A Secure k-nearest Neighbor Query Processing Method Based on Extended Range of Cipher-Text Search

Yong-Bing Zhang[1,2], Qiu-Yu Zhang[1], Yi-Long Jiang[2], and Jun Yang[2]
(Corresponding author: Qiu-Yu Zhang)

School of Computer and Communication, Lanzhou University of Technology[1]
No. 287, Lan-Gong-Ping Road, Lanzhou 730050, China
(Email: zhangqylz@163.com)
Gansu Institute of Mechanical & Electrical Engineering[2]
No. 107, Chi-Yu Road, Tianshui, Gansu 741001, China

(Received Jan. 12, 2022; Revised and Accepted June 21, 2022; First Online July 3, 2022)

Abstract

To improve the effect of privacy protection and the query service quality in a cloud environment, a secure k-nearest neighbor (kNN) query processing method based on an extended range of cipher-text search is proposed. Firstly, the area is divided into several square grids, and the Morton code of each location is calculated. Then, the location data is encrypted using public-key encryption and scrambling encryption, which is outsourced to the cloud server. When querying kNN interest points, a query request is sent to the data owner by the user. After obtaining the query trapdoor and private key, the query trapdoor is sent to the cloud server for the query. On the server-side, the grid region of the current location is found according to the approximate matching between the Morton codes. The current grid is used as the center to expand the search area until kNN interest points are found and sent to the querying user. Finally, the cipher-text query result is decrypted into plaintext with the private key on the user side. Experimental results show that the method can ensure location privacy and improve query efficiency. At the same time, the balance between privacy protection security and query service quality is achieved.

Keywords: Extended Range of Cipher-Text Search; K-nearest Neighbor (KNN); Location Privacy Protection; Location-based Service (LBS); Morton Code

1 Introduction

In recent years, location-based service (LBS) has become an indispensable part of people's daily life. The query of kNN interest points is one of the most important applications of LBS service [27], such as finding nearby hotels, hospitals, markets, restaurants and other geographical locations. With the development of cloud computing and communication technology, location big data is outsourced to cloud server by data owner [5], peoples can enjoy the convenience of location services through mobile devices anytime and anywhere. But the cloud server is not secure, which increases the risk of users' privacy leakage [6]. In order to protect the location privacy from being leaked, encrypting the location data before outsourcing and querying the location data in the cipher-text state are better methods to protect the location privacy [1,9].

However, in the kNN location cipher-text query, when the query range is too large, it will cause too many invalid queries and reduce the query efficiency. If the query range is too small, the number of interest points found in the selected area cannot meet the user's query requirements. Therefore, the extended range query method [13,23] can be used for cipher-text search, the spatial area is divided into several square grids, takes the grid where the query user is located as the current grid, and continuously expand the query area to the adjacent grid according to the query requirements, until kNN interest points are found.

In order to improve efficiency and service quality of cipher-text query, a secure kNN query processing method based on extended range of cipher-text search is proposed. The data record set composed of Morton code, location coordinate and placename information are generated through spatial location conversion by data owner, which is encrypted by a combination of public key encryption and scrambling encryption, and outsourced to the cloud server. When querying location data, a query request is first sent to the location data owner by query user. The data owner generates a query trapdoor according to the current location and query requirement of the query user, and sends it to the query user together with the private key. The query user sends the query trapdoor to the cloud server for query in cipher-text state. On the server side, the current grid region is found according to the Morton code, the kNN interest points are retrieved by neighborhood grid expansion search method, and the cipher-text query results are returned to the query user. At last, the query results are decrypted into plaintext with the private key on the user side. The contributions of this work are as follows:

1) The spatial location coordinates are transformed in this paper, and the spatial two-dimensional location coordinates are transformed into Morton code, which not only ensures the privacy of location coordinates, but also facilitates the encryption and approximate calculation of location data.

2) The spatial area is divided into several grids, so that all locations are divided into different square regions. The kNN interest points are found through neighborhood grid expansion search, which reduces invalid queries and further improves the query efficiency.

3) A safer encryption method is provided by combining public key encryption and scrambling encryption. The public key encryption mechanism is adopted for the location coordinates and place name information to facilitate the encryption and decryption of location data. The Morton code is encrypted by scrambling encryption scheme, and the neighborhood grid expansion can be calculated in the cipher-text state.

4) The neighborhood grid expansion search method is used to search the cipher-text location data, and the kNN query results are approximately sorted according to the neighborhood relationship.

The remaining part of this paper is organized as follows. Section 2 reviews related work of location privacy protection. Section 3 gives system model of this study. Section 4 describes algorithms and structure analysis. Section 5 gives the experimental results and performance analysis as compared with other related methods. Finally, we summarize our work in Section 6.

