ATLANTIS PRESS
Advances in Computer Science Research, volume 89
2019 International Conference on Wireless Communication, Network and Multimedia Engineering (WCNME 2019)

Nearest Neighbor Secure Query Research for Spatial Road Network Data

Ying Li
Maanshan Teacher's College, Ma'anshan, CHINA

***Abstract*—The nearest neighbor secure query of spatial road network data is a combination of cloud storage technology and geographic location-based services. According to the position information of interest points in space, the Voronoi diagram is used to construct a polygon block to divide space into multiple adjacent spatial units and replace the space midpoint information with block information for achieving road network data position information protection. The spatial region can be divided into multiple sub-regions based on the spatial unit. After receiving the request from the user to issue a query, the part half cut method (PHC) is proposed to divide the entire space area according to the attributes of the interest points. And then using a structured way to create a memory index by Quad-tree and R-tree, then realize nearest neighbor query. In order to improve the query efficiency, the memory hash table for establishing space units and partitions is proposed, and a fast search is performed by directly searching the storage locations of the space units. The comparative experiments show that PHC has the highest query efficiency.**

Keywords—location-based services; information protection; PHC; nearest neighbor query

I. INTRODUCTION

In recent years, due to the rapid development of mobile Internet, the spatial data service represented by the query of European space and road network space has been widely used and brought great convenience to users. However, this technology also has hidden dangers in information security. How to ensure the privacy of user information while realizing spatial data query? Such research has become a research hotspot in the field of spatial data applications.

In recent years, the research results in this field at home and abroad are as follows: In 2009, Wong et al. implemented ciphertext-based k-nearest neighbor calculation by establishing virtual dimensions on the server side, but the overhead of traversing ciphertext data points based on virtual dimensions Larger, so the efficiency of this scheme is very low, it is not suitable for application services [1]; in 2010, Hu et al. proposed a data privacy protection method for ciphertext fuzzy query of strings, which has certain privacy protection and query. The practical significance, however, only for string data processing does not meet the needs of query diversification [2]; in 2015, Huang et al. through the silent cascade to balance the delay of users in the mixed area and application area, to achieve privacy query of encrypted data, however, when the data does not pass through the mixed area, its pseudonym does not change, so the pseudonym can be cracked through its location, resulting in data encryption failure [3]; in 2013, domestic scholars Chan et al. proposed three perspectives on cloud computing security. Research on cloud computing data security privacy protection provides ideas and methods [4]; in 2013, Ken et al. based on geographical location external services The multi-dimensional trajectory anonymity method in personal mobile data collection is studied. This method provides a more advanced idea for the study of personal motion trajectory encryption, and more effective encryption processing of personal information [5]; in 2015, Lin et al proposed Active diffusion type location information privacy protection method [6], this method will use the location information sharing mechanism to achieve adjacent location sharing, which can reduce the user's dependence on the server, thereby enhancing the security of the user's mobile location information, and ensuring the security. Cloud data and user security, but the method is less efficient. Based on the existing research, this paper proposes a Secure Data Nearest Neighbor (SDNN) query method for spatial network data, which can meet the needs of both information security query and query efficiency.

II. RELATED TECHNOLOGIES

A. Nearest Neighbor Query

The Nearest Neighbor (NN) query is a type of spatial data query. The road network data nearest neighbor query is to query and obtain the point p(p ∈ K) closest to the query point q in the data set according to a given query point q in a given road network spatial data set K. The point p closest to the query point q in the data set K is the nearest neighbor of the point q, and the p point is the nearest neighbor of q. The road network spatial data set K contains a plurality of points of interest (POI), and |p1, p2| represents the distance between two points p1 and p2 in the space, and the nearest neighbor of the road network data is based on k nearest neighbors (K Nearest Neighbor, kNN) The concept proposed by the problem, definition 1 represents the meaning of the kNN query:

Definition 1: In the road network spatial data set K, given the target query point p, find the k-nearest neighbor of p, the query is to find the k POIs closest to the target point p in K, and use the set kNN(p) Said, as shown in Equation 1:

$$kNN(p)=\{ pi\in K\} |p1, p|, |p2, p|, |p3, p|, \ldots, |pk, p| \quad (1)$$

When k in the k-nearest neighbor is 1, the k-nearest neighbor query is the nearest neighbor query.

Definition 2: In the road network spatial data set K, given the target query point p, find the nearest neighbor of the p point.

