to the IoT should be closely studied, and then the integration of fog/edge computing and the IoT should be explored [5]. In the context of the large-scale proliferation of the IoT, Singh et al.(2017) focused on the security issues associated with the IoT from the perspective of cloud tenants, end users and cloud providers (whether it is the IoT or the entire IoT subsystem). to the researchers analyzed the current state of the cloud-supported IoT to identify the security considerations that require further work [6]. In order to protect the security of IoT devices, Yang et al. (2017) conducted research on ways to deal with these problems and find better ways to eliminate these risks, or at least minimize their impact on user privacy and security requirements. Their survey consisted of four parts. The first part explored the most relevant limitations of IoT devices and their solutions. The second one introduced the classification of IoT attacks. The next part focused on the mechanism and architecture of authentication and access control. The last part analyzed different levels of security issues [7]. Mosenia and Jha (2017) stated that the IoT, also known as the Internet of Objects, is a transformative means of providing multiple services. Compact, smart devices are an important part of the IoT. They have a wide range of uses, size, energy capacity, and computing power. However, integrating these smart devices into the standard Internet will pose several security challenges because most Internet technologies and communication protocols are not designed to support the IoT [8]. The above-mentioned papers have conducted valuable experiments related to their research topics. They all support their conclusions by means of practical applications, although they are not detailed enough for the relevant IoT technologies mentioned in the article. The use of related technologies is not comprehensive, but remains relatively superficial.

Compared with previous studies, the innovation of this paper is that before protecting the security and privacy of users' personal information captured by the IoT, first, research is conducted to determine the extent of the current IoT infringements of personal privacy. Then targeted research is conducted on related design infringements, and then design algorithms based on the infringing content, so that the designed algorithm in this paper can resist infringements in a more targeted manner, thereby protecting users' personal privacy.

2. PRIVACY PROTECTION METHOD

2.1 The Challenge of the IoT to the Protection of Privacy by Law

In the IoT environment, there are many nodes through which privacy is leaked. At any point during the processes of information collection, transmission and application, personal privacy may be leaked and stolen. Each of the three processes has different equipment for the data processing task, and because they have different features, there can be different types of privacy infringements [9].

(1) The characteristics of private information in the IoT environment

The IoT is a combination of various information technologies, and an IoT network covers various objects in the real world. Objects can be connected and accessed at any time and from any place. Compared with the mobile communication network and the Internet, the security and privacy risks of IoT are particularly serious. The information that is processed in the IoT has the following characteristics [10].

1) High sensitivity

The information and data flowing in the IoT mainly come from the environment, events, objects, etc.

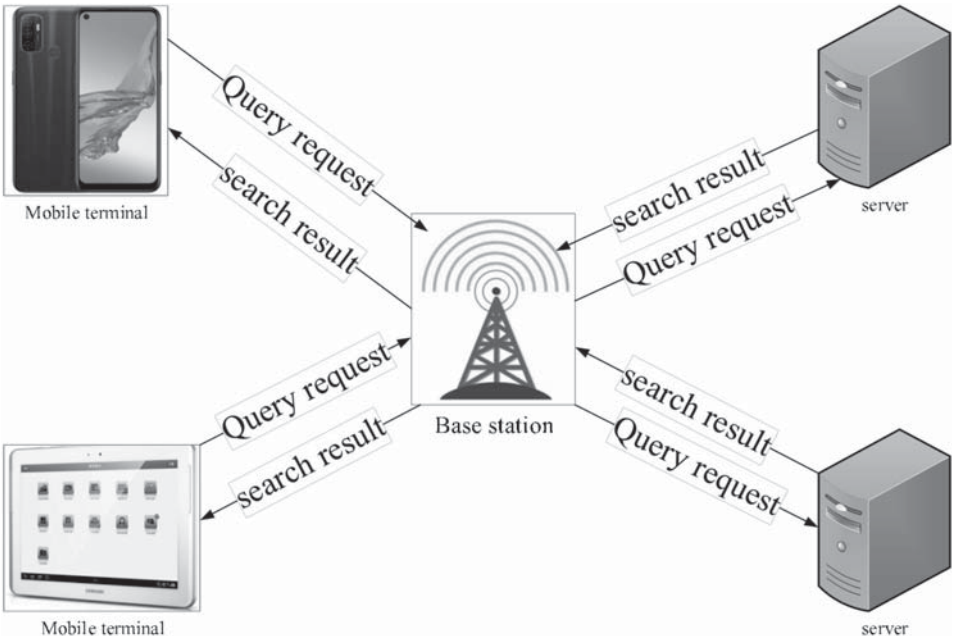

Figure 1 Independent structure.

that are closely related to people's lives, such as personal identity information, real-time location, objects carried, and individual ongoing activities, living habits, and shopping preferences etc. Much of this information is sensitive information about individuals. Of course, people do not want their whereabouts to be tracked at any time, and personal identification information can be obtained maliciously. In addition to the sensitivity of the information itself [11].

2) Authenticity

The information in the Internet of Things comes from the collection object itself, which is a reflection of the real state. Its data is obtained through sensor equipment. Such first-hand data is difficult to forge. Moreover, the structural system of the IoT is very logically related. Effective control is required so that data shared between individuals in the network is accurate and reliable. It can be said that the authenticity of data is the basic requirement of the IoT [12].

3) Systemic

The data in the IoT has a strong logical relationship. The user's data information may be distributed among multiple different databases and application systems. As the scope of user activities expands, the information is constantly changing. At the same time, the various pieces of distributed information will be linked together according to usage needs. In addition, similar information on different individuals is also related to adapt to the intelligent data processing in the IoT to complete the orderly and harmonious integration of environment, society and people [13–14].

(2) Types of privacy infringements of the IoT information-collection layer

Traditional sensors place the reader in a specific location secretly, which may lead to two types of privacy threats. According to the different service architectures, the common architectures designed to protect the privacy of the user's location are quite different [15]. The independent structure is shown in Figure 1 below.

The independent structure is composed of mobile terminals, base stations and location servers, and is a structure from client to server. In the independent structure, the mobile terminal is the main location privacy protection tool. It needs to have functions such as autonomous positioning and storage of information, and install the location service applications required by the user [16]. The distribution structure is shown in Figure 2.

The distributed structure comprises a collection of various mobile terminals of users and LBS service providers. It is characterized by an anonymous collection of peers between mobile devices. This collection is the basis of information anonymization [17]. When a terminal initiates an LBS query, the terminal and nearby mobile terminals are combined to form an anonymous set, and then the location query is initiated by the terminal itself or the agent terminal to the LBS service provider, and the LBS service provider returns the result after querying in the database to the terminal that initiated the request [18]. The third-party-based structure is shown in Figure 3 below.

The third-party-based structure consists of terminal equipment, third-party trusted anonymous servers, and LBS service providers. Its main function is to store and process user parameters, location information, location anonymization, and processing of query results. It is the main tool for the protection of location privacy [19].