2 Related Work

In order to prevent the leakage of privacy information, and ensure the secure query processing of kNN interest points, a variety of location privacy protection methods is proposed by experts and scholars, including k-anonymity method [26], dummy location method [20] and encryption method [11]. In these location privacy protection methods, k-anonymity [17] method was born in the relational database, and its key attribute is dealt with using generalization and fuzzy technology. None of the records can be distinguished from other k-1 records, and the location anonymity is realized. Xu et al. [21] proved that the size of k-anonymous region directly affects the accuracy of location query results, which provides guidance for the subsequent research on anonymous region construction methods. On the basis of this research, many anonymous region construction methods [16, 19, 22] are proposed. Zhang *et al.* [28] proposed a scheme to enhance user privacy through cache and spatial k-anonymity. By applying multi-level cache and spatial k-anonymity, the effect of privacy protection is improved. However, these methods have two serious shortcomings: Firstly, it must rely on TTP, but TTP is not absolutely secure, and it's easy to become the bottleneck of the system. Secondly, the size of anonymous region and the accuracy of query results are a pair of contradiction, and the larger the anonymous region, the better the effect of privacy protection, but the accuracy of the query results will be reduced.

The dummy location method has been widely studied because it does not need TTP, and does not need to construct anonymous region. Kido *et al.* [7] first introduced the method of dummy location into location privacy protection in 2005. Dummy locations are generated in Mobile client, and then be sent to LBS server with the real location for query. Niu *et al.* [14] proposed a mobile location privacy protection scheme named DUMMY-T, which aims to protect user's location privacy from background attacks, the dummy is generated by the dummy location generation (DLG) algorithm, and dummy path is generated by the dummy path construction (DPC) algorithm, which ensures the security of location privacy. Song *et al.* [15] proposed a privacy protection method of moving trajectory based on dummy location under continuous query conditions. Considering the semantic similarity and physical dispersion between dummy locations, Zhang *et al.* [29] selected locations with small semantic similarity and good physical dispersion to generate dummy locations, which improves the privacy protection effect of the dummy location method. Taking full account of the semantic similarity and query probability of location data, Wang *et al.* [18] proposed a maximum minimum dummy location selection based on location semantics and query probability, which avoids the attacker from filtering dummy locations combined with background knowledge. In summary, in the generation of dummy location, how to select the dummy location with good indistinguishability is the main problem.

In recent years, encryption method [30] has become the preferred method for location privacy protection because of its good privacy protection effect. When massive encrypted data is outsourced to the cloud server, the most important thing is to implement efficient encryption scheme, and ensure the data security and query security. Zeng *et al.* [25] proposed a location query method for cipher-text circle detection, which uses the method of encrypting vectors to judge whether a location point is in a circular area. Li *et al.* [10] proposed a cipher-text location query method in square area. Du *et al.* [3] proposed a location privacy protection method based on attribute privacy information retrieval. These methods improve the effect of privacy protection, but do not solve the problem of sorting query results. We focus on kNN query which protects privacy in cloud computing environment, and several typical kNN query processing schemes for use with encrypted databases are described below.

Elmehdwi *et al.* [4] proposed a kNN query processing scheme for encrypted database based on Paillier cryptosystem, and proposed two SkNN protocols for encrypting data in cloud environment. Yang *et al.* [24] present a verifiable privacy preserving kNN query scheme, using network Voronoi diagram and some cryptographic algorithm primitives, including pseudo-random function, Paillier cryptosystem, compressed RSA digital signature, etc. It can protect the privacy of spatial data and kNN query, and verify the reliability of query results. Kim *et al.* [8] proposed a security kNN query method based on homomorphic encryption, and the sorting of query results is realized, but the efficiency is too low. Cheng *et al.* [2] proposed a cipher-text query method based on point-rectangle distance to process kNN query of encrypted spatial data. Lian *et al.* [12] adopted Moore curve for one-way conversion of spatial data, which used AES encryption technology to protect the converted data, and kNN secure query in the conversion spatial is realized.

In the above methods, the secure kNN query of interest points in the cipher-text state is realized. Most of them carry out cipher-text search on the whole location dataset, or only carry out cipher-text location data search in a certain region, the method of dynamically expanding the query area is not used to realize cipher-text data search according to the query requirements of users. A secure kNN query processing method based on extended range of cipher-text search is proposed in this paper, which improves the query efficiency and realizes the approximate sorting of the kNN query results in cipher-text state.

3 The Proposed System Model

In the cipher-text retrieval scheme, the data owner transforms and encrypts the location data, and then outsources it to the cloud server. When user queries location data, a query request is first sent to the data owner. The data owner generates the query trapdoor according to the current location and query requirements of the query user, and then sends it to the query user together with the private key. The query user sends the query trapdoor to the cloud server for query, the kNN interest points are found through the neighborhood grid expansion search in the cipher-text state, which is returned to the query user after being approximate sorted. Finally, the cipher-text query results are decrypted into plaintext with the private key on the user side. The system structure is shown in Figure 1.

