A k-Nearest Neighbor Search Algorithm for Privacy Preservation in Outsourced Spatial Databases

Miyoung Jang, Min Yoon, and Jae-Woo Chang

Department of Computer Engineering, Chonbuk National University
567 Baekje-daero, Deokjin-gu, Jeonju-si, Republic of Korea
{brilliant, myoon, jwchang}@jbnu.ac.kr

Abstract. Traditional spatial databases owners outsource their resources to a cloud computing environment so that they can reduce cost for storage and management. However, the issue of privacy preservation is crucial in spatial database outsourcing since user location data is sensitive against unauthorized accesses. Existing privacy-preserving algorithms may reveal the original database from encrypted database and the query processing algorithms fall short in offering query processing on road networks. In this paper, we propose a privacy-preserving query processing algorithm which performs on encrypted spatial database. A new node-anchor index is designed to reduce unnecessary network expansions for retrieving k-nearest neighbor (k-NN) objects from a query point. Performance analysis shows that our algorithm outperforms the existing algorithm in terms of query processing time and the result size.

Keywords: Outsourced spatial database, Location-based services, K-nearest neighbor search algorithm, Privacy, Query processing

1 Introduction and Related Work

Due to the advancement of cloud computing technologies, the research on outsourced databases has been spotlighted. In the outsourced database environment, a data owner attempts to outsource his/her database to a service provider, in order to reduce cost for data storage and management. Only both authorized users and a data owner are allowed to access the outsourced data, but not the third parties. With the popularity of LBS, the traditional spatial databases owners want to outsource their databases to the service provider so that they can manage the spatial data efficiently. In this context, the issue of privacy preservation is very important in spatial database outsourcing because a user's location data is valuable and sensitive against unauthorized accesses.

In the literature, protecting outsourced spatial database has been actively studied [1-3]. The distance-oriented transformation technique [3] is proposed where the metric preserving transformation (MPT) converts an original spatial database into a numeric database by using a distance between POIs. Hence, the service provider cannot assume the original coordinate of POIs while guaranteeing the accuracy of the query processing result. However, because this technique only considers Euclidean distance between POIs, it cannot be directly applied to k-NN query processing in road networks. In real location-based applications, users move along with the road network and so it is crucial to consider road network restrictions for k-NN query processing.

In this paper, we propose a spatial database encryption scheme that produces a transformed database from an original database by using network distances among POIs. We randomly select anchors for grouping POIs and devise an anchor-node index in order to store both network distances between POIs as well as those between anchors and network nodes. To generate the index, we encrypt both distances and anchor information by using an order preserving encryption scheme (OPES[4]). In addition, we propose a novel k-NN query processing algorithm performs on transformed data in road network by considering not only the spatial data privacy but also the query processing efficiency. Our anchor-node index can greatly reduce the network expansion cost when retrieving the k-NN POIs from the user's location. Moreover, we reduce the number of retrieved POIs by incrementally reducing a k-NN search range.

The rest of the paper is organized as follows. Section 2 describes proposed method. Performance analysis is provided in section 3. Finally, section 4 concludes this research with future research directions.

2 K-NN Query Processing Algorithm for Outsourced Databases

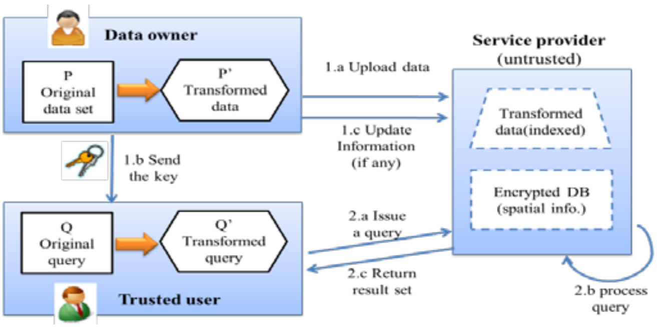
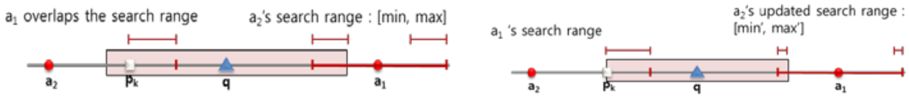