With the rapid development of IoT technology, privacy infringement methods have become diversified, concealed and intelligent, which has greatly challenged existing legislation for the protection of privacy. There is a need to respond by:

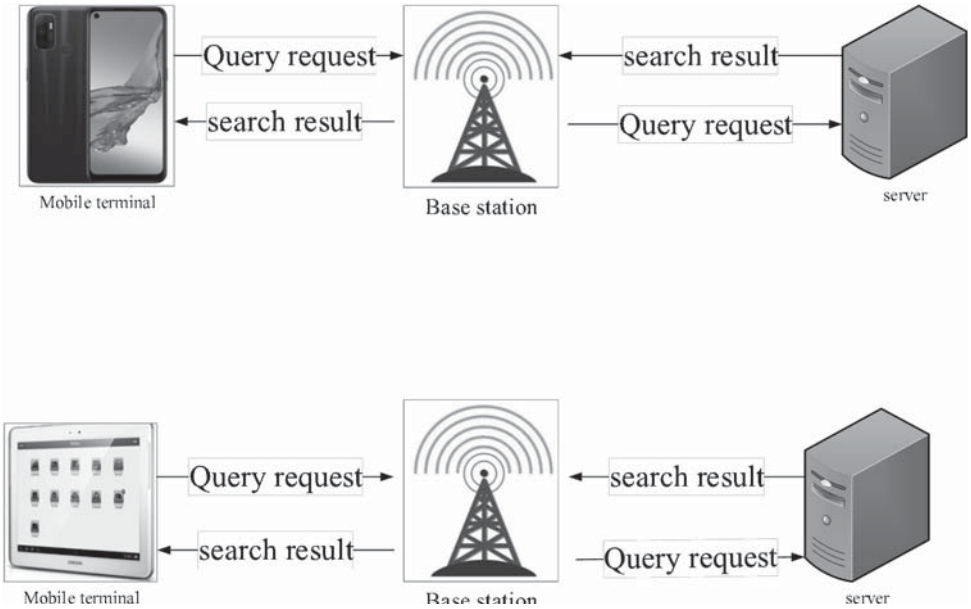

Figure 2 Distributed structure.

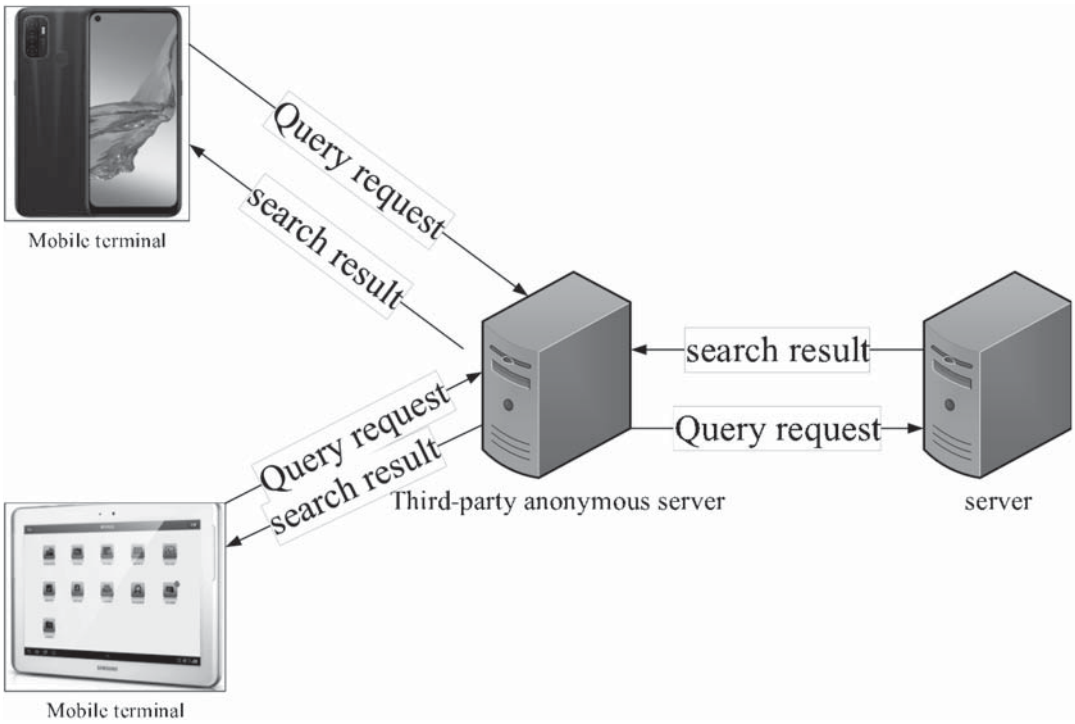

Figure 3 Third-party-based structure.

upgrading technical security, improving laws and regulations, and strengthening government supervision.

2.2 Privacy Protection Recommendation Algorithm Design

Recommendation based on content filtering uses feature extraction from item descriptions to construct item profiles (ItemProfile). At the same time, through the analysis of the items in the user's historical data, the user's degree of interest in the project features is extracted to construct a user profile (UserProfile), and at the same time, algorithm is recommended based on the project profile and the user profile [20]. Generally, the problem solved by the recommender system can be represented by Formula 1:

$$r_{ij} = F_{\Theta}(i, j) \tag{1}$$

The neighborhood-based recommendation system is the most intuitive. If it wants to determine whether a user will like a specific item, it finds the number of individuals similar to the user and see if these users liked the item. Similar to interest groups, group preferences can often indicate the preferences of users in the group. The process is the same when searching

for similar items. Here, it is necessary to determine the degree of similarity between "user-user" or "item-item", and predict user interest based on the similar user preference of a specific user or the user's preference for similar items [21].