As shown in Figure 1, the system structure in this paper consists of location data owner, cloud server and query user, the description is as follows:

Query user (QU): The accessor of location data, has its own location attribute and private key. Users can only use their own private key to decrypt the encrypted location data into plaintext.

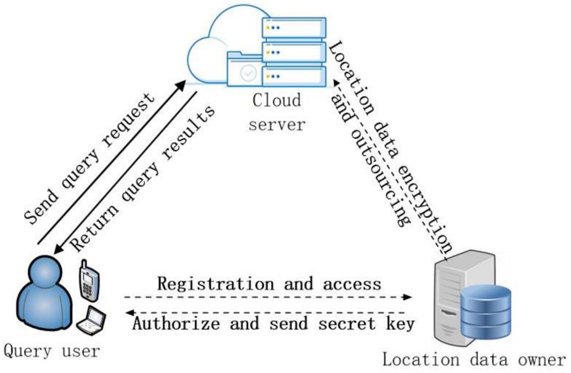

Figure 1: System structure model

Data owner (DO): Transformes the spatial location data, and generates location data records. The location data records are encrypted and outsourced to the cloud server. Moreover, the data owner sets the access rights, and generates the keywords trapdoor and private key.

Cloud server (CS): The provider of cloud service, providing data storage services for data owners. After receiving the user's query request, verifies the user's access rights, retrieves the cipher-text data through the keywords trapdoor, and sorts it approximately.

3.1 Spatial Location Conversion

In the system structure of Figure 1, the data owner obtains the geographic information of the region where the current location is located, as shown in Figure 2(a). The area is divided into $m \times m$ square grids, and all the location points of the area are divided into different grids, as shown in Figure 2(b).

The spatial location data is transformed, and two-dimensional coordinates of location points are converted into binary Morton codes: According to the grid line, along the abscissa from left to right, mark 1 above the grid line, and mark 0 below the grid line. From top to bottom along the ordinate, mark 1 on the left side of the grid line, and mark 0 on the right side of the grid line. The abscissa code values are placed in odd digits, and the ordinate code values are placed in even digits for coding.

3.2 Establishment of Spatial Data Structure

Two-dimensional location coordinates are transformed into binary Morton codes. According to the grid line, all the location points are divided into different grids, and the coordinates of all location points are converted to Morton codes. Because the Morton codes of all location points in the same grid must be the same, the grid area where the current positioning point is located can be found according to the Morton code value.

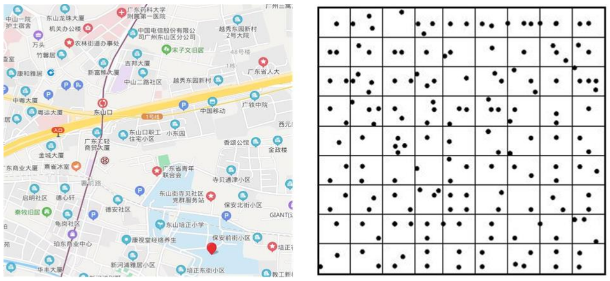
(a) (b)

Figure 2: The location area is divided into several square grids. (a) Selected location area, (b) Grid division.

The system consists of data owner, query user and cloud server. The data owner owns the original database with n location data records, and the attributes of each record include Morton code, location coordinates, and placename information.

3.3 Extended Grid Range of Cipher-text Search

In the cipher-text interest point query, according to the Morton codes of the data record, the grid area where the current position located is found first, and the cipher-text interest points are searched through the keywords of placename information in the grid. When the number of kNN interest points found meets the query requirements, the query results are returned to the query user. When the number of interest points found is less than the query requirement k, the query range is expanded to the adjacent grid, and the grid area adjacent to the current grid is further found until the number of interest points found is larger than or equal to the number of query requirements. Finally, the kNN interest points is returned to the query user and decrypted as plaintext data with the private key on the user side.

Definition 1. *Let R_s represents the selected rectangular area, R_s can be defined as $R_s = \{\mathrm{m \times m, S}\}$. among them,* m *represents the number of rows and columns of the grid divided in the R_s region.* $\mathrm{S} = \{S_1, S_2, S_3, ..., S_n\}$ *represents all the location datasets contained in the R_s region.*

Definition 2. *Let* $\mathrm{L} = \{\mathrm{M}, (\mathrm{x}, \mathrm{y}), \mathrm{W}\}$ *represents the placename information, (*x, y*) represents the location coordinates,* M *represents the Morton code. C_L represents cipher-text location data record.*

Definition 3. *Let* c *represents a positive integer. Let* a *and* b *represent integer, if* $(\mathrm{a} - \mathrm{b})\%\mathrm{c} = 0$*, then integers* a *and* b *are congruent to module* c*, denote by* $\mathrm{a} \equiv \mathrm{b}(\mathrm{mod}\ \mathrm{c})$*.*