The nearest neighbor query is to obtain a Candidate point set that is closest to the p point in the data set K, and then obtain the nearest neighbor q by determining the distance from each candidate point to the p point in the candidate point set. Use the set CS(K) to represent as shown in Equation 2:

$$CS(K)=\{pi \in K\}, \min(|p, pi|)=|p , q| \quad (2)$$

B. Voronoi Diagram

A Voronoi diagram is a group of continuous polygons formed by connecting two adjacent points in the data set into a straight line and then drawing a vertical bisector to form a polygon. Voronoi diagram based on Delaunay triangulation algorithm plays an important role in computational geometry. The first step in the construction of Voronoi diagram is to construct the dual Delaunay triangulation network of the Voronoi diagram. The second step is to find the center of the circle formed by all the triangles in the Delaunay triangulation network. The third step is to connect all the adjacent centers to obtain the Voronoi diagram.

According to the characteristics of dividing regions by distance, the Voronoi diagram processes the spatial regions. The Voronoi diagram is constructed by the vertical bisector between POI in the spatial region. In the spatial region, it is divided into multiple Voronoi units based on points in the spatial data set K.

C. Space Division Based on Voronoi Diagram

The Voronoi diagram has the feature of dividing adjacent regions according to distances. The Voronoi diagram can be used to divide the spatial data set into n spatial units. Different id Numbers are used to distinguish and name different units. Given the target query point q, get the Voronoi unit closest to the query point q, and the POI in the Voronoi unit is the nearest neighbor of the target query point. Each Voronoi unit corresponds to an id number, as shown in table 1, and the unit where the POI is located corresponds to its id number.

TABLE I. CELL LOCATION OF POI AND CORRESPONDING ID NUMBER

POI location	id number
p_1	C_1
p_2	C_2
p_3	C_3
……	……
……	……
p_i	Ci

In the road network data set, the Voronoi diagram nearest to the target point P is represented as ND(P), and the point in the nearest Voronoi unit is the nearest neighbor point to the target query point relative to other points in the data set K. Although the method of dividing space by Voronoi cell can solve the nearest neighbor query problem, if the process of solving the nearest Voronoi cell with this method requires traversing all POI in the spatial data set, it will incur a huge cost in terms of time and space, so the optimization strategy should be adopted to deal with spatial regions.

III. SPACE PARTITION AND STORAGE MODE

In order to reduce the cost of traversing the spatial data set and improve the query efficiency, a query optimization strategy was proposed based on the space division of Voronoi diagram. The query cost was reduced by partitioning the space, and the road network data was queried in an appropriate area according to the location of the target points. Therefore, a space partitioning and data storage strategy was proposed, and the whole space was segmented into B1 and B2, , Bn and other partitions using the space equipartition method, as shown in Figure 1.

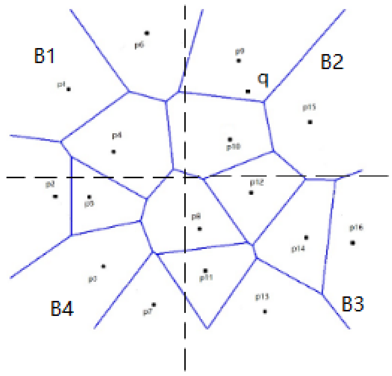

FIGURE I. SPATIAL REGIONS ARE DIVIDED USING THE SPATIAL EQUIPARTITION METHOD

Space partitioning method (the part half cut method, PHC) split is based on the spatial location, first of all, according to the location according to the space width of the median line of the space is divided into uniform four blocks, and then use this way to iteration, until the integral density satisfies a certain requirement to stop the iteration and establish the index of the tree structure, using 4 - tree structure to store data, each POI belongs to a block in a tree structure, mutually disjoint between block, block contains multiple POI and a partition is divided to different blocks, So using POI tagging of partition blocks, according to the position of the block to block is arranged to build 4 - tree, the leaves of the tree node after segmentation for the space block, not shard interloper in spatial segmentation, otherwise there will be a partition is occupied by a multiple blocks together, with POI tagging each node and construct the index tree, based on 4 - complete nearest neighbor query tree. As shown in Figure 2, the POI is stored in a 4-fork tree structure, with leaf nodes as the underlying structure to divide the space.

Use space partitioning method of segmentation space, the space is divided into 4 n space area, because the method divided the space has 4 n leaf nodes, so using 4 - quadtree structure to structured data storage and index, on a leaf node in the structure of 4 - tree storage divided the space block, after 4 - each leaf node in the tree to the POI said, used to indicate in the space area of POI, only when the query need to deal with data for regional block, at query time reduced the POI more times, improve the query performance.