In the neighborhood model, taking the item-similarity-based recommendation system as an example, the weighted and formal expression of the pairwise similarity is shown in Equation 2:

$$r_{ij} = F^{NN}(i, j) = \sum_{k \in R_i} r_{ik} \cdot S_{kj} \tag{2}$$

The implicit vector dot product solution corresponding to the user is formalized as:

$$r_{ij} = F^{MF}(i, j) = u_i \cdot v_j \tag{3}$$

In the specific implementation process, it is necessary to construct an objective function based on the observed user rating data to minimize the sum of squared errors between the estimated rating and the original rating.At the same time, in order to avoid the recommendation system model over-fitting to the training data (that is, the existing historical score), the objective function needs to be regularized [22]. This is very important, because the primary goal of the recommendation system is to predict unobserved data, not to ensure that the observed score estimation error is minimal. Referring to Equation 4 for the minimized objective function,

$$loss(U, V) = \min \sum_{(i,j) \in R} \left(r_{ij} - u_i \cdot v_j\right)^2 \tag{4}$$

The stochastic gradient descent algorithm iteratively updates u and V according to the update rules shown in formulas 5 and 6. The parameter t represents the t-th iteration, and the initial value is 0.

$$u_i(t) = u_i(t-1) - \gamma(\nabla_{ui}(U(t-1),\ V(t-1))) + 2\lambda u_i(t-1) \tag{5}$$

$$v_j(t) = v_j(t-1) - \gamma \cdot (\nabla_{vj}(U(t-1), V(t-1))) + 2\mu v_j(t-1) \tag{6}$$

In,

$$\nabla_{ui}(U, V) = -2 \sum_{j:(i,j)} v_j \left(r_{ij} - v_i^T v_j\right) \tag{7}$$

$$\nabla_{vi}(U, V) = -2 \sum_{i:(i,j)} v_i \left(r_{ij} - u_i^T v_j\right) \tag{8}$$

At the same time, assuming that the recommendation system itself is credible and can guarantee the confidentiality of the user matrix u, then the previously solved u can be substituted into Equation 9, thereby converting the problem into a univariate problem, greatly simplifying the process of model refinement [23].

$$loss(U, V) = \min \sum_{x,j \in R} \left(r_{ij} - u_i \cdot v_j\right)^2 \tag{9}$$

The following is a formal expression of collaborative filtering using an automatic encoding machine. The user rating vector $U \in R$ is taken as input (m is the number of items), and the encoder is used to map it to the hidden layer representation. The mapping method is as follows:

$$u = h(W^T u + b) \tag{10}$$

In the output layer, the hidden variable surface needs to be mapped back to the input vector space for the reconstruction of the original input. The mapping method of the output layer is as follows:

$$u = f(W'^T u + b') \tag{11}$$

In:

$$W' \in R^{k \times m} \tag{12}$$

is expressed as the weight vector from the hidden layer to the output layer, and:

$$b' \in R^m \tag{13}$$

is expressed as a bias vector.

The autoencoder is used to complete the sparse user scores to minimize the error between the reconstructed score vector and the original score vector. The objective function is expressed formally as:

$$L = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{u} - u\right) + R(W, W', b, b') \tag{14}$$

where n represents the number of users, and R is a two-paradigm regularization term, which controls the complexity of the model to prevent over-fitting. The regularization term is expanded as:

$$R(W, W', b, b') = \frac{\lambda}{2} \left(\|W\|_2^2 + \|W'\|_2^2 + \|b\|_2^2 + \|b'\|_2^2\right) \tag{15}$$

So far, the problem of predicting user preferences has been transformed to minimize the objective function. The most commonly used method of objective function minimization is stochastic gradient descent [24–25].

$$\frac{\partial h(x)}{\partial x} = h(x)(1 - h(x)) \tag{16}$$

$$\frac{\partial f(x)}{\partial x} = 1 \tag{17}$$

The parameter update algorithm is:

$$\frac{\partial L}{\partial W'} = \frac{\partial L}{\partial u} \frac{\partial u}{\partial W'} + \lambda W' \tag{18}$$

$$\frac{\partial L}{\partial b'} = \frac{\partial L}{\partial u} \frac{\partial u}{\partial b'} + \lambda b' \tag{19}$$

and because there are:

$$\frac{\partial L}{\partial u} = \frac{\partial L}{\partial \hat{u}} \frac{\partial \hat{u}}{\partial u} + \lambda \overline{u} \tag{20}$$

there is an update algorithm for parameters W and b:

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial \overline{u}} \frac{\partial \overline{u}}{\partial W} + \lambda W \tag{21}$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \overline{u}} \frac{\partial \overline{u}}{\partial b} + \lambda b \tag{22}$$

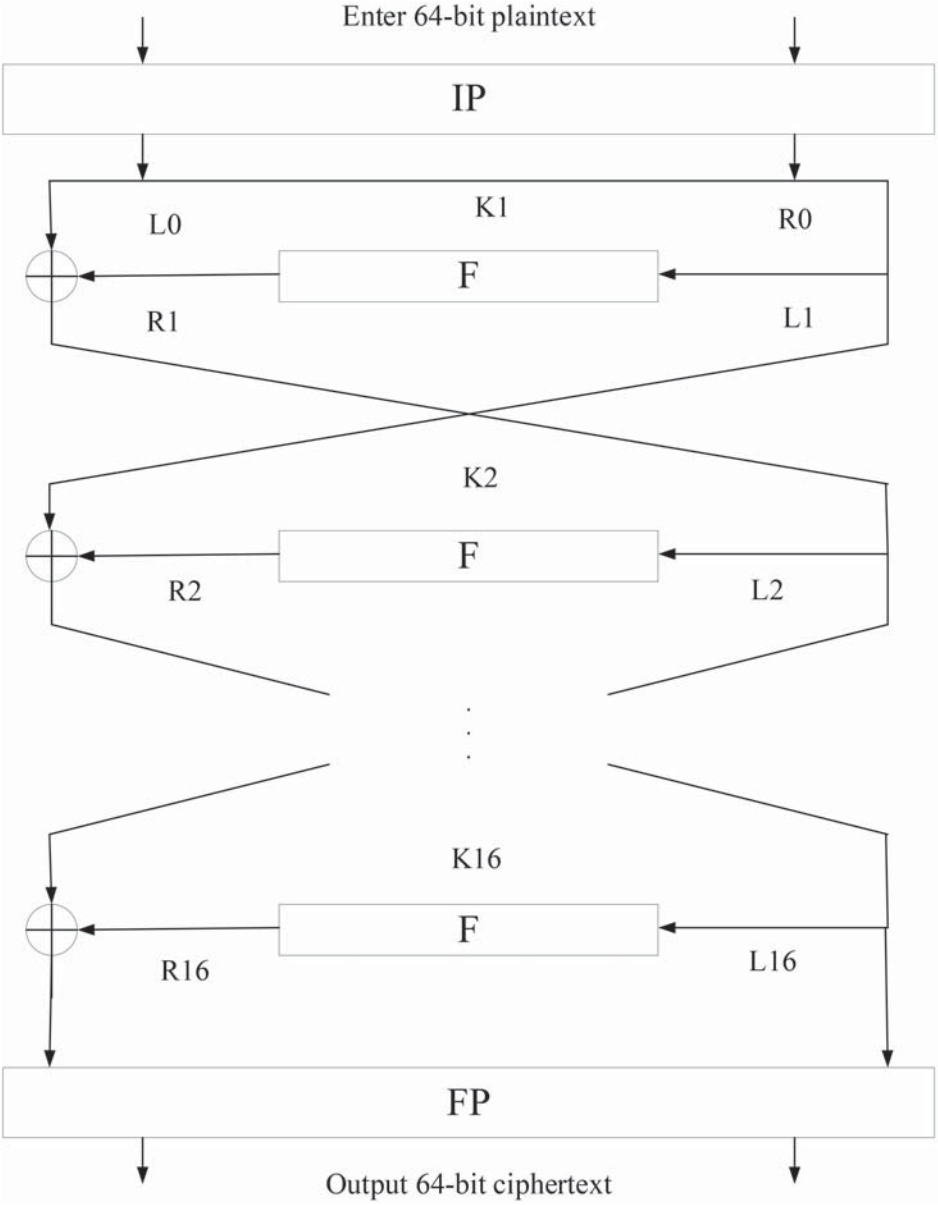

Figure 4 Flow of DES algorithm.