Definition 4. *Let* n *represents a positive integer, Euler function refers to the number of positive integers that is less than* n *and is coprime with* n*, recorded as* $\varphi(\mathrm{n})$*.*

4 Algorithm Description and Structure Analysis

4.1 Algorithm Description

The data owner divides the spatial area into several square grids, and all location points are divided into different grids. The Morton codes of all location points are calculated according to the grid lines. Then, the location coordinates and placename information are encrypted by public key encryption mechanism, the Morton codes are scrambled encrypted respectively, and then outsourced to the cloud server. When the user queries the location data, the query request is first sent to the location data owner. The data owner generates a query trapdoor according to geographic information and query keywords of the query user, and sends it to the query user together with the private key. The query user sends the query trapdoor to the cloud server for query. On the server side, the area where the current location located is obtained according to the approximate matching calculation between Morton codes. In the grid region, the kNN interest points are found through placename keywords matching. When the number of interest points found does not meet the query requirements, the search continues through the adjacent grid expansion search until the number of interest points found meets the requirements is.

1) Generate location data record:

   Input: All location points in the area.

   Output: Generate location data record L.

   Step 1: R_s, S, m,$(x_i, y_i) = P_i$.

   Step 2: The area R_s is divided into $m \times m$ square grids.

   Step 3: According to the grid line, along the abscissa from left to right, mark 1 above the grid line, and mark 0 below the grid line. From top to bottom along the ordinate, mark 1 on the left side of the grid line, and mark 0 on the right side of the grid line. The abscissa code value is placed in odd digits, and the ordinate code value is placed in even digits for coding. The Morton codes M Generated.

   Step 4: Repeat step 3, and all the location points in the region are converted.

   Step 5: Extract the keywords of placename information and generate a new placename information list of "keywords+placename information".

   Step 6: Generate location data record $L = \{M, (x, y),\ W\}$.

2) Generate secret key: Select prime numbers p and q, calculate $n = p \times q$, and calculate Euler function $\varphi(n) = \varphi(pq) = \varphi(p) \times \varphi(q) = (p-1)(q-1)$. Randomly select the number e satisfying the condition $1 < e < \varphi(n)$ and $\gcd(e, \varphi(n)) = 1$, then the public

key is $(e,\ n)$. Calculate the private key d, that is $d \equiv e - 1 \bmod \varphi(n)$.

3) Encryption phase: This method adopts the encryption scheme combining public key encryption and scrambling encryption: the location information $\{(x,\ y),\ W\}$ is encrypted with the public key$(e,\ n)$, and Morton code M is encrypted by scrambling encryption mechanism. The algorithm is as follows:

   a. Public key encryption algorithm: $(x,\ y)$ and W of $\{M,\ (x,\ y),\ W\}$ are encrypted with public key encryption algorithm.

   Placename information W and location coordinates $(x,\ y)$ are encrypted with public key $(e,\ n)$ respectively, and the cryptograph data $C_{(x,y)}$ and C_W are obtained. Among them, Equation (1) is the encryption formula of location coordinates $(x,\ y)$, and Equation (2) is the encryption formula of placename information W.

$$C_{(x,y)} = E(x,y) = ((x,y))^e \bmod n \quad (1)$$
$$C_W = E(W) = (W)^e \bmod n \quad (2)$$

   where n is the product of two large prime numbers p and q, that is, $n = pq$, $\varphi(n) = (p-1)(q-1)$; e is a randomly selected prime number, satisfying $1 < e < \varphi(n) - 1$, $(e, \varphi(n)) = 1$.

   b. Scrambling encryption algorithm: M of $\{M,\ (x,\ y),\ W\}$ is encrypted by scrambling encryption algorithm.

      i. For a 32-bit Morton code binary string, $a[i]$ represents each bit value of Morton code.

      ii. Randomly generate a random number j, j must satisfy $0 \le j \le 31$.

      iii. Swap $a[i]$ and $a[j]$.

      iv. Execute $i++$, j does not repeat the value every time. Until i=31, the exchange ends.

   The data record is encrypted by public key encryption algorithm a) and scrambling encryption algorithm b) to generate the cipher-text location data record $C_L = \{C_M, C_{(x,y)}, C_W\}$.

4) Trapdoor generation stage: When querying the kNN interest points, the query user sends a query request with the current location coordinates (x_u, y_u) and the query interest point keywords to the data owner. According to the location coordinate transformation method, the data owner calculates the Morton code M_u, and encrypts it according to the scrambling function. The current location coordinates and query location keywords are encrypted with public key (e, n), and query trapdoor $F = (C_{M_u}, C(x_u, y_u), C_{W_u})$. Finally, the data owner sends trapdoor F and private key d to the query user.