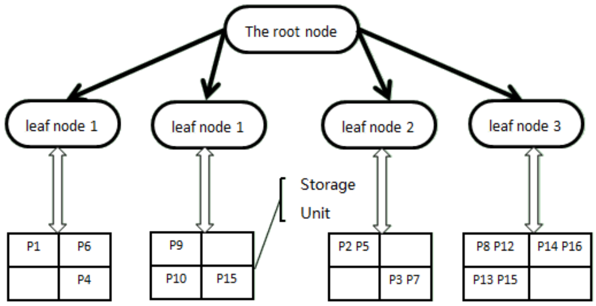

FIGURE II. USE A 4-FORK TREE TO STORE THE SPACE POI

IV. SAFE NEAREST NEIGHBOR QUERY PROBLEM DESCRIPTION

After the spatial data is structured, the query list is constructed based on the index table structure for the POI storage structure and the Voronoi partitioning unit. Since each POI belongs to a unique structural partition, a structural partition generally contains at least one POI, and one partition structure may be included in multiple Voronoi partitioning units, and there may be multiple in one Voronoi partitioning unit. The partition structure, as shown in Table 2, is a memory hash table established between the structure partition and the partition unit, and the Voronoi partition unit is used to indicate that the structure partition can display the most recent interest point unit more intuitively. The tree structure stores structured data. Each leaf node of the tree represents a structural partition of the POI, and the Voronoi partitioning unit overlapping with the structure partition is used to determine the POI to be queried.

TABLE II. HASH TABLE BETWEEN STRUCTURE PARTITION AND PARTITION UNIT

Serial number	Space area number	Voronoi unit
1	B_1	C_1、C_2、……、C_m
2	B_2	C_p、……、C_r
……	……	……
n	B_n	C_x、……、C_z

V. IMPLEMENTATION OF SECURE NEAREST NEIGHBOR QUERY METHOD BASED ON SPACE PARTITION

The idea of the index model based on hash table is the index structure model based on road network data information storage, including Voronoi unit information and spatial region, the structure of hash is used to map information, and the region information and corresponding unit information are stored in index table structure. As shown in Figure 3, the index table is stored. One spatial region corresponds to multiple Voronoi units. Based on the storage structure, all corresponding Voronoi units in a region can be quickly found. Table 1 can find corresponding for each Voronoi unit. POI, according to the nature of Voronoi, can implement nearest neighbor queries.

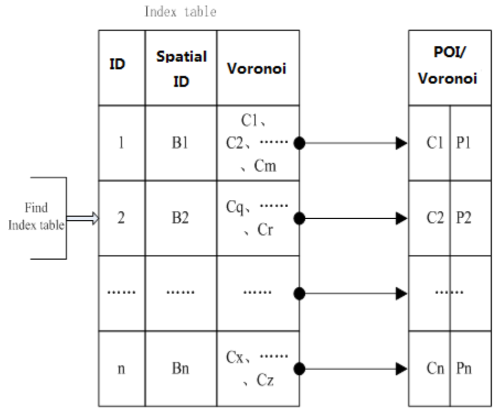

FIGURE III. SCHEMATIC DIAGRAM OF THE INDEX STRUCTURE

Secure Data Nearest Neighbor Query Algorithm (SDNN), which uses a tree structure to store structural partitions partitioned by spatial data, builds an index structure based on structural partitions and Voronoi elements, and implements iterative search and corresponding lookups. Neighbor query. First, iteratively finds the leaf node where the query point q exists from the root node in the TREE, and obtains the structure partition where q is in the TREE. The corresponding Voronoi unit is found through the structure partition and the index table of the Voronoi unit. If the neighboring unit Ci is found to correspond to a plurality of Voronoi units, it is necessary to perform position comparison to obtain the nearest neighbor Ci. According to the characteristics of the Voronoi unit, it can be known that a POI corresponds to a Ci, and the nearest neighbor point can be obtained, and the shortest distance is used. The query algorithm finds the POI closest to the query point, and the nearest neighbor query algorithm is shown in Table 3:

TABLE III. SDNN QUERY ALGORITHMS

Algorithms (SDNN Query Algorithms)
IN：query point *q*,Voronoi cell C, TREE T , partition B ,index table L; OUT：NN point P_i
1 The client uses an encryption scheme to encrypt *q* into E(*q*) and use the Voronoi method to different encrypted *POI* E(P_i) to different cells ; 2 send E (*q*) to the server;
3 Decrypt the E (*q*) to *q* on the server side;
4 search the partition of *q* the in TREE T;