When the stochastic gradient algorithm is used to solve the parameters, the user score vector (a small subset of all user score vectors) is randomly selected to participate in the calculation at each step of the parameter update in order to achieve a more unbiased estimation. It is worth noting that even if the low-dimensional representation of the hidden layer of the sparse input is finally learned, since the automatic encoding machine based on the user rating vector and the item rating vector is not co-trained, it is not suitable to transfer hidden vectors in matrix decomposition. The score is estimated in the form of a product [26–27].

2.3 Encryption System

(1) DES algorithm

The DES algorithm is a widely-used block cipher algorithm. The flow of the DES algorithm is shown in Figure 4. The algorithm has 16 rounds of the same processing of plaintext input data, called "return" processing, and at the same time, there is one replacement in the first place called IP and FP.

Before the main return processing of the DES algorithm, the plain text input data is divided into 64-bit data packets, and each 64-bit data packet is divided into two and a half data packets of 32 bits, which are then cross-processed. The f function processes half of the data packets of 32-bit data at a time, as shown in Figure 5 for the f function process.

Key scheduling is the process of generating subkeys. This process can be described as follows. The selected 56 bits are divided into two 28-bit half-key blocks, and each half-key block is processed separately. In the next process, the two half-key blocks are moved 1 or 2 bits to the left. For comparison, Table 1 shows the performance of typical symmetric encryption algorithms and asymmetric encryption algorithms.

(2) Fine-grained encryption technology

The fine-grained encryption technology comes from various levels of access control in the database. In database

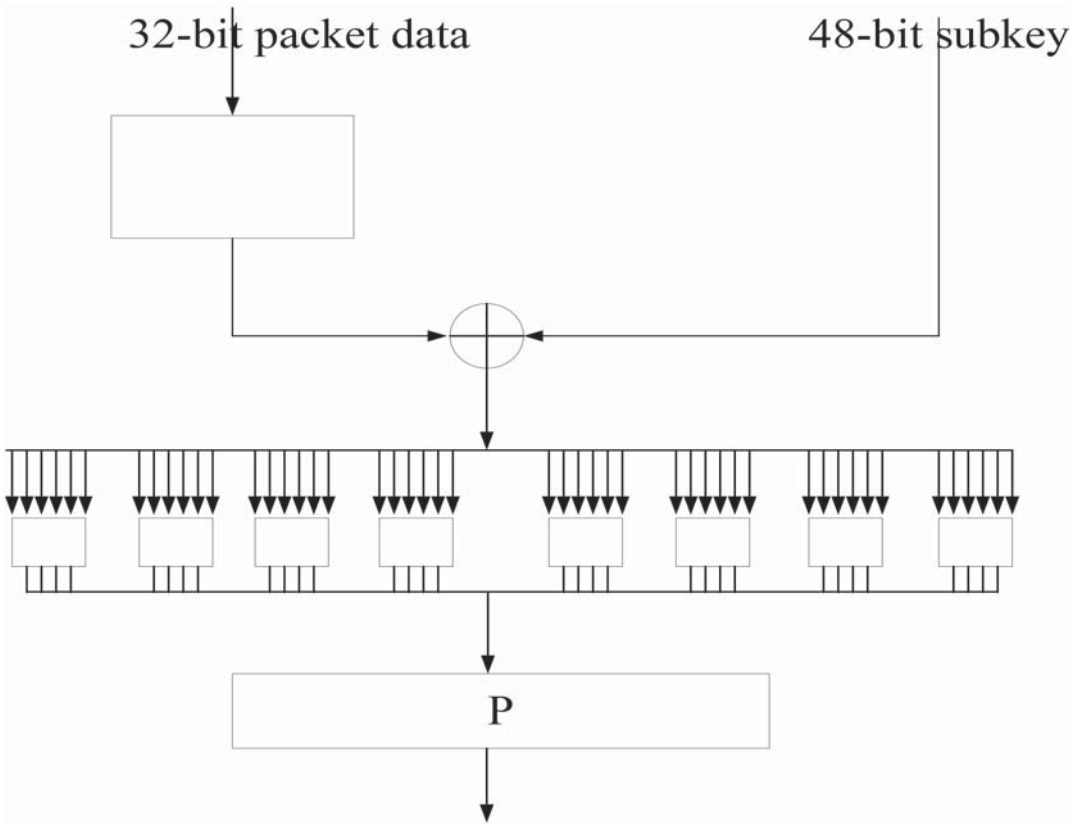

Figure 5 Flow of the f function.

Table 1 Comparison of encryption algorithms.

name	Algorithm type	Key length	speed	LF	safety
DES	symmetry	56 bits	middle	low	middle
IDEA	symmetry	64/128 bits	high	middle	high
RSA	asymmetrical	support variable length key	low	high	high

encryption, basic encryption operations can be performed on the entire database table. A brief schematic diagram of fine-grained encryption is shown in Figure 6 below.

3. THE PRIVACY PROTECTION EXPERIMENT OF RELATIONAL DATA

3.1 Privacy Protection of Relational Data

In relational data privacy protection, privacy refers to sensitive information that data owners, such as individuals and organizations, do not want outsiders to know. In specific applications, privacy includes sensitive data and the characteristics of data representation. Generally speaking, a person's salary, patient's medical information, company's financial data, etc. [1] can be regarded as private information that needs to be protected. Since different users have different perspectives on personal privacy needs, the meaning of privacy in different situations is not completely the same. For example, conservative patients regard personal health information as private, while more open patients do not. Generally speaking, from the perspective of the owners of private information, privacy can be divided into two types: personal privacy and common privacy.

(1) Personal privacy: Personal privacy includes all personal information, affairs, and areas that can be determined to belong to a specific individual or that can be determined to be related to a certain entity who is unwilling to disclose the information. This information includes one's ID number, social security number, telephone number, salary, health status, hobbies, etc.

(2) Common privacy: Common privacy includes not only personal privacy, but also all information that individuals own but do not want to be known to the outside world. Sensitive information such as the location of the company's employees, salary distribution and other sensitive information comes under this category.