Figure 1. Overall architecture

Figure 1 depicts the overall architecture of our spatial data encryption and k-NN query processing. In this paper, we improve MPT algorithm to support k-NN query processing in the road network. The basic idea behind our method is to randomly choose a subset of the database a set of anchors and then assign each POIs in the original database to its nearest anchor. For each POI, we compute network distance from its anchor and then apply OPE on the distance value. These distance values are sorted and stored in the service provider in order to provide the k-NN query processing. The benefit of using OPE is that it hides the original distance values and while allowing comparisons to be correctly evaluated at the server side. After encrypting the original database, a data owner generates an anchor-node index which stores distances between anchors and network nodes that lay on their anchor ranges. Hence each node stores its nearest anchor id and the distance from the anchor. By using the anchor-node index, a query issuer can retrieve the nearest anchor from its location by simply extending a network edge where he/she stands. In the query processing phase, we set the search range based on the value of anchor's coverage, distance between a query and an anchor, and the distance between a query and k-NN POI. First, a trusted user applies the encryption method to its location data by using the same data encryption metric. Second, the trusted user finds the nearest anchor from its location and set the search bound by using the distance to the anchor. The user can easily retrieve the nearest anchor from its location by using anchor-node index. Third, after finding the nearest anchor, the trusted user requests r(%) of sample POIs from the nearest anchor to the service provider. Then, the user sets the tight search range of anchors to reduce the retrieved number of POIs. The search range can be defined by using the distance of the sampled k-NN POI from its anchor, i.e., dist(p_k, a_i), the distance between the query point and the anchor, i.e., dist(a_i, q), and the anchor range (a_i, r_i). There are two cases of search range setting for anchors considering the position of q. When a query point resides outside the nearest anchor range, the query range can be defined as $[\min, \max] = [dist(a_i, p_k), dist(a_i, r_j)]$, and if a query point is located within the nearest anchor range, the search range for k-NN POIs can be defined as $[dist(a_i, p_k), dist(a_i, q)]$. In order to guarantee a correct query result, the query issuer should retrieve all anchors that overlap the query range. When an additional anchor is found, the user repeats the first round for the anchor found. If the anchor returns POI that is closer to the query point than p_k in a_q, user should update the query range by using the newly retrieved k-NN POI. Figure 2 describes the basic idea behind the query range update. Since the service provider only stores the encrypted value, it retrieves the set of potential candidates that can be the k-NN of query point by using OPE values. The user decrypts the result and computes the real network distance from candidate POIs and prunes out unnecessary POIs. When computing the real network distances from the query to the POIs, we make use of a heuristic for network expansion in order to reduce the number of unnecessary network expansion. Finally, the user sends the search range of found anchors and receives the result POIs.


a_1 overlaps the search range
a_2's search range : [min, max]
a_2
p_k
q
a_1
a_1 's search range
a_2's updated search range : [min', max']
a_2
p_k
q
a_1

a) k-NN search range with the nearest anchor a1 (b) update of the k-NN search range with a2

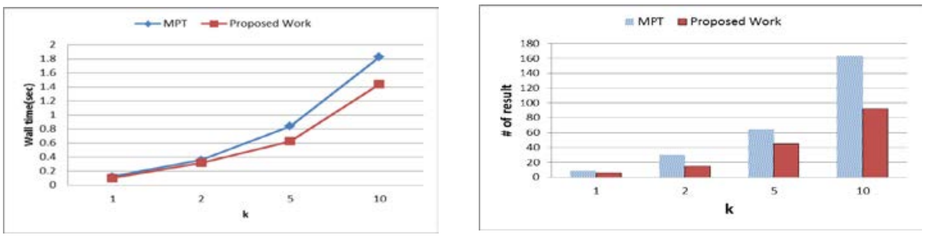
Figure 2. Update of k-NN search range

3 Experimental evaluation

We compare our algorithm with the existing MPT algorithm in terms of preprocessing time for query processing time and returned candidate set size. For fair comparison, we extend the existing MPT method to support k-NN query processing in road network environment in a naïve way. The naïve extended MPT indexes the road network and POIs by R-tree which is commonly used in the traditional spatial network query processing. Both algorithms are implemented in Visual studio C++

2005 on Window XP professional sp3 operating system with Intel Core(TM) i3 CPU 530 @ 2.93GHz and 2GB RAM. The San Francisco bay area map data consisting of 220,000 edges and 170,000 nodes are used. In addition, we made use of 22,025 POI data which is generated by using network-based generator of moving objects [5]. Every result reported in this paper is the average value of 100 k-NN query processing.

Figure 3 plots query processing time and query result size for both algorithms with different k. Our algorithm greatly reduced the query processing time since we utilize the anchor-node index. This is because when retrieving the nearest anchor from the user's location, our algorithm does not need to expand the network but only search the anchor-node index that stores the network distance between each node and its nearest anchor. On the other hand, the extended MPT performs direct expansion of network edges from the user's location to the nearest anchor. In terms of query result size, our algorithm retrieved fewer candidates then that of the exiting algorithm. This is because our algorithm gradually reduces the anchor search range by updating the distance between a k-NN POI and the user.



(a) Query processing time varying k (b) Query result size varying k

Figure 3. Query processing performance

4 Conclusion

In this paper, we design a spatial database encryption scheme that produces a transformed database from the original database by using network distances among POIs. As a future work, we plan to study on a pruning technique to improve the performance of our method by reducing the size of the returned candidate set.

References

1. D. Sacharidis, K. Mouratidis and D. Papadias: "k-Anonymity in the Presence of External Databases," IEEE TKDE, 2010, 22, (3), pp.392-403
2 M. L. Yiu, G. Ghinita, C. S. Jensen and P. Kalnis: "Enabling Search Services on Outsourced Private Spatial Data," VLDB Journal, 2010, 19, (3), pp.363-384
3 M. L. Yiu, I.Assent, C. S. Jensen and P. Kalnis: "Outsourced Similarity Search on Metric Data Assets," IEEE TKDE, 2010, 24, (2), pp.338-352
4 R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu: "Order-Preserving Encryption for Numeric Data," In SIGMOD, 2004, pp. 563–574.
5 T. Brinkhoff, "A Framework for Generating Network-Based Moving Objects," GeoInformatica, 2002, Vol.6 No.2, pp.153-180