5) Cipher-text query stage: The query user sends the query trapdoor F to the location server for query in cipher-text state. Firstly, the location data in the current region is found by matching calculation between the Morton codes. Then, the kNN interest points are retrieved according to the key words of placename information. When the number of interest points does not meet the query requirement, the interest points in adjacent grid regions are searched through extended range of cipher-text search until the number of interest points meets the query requirement, which is sorted approximately according to the adjacent relationship. At last, the kNN ciphertext location records $\{C(x_u, y_u), C_W\}$ are sent to the query user.

6) Data decryption stage: After receiving the ciphertext result set $\{C(x_u, y_u), C_W\}$, which is decrypted into plaintext set $\{M(x_u, y_u), M_W\}$ with private key (d, n) on the user side. Among them, Equation (3) is used to decrypt the location coordinates, and Equation (4) is used to decrypt the placename information.

$$M_{(x,y)} = D(C_{(x,y)}) = (C_{(x,y)})^d \bmod n \quad (3)$$
$$M_W = D(C_W) = (C_W)^e \bmod n \quad (4)$$

4.2 Algorithm Structure and Analysis

The proposed algorithm consists of location data generation, secret key generation, data encryption, trapdoor generation, cipher-text query, and data decryption. In the stage of location data generation, the selected area is divided into several grids, and the Morton code of each location coordinate is calculated according to the grid line. Because the Morton codes of all location points in the same grid are the same, the spatial region where the current location point located can be determined according to the Morton code value. In order to improve the query efficiency and query accuracy, the keywords of placename information are extracted, such as “school”, “hospital”, “station”, “Hotel”, “shopping mall”, “movie”, “food”, “KTV”, etc., and the placename information records in the format of “keywords+place name information” are generated.

In the data encryption stage, public key encryption mechanism is used to encrypt the placename information and coordinates, and scrambling encryption is used to encrypt Morton codes. Since the Morton code is a string of binary numbers, the scrambling of corresponding digits does not affect the calculation of Hamming distance between two binary strings. For example, for binary strings A = “1011010100” and B = “1001101101”, there is $\sum(A \oplus B) = 5$. Suppose that for A and B, $a[i]$ represents each binary bit. If $a[i]$ and $a[n-i]$ are exchanged with each other, then A_1 = “0010101101”,B_1 = “1011011001”,and $\sum(A_1 \oplus B_1) = 5$. It can be seen that the Hamming distance before and after scrambling is the same. Therefore, scrambling encryption does not affect the calculation of

Hamming distance between Morton codes. In the trapdoor generation phase, the query user sends a query request to the data owner. After authentication, the current location coordinate and query keywords are sent to the data owner. The two-dimensional coordinates of the current user's location is transformed, and encrypted query trapdoor is generated. Among them, location coordinates and placename information are encrypted with public key (e, n), and the Morton code is encrypted with scrambling. The scrambling order used here is exactly the same as the scrambling of the dataset records.

In the extended range of cipher-text search algorithm, firstly, according to the Morton code of the query user's current location, the grid region of the current location is found through approximate matching calculation, and the kNN interest points are found through the approximate matching of place name information in the grid region. If k is the required number of query interest points, and k' is the number of currently location interest points found, there is:

1) When $k' \geq k$, the number of interest points found meets the query requirements, then the search is ended;

2) When $k' < k$, expand the search area and search the interest points in the adjacent grid regions.

In (1), when the number of found interest points k' is larger than the number of required interest points k, and the number of found interest points meets the query requirements, k' interest points are sent to the query user. The query user decrypts location data with the private key.

In (2), when the number of interest points found k' is less than the number of interest points required k, expand the range of adjacent grids by approximate matching between Morton codes, until the number of found interest points meets $k' \geq k$, the search is stopped and k' locations are send to the query user. The query result is decrypted to plaintext with the private key on the user side. As shown in Figure 3.

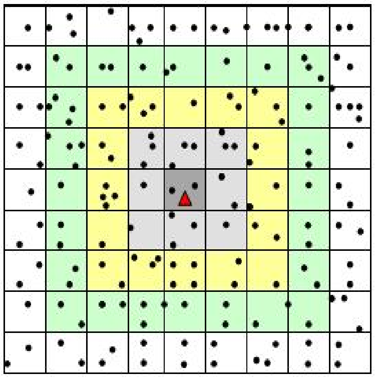

Figure 3: Adjacent grid expansion search

4.3 Safety Analysis

In the encryption process, the public key encryption mechanism is used to encrypt the placename information and location coordinates in the location data record, encrypted location data cannot be decrypted without a private key. Its secrecy is better, which is proved by experiments. The Morton code in data record is obtained by location coordinate transformation. Before location data outsourcing, the Morton code is encrypted through scrambling technique and to ensure the security of location data. In the generation of trapdoor, the data owner transforms the query user's location coordinates according to the generation method of Morton code, the query record is encrypted by public key and scrambling function, the privacy of the query trapdoor is guaranteed. The query process and the calculation of approximate distance are all completed in the cipher-text state. Moreover, in the query process, the grid region and location data are mainly determined through the matching calculation between Morton codes, but Morton codes are obtained through location coordinate conversion and scrambling encryption, which is equivalent to secondary encryption.