In relational data privacy protection, a relational data set S can be regarded as a multi-attribute two-dimensional table composed of n records, and each record contains w attributes. These attributes are usually divided into the following three types according to their characteristics and functions:

1) Quasi-identifiers (QI): this type of attribute can be combined with others to determine the identity of an

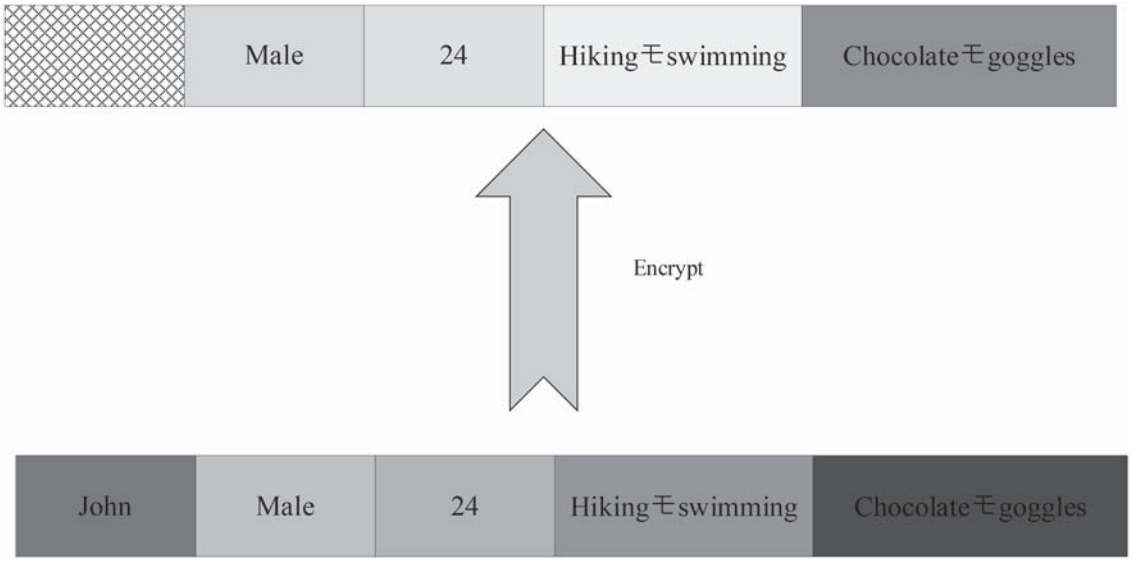

Figure 6 Schematic diagram of fine-grained encryption.

Table 2 Original medical information sheet.

Name	Age	Gender	Country of Citizenship	Disease/Condition
Tom	52	male	UK	diabetes
Allen	59	female	Ireland	stomach cancer
Steven	57	male	Spain	depression
Jim	35	male	America	AIDS
Lin	38	male	China	AIDS
Lucy	29	female	Germany	breast cancer
James	27	male	Germany	influenza

individual, such as gender, age, ethnicity, height, weight, date of birth, postal code, etc.

2) Sensitive attributes (SA): this type of attribute relates to the sensitive information of the individual. It needs to be protected in the published data set to prevent attackers from obtaining the value of this type of attribute, such as personal salary, disease information, etc.

3) Other attributes (Elseattribute, EA): this type of attribute can be made public without endangering individual privacy.

Table 2 is the original medical information table, where "name" is an individual identifier attribute, "age", "sex", and "nationality" together constitute a quasi-identifier attribute, and "disease/condition" is a sensitive attribute that needs to be protected.

(3) Attack method

When relational data is being published, it may be vulnerable to a number of different attacks. The main attack methods are described below.

1) Link attack

Link attacks are the most common method of using published data sets to illegally obtain individual private information. In the process of data release, only starting with the destruction of the relationship between the individual and sensitive information, removing or hiding the identifier attribute of the individual's corresponding record, cannot fundamentally protect the individual's private information. As shown in Table 2, "age", "sex", and "nationality" together constitute a quasi-identifier attribute, and "disease/condition" is a sensitive attribute that needs to be protected. Even if the published data table does not contain individual identification, if an attacker can discover the age, gender, and nationality of an attack target through other means, the attacker can still easily obtain the disease/condition information about the attack target.

2) Homogeneous attack

Homogeneity attack is proposed against the privacy disclosure risk still existing in published data tables that meet the anonymous model, as shown in Table 3, the medical information table that satisfies 2-Anonymity. In the equivalence class formed by t4 and t5, the disease attribute values are all AIDS. Once the attacker can locate the target individual in the equivalence class, he can infer the disease information of the target individual with 100% probability.

3) Similarity attack

Considering that the diversity model can not handle the semantic approximation of sensitive values in equivalence classes, and there is a risk of privacy information leakage, this algorithm proposes a similarity attack. Although the sensitive attribute values of the equivalence classes in the released data set satisfying the diversity are not the same, they may be very similar in semantics, and attackers can still obtain very important privacy-related information based on this. As shown in Table 4, the 4-anonymous medical information table,

Table 3 Satisfaction 2-Anonymous Medical Information Form.

Serial number	Age	Gender	Country of Citizenship	Disease/Condition
T1	[52–57]	*	*	diabetes
T2	[52–57]	*	*	stomach cancer
T3	[52–57]	*	*	depression
T4	[35–38]	male	*	AIDS
T5	[35–38]	male	*	AIDS
T6	[27–29]	*	Germany	breast cancer
T7	[27–29]	*	Germany	influenza
*Means not visible				

Table 4 Meets 4-Anonymous Medical Information Sheet.

Serial number/equivalent class number	Age	Gender	Country of Citizenship	Disease/condition
T1/1	[36–41]	*	America	diabetes
T2/1	[36–41]	*	America	stomach cancer
T3/1	[36–41]	*	America	depression
T4/1	[36–41]	*	America	AIDS
*Means not visible				
…				

Table 5 Table showing disease information released at time T1.

Name	Serial number/equivalent class number	Age	Country of Citizenship	Disease/Condition
Lucy	T1/1	[31–35]	UK	diabetes
Allen	T2/1	[31–35]	Ireland	stomach cancer
Tom	T3/1	[52–58]	*	depression
Jim	T4/1	[52–58]	*	AIDS
James	T5/1	[62–70]	China	AIDS
Wilson	T6/1	[62–70]	Germany	breast cancer
*Means not visible				

even if the attacker is not sure of the target patient's specific disease information , he can still infer that the patient's disease is related to gastric disease.

4) Value equivalence attack

Value-equivalence attacks are mainly aimed at anonymized data sets that are dynamically published. Although each independently published data set satisfies the basic anonymity requirements, the attacker can combine the continuously published anonymous data sets to obtain the recorded sequence belonging to the same attribute set, and then infer the sensitive attribute value of one or several individuals. For example, Table 5 is the disease information table released at T1 time. Obviously, the released sequence satisfies the 2-invariance principle. However, the attacker may still find that "Allen" and "James" have the same disease through other means. If at T3, "Jone" is recovered from the illness and then deleted from the data table. "Tom" was added to the equivalence class 1 because of the "flu" disease. The attacker can still infer that "Lin", "Allen" and "James" have the same disease information. As long as the attacker can infer the identity of any person in the attribute set, the sequence anonymity of equivalence class 1 will be completely invalid.