5 while（the partition != leaf node in the TREE T）
6 {
7 If（*q* in the partition B）
8 {
9 get the partition B;
10 }
11 }

12 FIND（B in index table L）//the cell Ci can be find by FIND 13 get the NN Voronoi cell ND(*q*)by the location information of *q* is C_i;
14 find the NN POI E(Pi) in C_i;
15 E(P_i) was encrypted by E, the server can get E(P_i); 16 return E(P_i);
17 send E(P_i) to the client and decrypt E(Pi) become P_ii;
18 the result P_i is got by us;

VI. Comparison of Experimental Results Safe Nearest Neighbor Query Problem Description

The nearest neighbor query method of road network data security implements data index based on partition optimization strategy. In addition to the spatial equalization method (PHC) proposed in this paper, the traditional partition methods include the iterative division between points method (IBP) and the half split method (HM). The following are separately from the partition. The influence of the number of interest points on partitioning time and query time is compared through experiments.

When the number of partitions is determined, in order to analyze the influence of the number of interest points on the partition time, a data set of 20,000 to 150,000 points is selected to carry out the experiment. As shown in Figure 4, the time consumption of three partition methods, namely, the spatial average partition method (PHC), the iterative partition method based on interest points (IBP) and the half-fold partition method (HM), is linear distribution. PHC calculates the unit length of each partition. As the number of data sets increases rapidly, the partition time of IBP increases faster than that of HM. From the experimental results, it can be seen that the efficiency of the folded partition method is relatively lower than that of the spatial equalization method, and the efficiency of the iterative partition method based on interest points is the lowest.

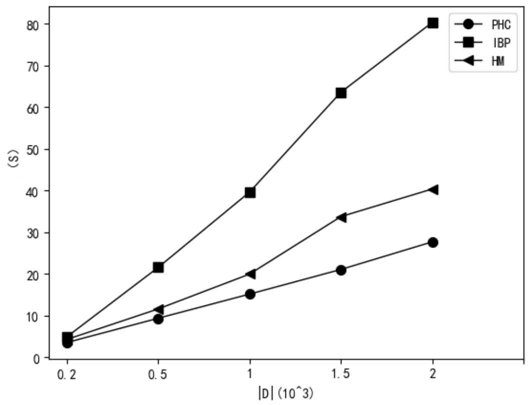

FIGURE IV. PARTITION TIME AND PARTITION SIZE

VII. Conclusions

This paper puts forward the problem of the secure nearest neighbor query of road network data and designs corresponding solutions. Through the comparative analysis of experimental results, it can be seen that the secure nearest neighbor query method of cloud-end road network data has security and validity. Using SDNN algorithm of spatial equalization method can reduce the number of POI returned by the query, reduce query overhead, and realize privacy protection while improving query performance.

In the future work and research, the following aspects can be further studied: (1) In practical applications, as user needs may not only be the nearest neighbor query requirements, how to solve the problem of secure k-nearest neighbor query on the road network has certain research significance; (2) If data sets not only consider location attributes, but also other attributes of data, how? It is also meaningful to do more detailed research based on various attributes.

References

[1] Wong W K, Cheung W L, Kao B, et al. Secure kNN computation on encrypted databases[C]// ACM SIGMOD International Conference on Management of Data. ACM, 2009:139-152.

[2] Hu H, Xu J. 2PASS: Bandwidth-Optimized Location Cloaking for Anonymous Location-Based Services [J]. IEEE Transactions on Parallel & Distributed Systems, 2010, 21(10):1458-1472.

[3] Lin X, Lu R. Vehicular Ad Hoc Network Security and Privacy [M]. Wiley-IEEE Press, 2015, 32(1):91-118.

[4] Chan N N, Dang T K. On efficient processing of complicated cloaked region for location privacy aware nearest-neighbor queries [C].International Conference on Information and Communication Technology. Springer-Verlag, 2013:101-110.

[5] Ken M, Kazuhiro M, Hiroshi M. Privacy-preserving Publishing of a Pseudonym-based Trajectory Location Data Set (Extended Abstract) [J]. 2013, 2013:1-8.

[6] Lin X, Lu R. Vehicular Ad Hoc Network Security and Privacy [M]. Wiley-IEEE Press, 2015, 32(1):91-118.