After retrieving kNN interest points, only the coordinates and place name information of the cipher-text query record are returned to the query user. The encrypted Morton codes are not returned to the query user, nor are decrypted. In the algorithm, the data conversion and scrambling encryption are confidential to the location data server and query users, and the privacy of location data is guaranteed.

5 Experimental Results and Analysis

In the experiment, synthetic dataset and real dataset are used for performance analysis, 10k records are randomly generated on synthetic dataset. we select the real location map data of Guangzhou in Google map as the real dataset, and divides the area into 32 × 32, the main parameter k of the experiment is $1 \leq k \leq 30$.

The hardware environment of the experiment is 3.2 GHz Intel Core i5 processor with 8 GB of memory. The operating system is windows 10, which adopts my eclipse development platform and is implemented in Java programming language. Table 1 shows the default parameter configuration of the experiment.

Table 1: Experiment default parameter configuration

Parameter	Value
k	$[1, 30]$
Number of grids	32×32
Location set	100000
Spatial scope (km^2)	10×10

5.1 Algorithm Performance Analysis

Firstly, the performance of the proposed method is verified by experiments. When the security parameter is 1024, the encryption time, decryption time, trapdoor generation time and query time of the proposed method are shown in Figure 4.

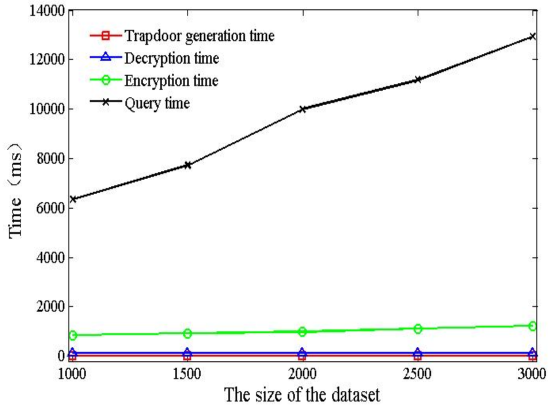

Figure 4: The performance of the proposed method

As can be seen from Figure 4, the time of cipher-text query is the most. This is because the proposed method needs to query the interest points in the grid where the current position is located in the cipher-text state. When the number of interest points could not meet the user's query requirements, it is necessary to expand the query range to the neighborhood grid until all interest points that meet the query requirements are retrieved, and the whole search process takes more time. The generation time of trapdoor is the least. The data owner transforms the current location coordinates provided by the query user, and generates the query trapdoor according to the public key encryption algorithm and scrambling function. Its algorithm is relatively simple, so it costs less time. As can be seen from Figure 4, Data encryption takes more time than data decryption. This is because data encryption encrypts all datasets, while data decryption decrypts only k interest points.

5.2 Comparison of Query Efficiency

5.2.1 Comparison of Query Time with the Dataset

When the security parameter is 1024, we compare the query processing time of SkNNm method [4], SkNNI method [8], SkNNb method [4] and the proposed method on different datasets through experiments, as shown in Figure 5.

As can be seen from Figure 5, the dataset is from 2k to 10k, the query processing time of several methods increases with the increase of the dataset. With the increase of dataset, the data query processing time of SkNNm method increases the fastest, the data query processing time of SkNNI method increases faster than that of SkNNb method, and the growth of the proposed method is the slowest. As can be seen from Figure 5, in the same size of the dataset, the data query processing time of SkNNm method is the most, SkNNI method takes more time than SkNNb method, the query processing time of the proposed method is less than that of the other three methods.

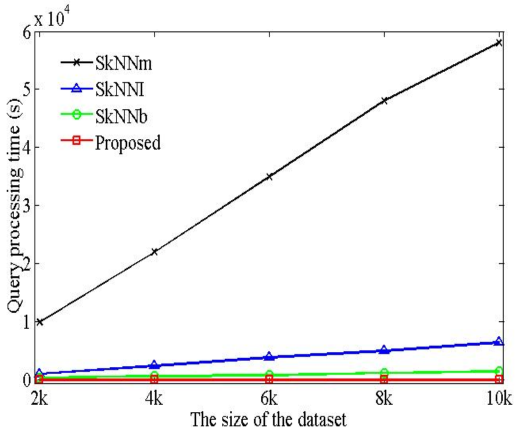

Figure 5: Query processing time varies with the size of the dataset

5.2.2 Comparison of Query Time with the Interest Points

When the security parameter is 1024 and the dataset size is 5000, the query efficiency of the proposed method is compared with SkNNm method, SkNNI method and SkNNb method through experiments. The results are shown in Figure 6.