4. ALGORITHM ENCRYPTION ANALYSIS

4.1 Comparative Analysis of LPPK Method and Existing Methods

The following experiment involves the kNN query and range query method proposed in this article and three other current methods. respec Among them, the LPPK method proposed in this paper is termed Kernel (kernel method), and the parameters of the three methods for comparison are set according to the default values in the related articles: In JL, $m = 10$, $\varepsilon = 0.5$ and $r = 2$km in GEO and DP, and $\omega = 6$ in Kernel. The comparison is shown in Figure 7 below.

It can be seen from the figure above that the lines representing Kernel in each similarity graph are above other methods, indicating that the kernel method has obtained higher similarity values for the four interest point queries of the two data sets. When k is small (for example, when $k < 30$), the similarity between GEO and DP methods is small, indicating that the query effect of these two methods is poor. When $k > 30$, the similarity between GEO and DP is slightly stable, but far less than the similarity between JL

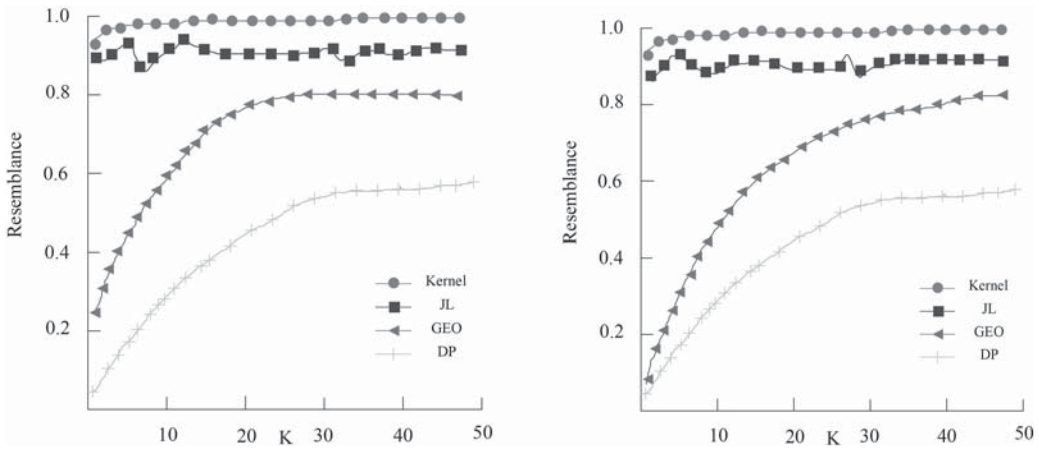

Figure 7 Comparison of kNN query results.

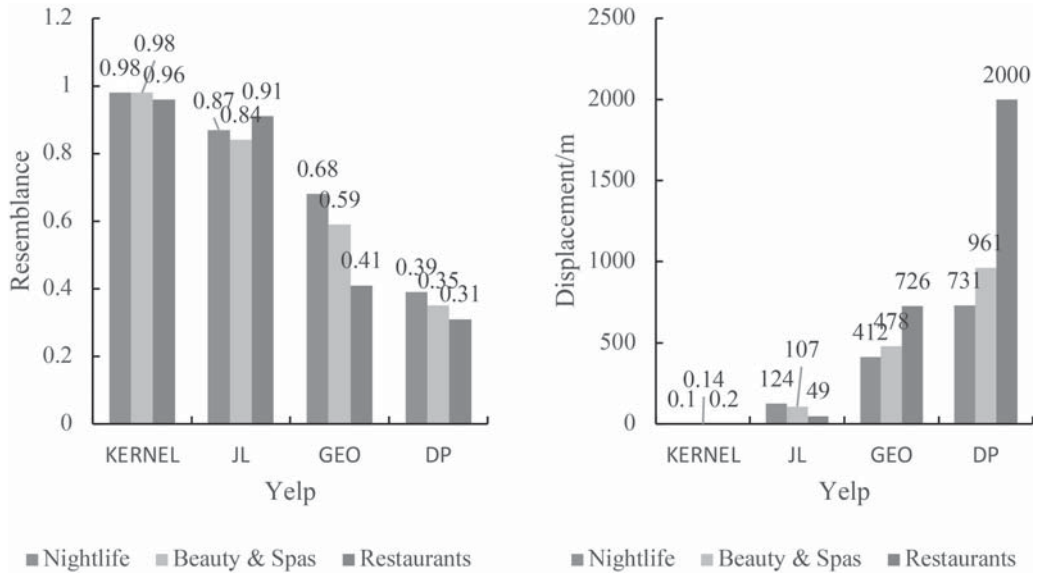

Figure 8 LPPK's kNN query performance under different interest point densities.

and Kernel methods. It can be seen from the figure that the similarity of JL is almost stable at about 0.85, which performs well, but the result of Kernel is close to 1. Therefore, it is still much higher than JL, indicating that Kernel's performance is significantly better than JL. In addition, from the trend of the curve in the figure, as the value of k increases, the similarity between GEO and DP first rises and then gradually stabilizes. This shows that the query effect of these two methods has a strong correlation with the value of k, while the Kernel and JL methods show close to perfect robustness in all configurations, indicating that the two methods are not sensitive to changes in the value of k. In terms of distance difference, the distance difference between Kernel and JL is always significantly lower than the other two methods. When $k < 30$, the distance difference between GEO and DP decreases rapidly with the increase of k, but tends to be stable when $k > 30$. On the contrary, JL and Kernel have always maintained a relatively low distance difference. Taking the Yelp data set as an example, the value of the distance difference between GEO and DP results is much higher than that of JL whose distance difference is about 100 meters, and the result of Kernel whose distance difference is less than 5 meters is close to perfect.

At the same time, the performance of the four methods is compared under different interest point densities. This article selects 3,029 points of interest from the Yelp data set covering an area of 340 square meters, including 404 night scenes, 848 beauty institutions and 1,777 restaurants. Therefore, the density of these three points of interest: restaurants>beauty institutions>night scenes. 2177 points of interest were extracted from the SimpleGEO data set covering an area of 12,368.193 square meters, including 79 restaurants, 666 stores, and 1434 medical institutions. The density of these three points of interest: Medical institutions>shops>restaurants. The statistics are presented in Figure 8 below.

From the above comparison, it can be shown that the increase in the density of interest points reduces the accuracy of these two methods. The results of these two data sets both show that GEO and DP are sensitive to changes in the density of interest points, while Kernel and JL are not sensitive to them. The query results are better under different density of interest points. And in contrast, Kernel's query results are more accurate.