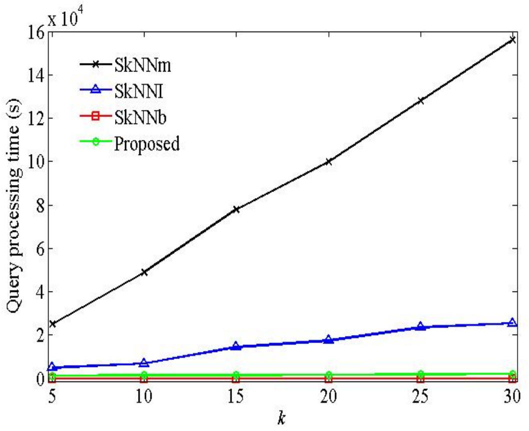

Figure 6: Query processing time varies with the number of the interest points

As can be seen from Figure 6, the query processing time of SKNNm method and SKNNI method increases with the increase of the interest points. With the increase of the number of the interest points, the query processing time of SKNNb method and the proposed method increases less and remains stable. As can be seen from Figure 6, under the same dataset, the query processing time of SKNNI method and SKNNm method are more, and the query processing time of SKNNm method is the most. Among these methods, the query processing time of the proposed method is the least. Moreover, when the dataset is larger, the efficiency advantage of the proposed method is more obvious.

Experimental comparison shows that the proposed method not only meets the effect of privacy protection as much as possible, but also has higher query efficiency, more stable performance, and can effectively improves the quality of location service.

6 Conclusions

Aiming at the problems of low query efficiency and low location service quality of most location privacy protection scheme based on cipher-text query, a secure kNN query processing method of cipher-text extended search is proposed to improve the effect of privacy protection and location service quality from three aspects. Firstly, the spatial location is transformed, and the two-dimensional position coordinates are transformed into Morton code, which is convenient for data encryption and cipher-text approximate calculation. Then, the encryption scheme combining public key encryption and scrambling encryption is used to encrypt the location data record, which is convenient for cipher-text search and calculation. Finally, using the extended range of cipher-text search scheme, the kNN interest points are queried according to the query requirements and conditions. The security of the proposed method is analyzed in this paper. The performance of the proposed method is analyzed through experiments, and compares the query processing time with SkNNI method, SkNNm method and SkNNb method. Experimental results show that the proposed method can effectively improve query efficiency and location service quality. This scheme mainly considers the location privacy protection of snapshot query. In the next step, we will study the location privacy protection of continuous query.

Acknowledgments

This work is supported by the National Science Foundation of China (No. 61762059, 61862041), Natural Science Foundation of Gansu (20JR5RE634), University Innovation Foundation of Gansu (2021A-291), and Natural Science Foundation of Tianshui (2020-FZJHK-2664). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which will improve the presentation.

References

[1] Z. A. Almusaylim and N. Z. Jhanjhi, "Privacy protection of user in location-aware services of mobile cloud computing," vol. 111, pp. 541–564, 2020.

[2] X. Cheng, S. Su, Y. Teng, and et al, "Enabling secure and efficient knn query processing over encrypted spatial data in the cloud," vol. 8, pp. 3205–3218, 2015.

[3] G. Du, L. Zhang, C. G. Ma, and et al, "Location privacy protection method based on attribute-based privacy information retrieval," vol. 42, pp. 680–686, 2021.

[4] Y. Elmehdwi, B. K. Samanthula, and W. Jiang, "Secure k-nearest neighbor query over encrypted data in outsourced environments," in *Proc of the 30th International Conference on Data Engineering(ICDE)*, p. 664–675, Chicago, April 2014.

[5] Y. Fu, Y. H. Yu, and X. P. Wu, "Differential privacy protection technology and its application in big data environment," vol. 40, pp. 157–168, 2019.

[6] Z. Y. Hu, S. L. Liu, and K. F. Chen, "Privacy-preserving location-based services query scheme against quantum attacks," vol. 17, pp. 972–983, 2020.

[7] H. Kido, Y. Yanagisawa, and T. Satoh, "Protection of location privacy using dummies for location-based services," in *Proc of the 21th International Conference on Data Engineering Workshops*, pp. 1248–1252, Tokyo, April 2005.

[8] H. I. Kim, H. J. Kim, and J. W. Chang, "A secure knn query processing algorithm using homomorphic encryption on outsourced database," No. 123, pp. 101602–101622, 2019.

[9] L. Li, Z. J. Lv, X. H. Tong, and R. H. Shi, "A dynamic location privacy protection scheme based on cloud storage," vol. 21, pp. 828–834, 2019.

[10] Z. D. Li, W. M. Li, and Q. Y. Wen, "An efficient blind filter: Location privacy protection and the access control in fintech," No. 100, pp. 797–810, 2019.

[11] Z. P. Li and D Wang, "Achieving one-round password-based authenticated key exchange over lattices," vol. 2019, pp. 1–14, 2019.

[12] H. Lian, W. Qiu, D. Yan, and et al, "Privacy-preserving spatial query protocol based on the moore curve for location-based service," No. 96, pp. 1–16, 2020.

[13] Z. Mei, H. Zhu, Z. Cui, and et al, "Executing multi-dimensional range query efficiently and flexibly over outsourced ciphertexts in the cloud," No. 432, pp. 79–96, 2018.

[14] B. Niu, S. Gao, F. Li, and et al, "Protection of location privacy in continuous lbss against adversaries with background information," in *Proc of the 3th International Conference on Computing, Networking and Communications (ICNC)*, pp. 1–6, Sanya, April 2016.

[15] C. Song, Y. D. Zhang, W. P. Peng, and et al, "Research on location privacy protection scheme based

on similar trajectory replacement," vol. 43, pp. 135–142, 2020.

[16] G. Sun, L. J. Song, D. Liao, and et al, "Towards privacy preservation for,".

[17] L. Sweeney, "k-anonymity: A model for protecting privacy," vol. 10, pp. 557–570, 2002.

[18] J. Wang, C. R. Wang, J. F. Ma, and H. T. Li, "Dummy location selection algorithm based on location semantics and query probability," vol. 41, pp. 53–61, 2020.

[19] T. C. Wang, Y. Liu, X. Jin, and et al, "Research on k-anonymity-based privacy protection in crowd sensing," vol. 39, pp. 170–178, 2018.

[20] X. Y. Xia, Z. H. Bai, J. Li, and R. Y. Yu, "A location cloaking algorithm based on dummy and stackelberg game," vol. 42, pp. 2216–2232, 2019.

[21] J. Xu, X. Tang, H. Hu, and et al, "Privacy-conscious location-based queries in mobile environments," vol. 21, pp. 313–326, 2010.

[22] G. H. Yan, T. liu, X. J. Zhang, and et al, "Service similarity location k anonymity privacy protection scheme against background knowledge inference attacks," vol. 54, pp. 8–18, 2020.

[23] J. M. Yang, Q. L. Wu, Z. Y. Qu, and et al, "Density clustering algorithm based on extended range query," vol. 34, pp. 2938–2992, 2017.

[24] S. Yang, S. Tang, and X. Zhang, "Privacy-preserving k nearest neighbor query with authentication on road networks," No. 134, pp. 25–36, 2019.

[25] M. Zeng, K. Zhang, J. Chen, and H. F. Qian, "P3gq: A practical privacy-preserving generic location-based services query scheme," vol. 51, pp. 56–72, 2018.

[26] M. J. Zeng, Z. L. Cheng, X. Huang, and et al, "Spatial crowdsourcing quality control model based on k-anonymity location privacy protection and elm spammer detection," vol. 10, pp. 1–10, 2019.

[27] J. D. Zhang and C. Y. Chow, "Enabling probabilistic differential privacy protection for location recommendations," vol. 14, pp. 426–440, 2021.

[28] S. B. Zhang, X. Li, Z. Y. Tan, T. Peng, and G. J. Wang, "A caching and spatial k-anonymity driven privacy enhancement scheme in continuous location-based services," No. 94, pp. 40–50, 2019.

[29] Y. B. Zhang, Q. Y. Zhang, Z. Y. Li, and et al, "A k-anonymous location privacy protection method of dummy based on geographical semantics," vol. 21, pp. 937–946, 2019.

[30] C. L. Zhou, Y. H. Chen, H. Tian, and et al, "Location privacy and query privacy preserving method for k-nearest neighbor query in road networks," vol. 31, pp. 471–492, 2020.

Biography

Yong-Bing Zhang He is currently a Ph.D. student in Lanzhou University of Technology, and worked at school of Gansu Institute of Mechanical & Electrical Engineering. He received his master degree in electronic and communication engineering from Lanzhou University of Technology, Gansu, China, in 2015. His research interests include network and information security, privacy protection.

Qiu-Yu Zhang Researcher/PhD supervisor, graduated from Gansu University of Technology in 1986, and then worked at school of computer and communication in Lanzhou University of Technology. He is vice dean of Gansu manufacturing information engineering research center, a CCF senior member, a member of IEEE and ACM. His research interests include network and information security, information hiding and steganalysis, multimedia communication technology.

Yi-Long Jiang Professor, graduated from Shanghai Technology University in 1989, and then worked at school of Gansu Institute of Mechanical & Electrical Engineering. His research interests include embedded control technology, intelligent manufacturing intelligent, manufacturing technology.

Jun Yang Associate professor, graduated from Xi'an University of Finance and Economics in 2003, and received his master degree in educational technology from Northwest Normal University, Gansu, China, in 2010. He worked at school of Gansu Institute of Mechanical & Electrical Engineering. His research interests include information security, information management.