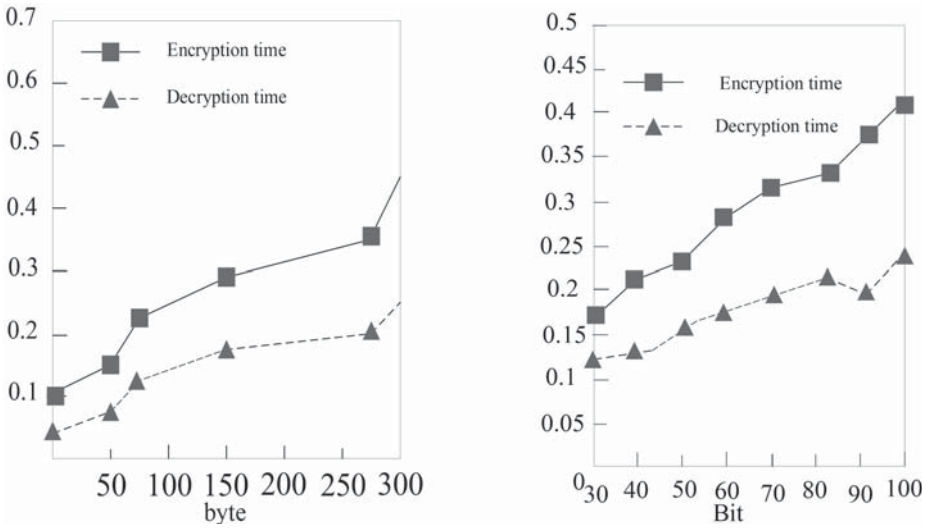

Figure 9 Relationship between encryption and decryption time and finite field.

4.2 Calculation of Consumption Analysis

The secure aggregation method includes three processes: encryption, additive aggregation and decryption. The encryption process of the algorithm proposed in this paper includes two scalar multiplication operations and one scalar addition operation. Compared with scalar addition, the calculation cost of scalar multiplication is greater. For the scalar multiplication rG and rK (KkG) in the ECC encryption process, let the length of the binary expansion of r be m and the length of the expansion of rk as l, using the simplest binary expansion method to calculate rG and rK will require m/1 times of point doubling and m/2 times of point addition, and 1 times of point doubling and 1/2 point addition. The additive aggregation process is the operation of scalar addition, and its computational cost is related to the number of nodes participating in the aggregation and the length of the finite field.

The calculation time of elliptic curve encryption, decryption and additive aggregation obtained by changing the finite field size and the length of the private key is shown in Figure 9 below.

It can be seen from the above data that the ECC homomorphic encryption and decryption time is related to the size of the plaintext, the length of the key, and the length of the finite field. For the purpose of ensuring security, reducing the length of the finite field and the length of the key can effectively reduce the encryption and decryption time. ECC homomorphic encryption requires two scalar multiplications, so the encryption time is significantly longer than the decryption time. Because the addition aggregation operation does not need to go through a scalar multiplication operation, it requires much less time than the encryption and decryption, and the difference is significant. The addition aggregation time increases with the increase of the finite field length.

4.3 Algorithm Encryption Space Analysis

The original DES algorithm, the algorithm without S-box and the algorithm with S-box are compared and analyzed. First, two groups of plaintext with only 1bit difference are input. After the encryption calculation of the algorithm, the two groups of ciphertexts obtained are obviously different, and they do not show the data similarity in the original text. The S-box-free algorithm performs 6 rounds of encryption on the same two sets of plaintext, and the two sets of ciphertexts obtained have a certain degree of similarity. However, the two sets of ciphertexts obtained by the same encryption of the two sets of plaintexts with the S-box algorithm are obviously different. This comparative analysis demonstrates that the S-box-free algorithm does not have a better "avalanche effect", and the improved algorithm after adding the S-box has a security close to the original DES. The specific comparison is shown in Figure 10:

In regard to encryption time, due to the frequent calculations of DES function rounds and shift operations, the processing time is relatively long. However, the execution time of S-box and S-box algorithms with similar algorithm structures are close to each other and shorter, which is about 1/3 of the time consumed by the DES algorithm. This ensures that the improved encryption algorithm has better real-time performance.

From the perspective of storage space occupied by encryption, DES occupies the most memory and external storage space, and the non-S-box algorithm occupies the least. The S-box algorithm is a compromise between the two algorithms. The storage space occupied by the S-box algorithm is about 2/3 of the storage space occupied by the DES algorithm. Therefore, it has advantages over the DES algorithm and is suitable for sensing nodes with less storage space.

After testing the memory of related algorithms, it is concluded that the proposed improved algorithm has increased the encryption and decryption time by 33%. The ability to protect personal information privacy in the context of the IoT has increased by 41.8%, which can provide effective protection of users' personal privacy on the basis of relationship-based encryption during actual application.

5. CONCLUSIONS

The main focus of this article is the protection of users' personal and private information in the context of the IoT.

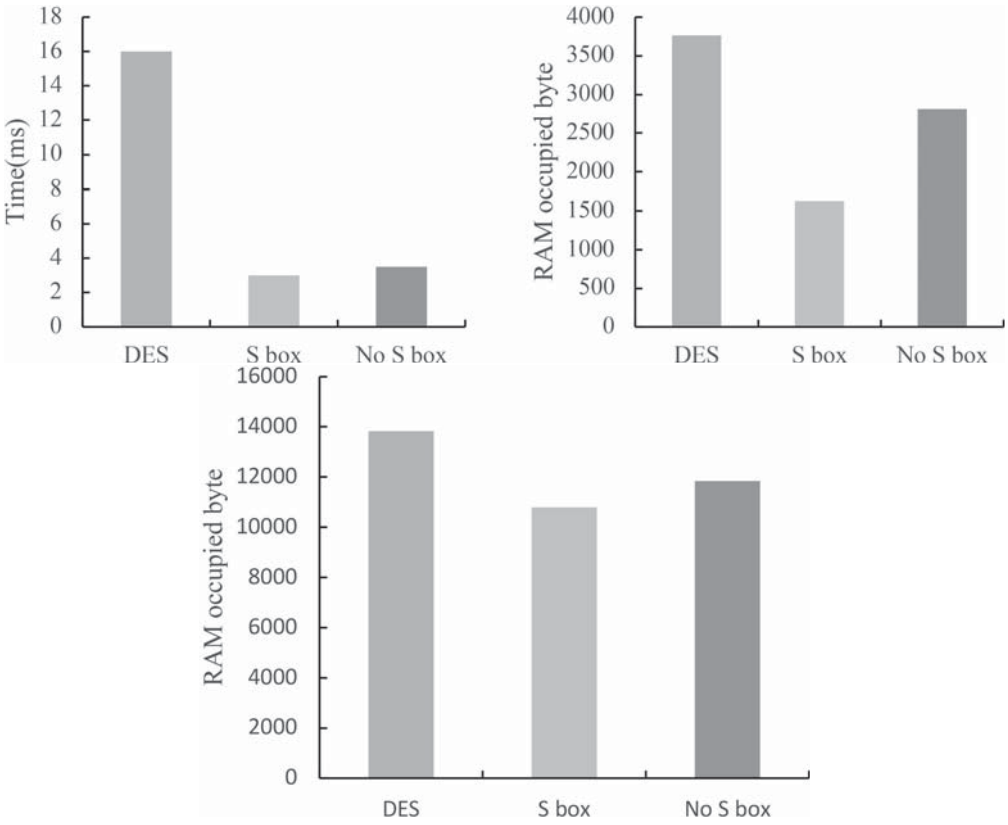

Figure 10 Comparison of encryption space used by three algorithms.

After a general understanding of personal user privacy violations in the context of the IoT, the related privacy protection recommendation algorithm is designed and, combined with the LPPK method, the user's personal location is also encrypted and protected, which effectively eliminates the problem of user information infringement in the context of the IoT. At the same time, the performance analysis of the algorithm and the analysis and reduction of the energy consumption are designed to improve the sustainable operation ability of the algorithm.

ACKNOWLEDGEMENTS

This work was supported by the Scientific Research Fund Project of the Education Department of Liaoning Province: Thinking on Public Law of Personal Data Protection in Government Data Opening (LN2020J08).

REFERENCES

1. Razzaque M A, Milojevic-Jevric M, Palade A, et al. (2017). Middleware for IoT: A Survey. IEEE IoT Journal, 3(1):70–95.
2. Stojkoska B, Trivodaliev (2017). K V. A review of IoT for smart home: Challenges and solutions. Journal of Cleaner Production, 140(pt.3):1454–1464.
3. Mishra D, Gunasekaran A, Childe S J, et al. (2017) Vision, applications and future challenges of IoT: A bibliometric study of the recent literature. Industrial Management & Data Systems, 116(7):1331–1355.
4. Mostafa H, Kerstin T, Regina S. (2017). Wearable Devices in Medical IoT: Scientific Research and Commercially Available Devices. Healthcare Informatics Research, 23(1): 4–15.
5. Lin J, Yu W, Zhang N, et al. (2017). A Survey on IoT: Architecture, Enabling Technologies, Security and Privacy, and Applications. IEEE IoT Journal, 4(5):1125–1142.
6. Singh J, Pasquier T, Bacon J, et al. (2017). Twenty Security Considerations for Cloud-Supported IoT. IEEE IoT Journal, 3(3):269–284.
7. Yang Y, Wu L, Yin G, et al. (2017). A Survey on Security and Privacy Issues in Internet-of-Things. IoT Journal, IEEE, 4(5):1250–1258.
8. Mosenia A, Jha N K. A (2017). Comprehensive Study of Security of Internet-of-Things. IEEE Transactions on Emerging Topics in Computing, 5(4):586–602.
9. Alrawais A, Alhothaily A, Hu C, et al. (2017). Fog Computing for the IoT: Security and Privacy Issues. IEEE Internet Computing, 21(2):34–42.
10. Zarpelao B B, Miani R S, Kawakani C T, et al. (2017). A survey of intrusion detection in IoT. Journal of Network & Computer Applications, 84(Apr.):25–37.
11. Zhang Y, Wen J. (2017). The IoT electric business model: Using blockchain technology for the IoT. Peer-to-Peer Networking and Applications, 10(4):983–994.
12. Kshetri, Nir. (2017). Can Blockchain Strengthen the IoT? IT Professional, 19(4):68–72.
13. Hahm O, Baccelli E, Petersen H, et al. (2017). Operating Systems for Low-End Devices in the IoT: a Survey. IEEE IoT Journal, 3(5):720–734.
14. Zhang D, Yang L T, Min C, et al. (2017). Real-Time Locating Systems Using Active RFID for IoT. IEEE Systems Journal, 10(3):1226–1235.

15. Ni J, Zhang K, Lin X, et al. (2018). Securing Fog Computing for IoT Applications: Challenges and Solutions. IEEE Communications Surveys & Tutorials, 20(99):601–628.
16. Collier, Steven E. (2017). The Emerging Enernet: Convergence of the Smart Grid with the IoT. IEEE Industry Applications Magazine, 23(2):12–16.
17. Atzori L, Iera A, Morabito G. (2017). Understanding the IoT: definition, potentials, and societal role of a fast-evolving paradigm. Ad Hoc Networks, 56(Mar.):122–140.
18. Ejaz W, Naeem M, Shahid A, et al. (2017). Efficient Energy Management for the IoT in Smart Cities. IEEE Communications Magazine, 55(1):84–91.
19. He L, Ota K, Dong M. (2018). Learning IoT in Edge: Deep Learning for the IoT with Edge Computing. IEEE Network, 32(1):96–101.
20. Khan A A, Rehmani M H, Rachedi A. (2017). Cognitive-radio-based IoT: applications, architectures, Spectrum related functionalities, and future research directions. IEEE Wireless Communications, 24(3):17–25.
21. Pikul J H, Ning H. (2018). Powering the IoT. Joule , 2(6):1036–1038.
22. Yang W, Mao W, Zhang J, et al. (2017). Narrowband Wireless Access for Low-Power Massive IoT: A Bandwidth Perspective. IEEE Wireless Communications, 24(3):138–145.
23. Mumtaz S, Alsohaily A, Pang Z, et al. (2017). Massive IoT for Industrial Applications: Addressing Wireless IIoT Connectivity Challenges and Ecosystem Fragmentation. IEEE Industrial Electronics Magazine, 11(1):28–33.
24. Ammar M, Russello G, Crispo B. (2018). IoT: A survey on the security of IoT frameworks. Journal of Information Security and Applications, 38(FEB.):8–27.
25. Lyu X, Wei N, Hui T, et al. (2017). Optimal Schedule of Mobile Edge Computing for IoT Using Partial Information. IEEE Journal on Selected Areas in Communications, 35(11):2606–2615.
26. R. Zhang, G. Wang. (2022). A Study of Online Educational Resource Recommendation for College Chinese Courses Based on Personalized Learning. International Journal of Engineering Intelligent Systems ; 30(5):335–340.
27. F. Wang, L. Shen, W. Li. (2022). A Target Detection Algorithm of Aerial Images in Power Grid Inspection Based on Transfer Learning. International Journal of Engineering Intelligent Systems ; 30(5):341–351.

Li Feng was born in Yakeshi, Inner Mongolia, P.R. China, in 1979. She received her PhD from the China University of Political Science and Law, P.R. China. Currently, she is working in the School of Law, Dongbei University of Finance and Economics. Her research interests include administrative contracts, and personal information protection.
E-mail: fadawaly@163.com