Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 4, April 2013, pg.65 – 72

RESEARCH ARTICLE

Enhanced Network Based Query Processing in Road networks

Rathinaeswari.S.P[1], R. Parvathi[2]
[1]Department of Information Technology, PSNA College of Engineering and Technology, Dindigul, India
[2]Department of Information Technology, PSNA College of Engineering and Technology, Dindigul, India

[1] rathina.sp@gmail.com; [2] paru_rpr83@yahoo.co.in

***Abstract*— The location aware portable devices are widely increased. This rise to the usage of location based services to reveal their own location and needed location. The existing query processing techniques uses the private information of the user. It threatens the user identity, privacy and confidentiality. To overcome this, instead of spatial region with one query reference point to process the query, it uses nodes and edges to process query. Privacy does not apply due to the spatial region. In this paper, we propose a technique which processes the query to provide nearest location of the user. Network based query processing technique is used for privacy and the nearest location is obtained by incremental evaluation. Digital Signature is provided along with each query processing. It contains specific id for each processed query and user for future reference.**

Key Terms: - kNN query; Privacy; location-based services; Digital Signature; K-anonymity; anonymizer

I. Introduction

The explosive growth of user needs results location based application. The registered user for location based services can send their location information to the location based server. It indexes the location and sends the information to the user. One or more location servers are used to make the data available to the user. For example, if a user, queries for a cancer hospital then the global positioning system tracks the user location and the *location based services* (LBS) to process the queries with it.

LBS perform mobile commerce when taking the form of coupons or advertising directed at customers based on their current location. The main threat comes from sensitive information of data accessed by the user. Example of sensitive nature of queried data are location of alcoholic groups, hospitals for cancer disease and less sensitive nature of data are location of garden restaurant, shopping mall. Although Location based application assures safety, it threatens privacy and security of the users. The location based query processor assumes that the user agree to reveal their location. If a user wants to keep his private information securely then he can turn off the portable device or unregister from location server.

There are many existing anonymity techniques to provide privacy to the user information. The user can query by using the fake id but the location of the user reveals the identity. For example if a user queries for a hospital using fake id from his home ,the location in which the user query reveals the identity i.e. his resident of the home. Hence the following general approach is followed. When a user poses a query, it is send to the trusted third party anonymizer which acts as a middle layer between the user and location server. It receives the location of the user and sends the spatial region to the location server to process the query. Location server sends the candidate set of queried object to the anonymizer. The anonymizer returns result to the corresponding user. Here the Euclidean distance is considered to process the query. It does not provide the exact nearest location of the object. The *Anonymizing Spatial Region (ASR)* is computed in the anonymizer and forwarded to the location server. This contains privacy requirements. But the candidate set does not provide objects relative to the user location.

In this paper we propose a technique based on the network structure of the road to provide privacy, increase in performance and to provide location relative to the current location of the user i.e. the user location is updated whenever the user moves, hence it answers different result in different location based on the current location of the user. *Anonymizing edge list (AEL)* is computed instead of spatial region. It contains nodes and edges information of the user location. It considers network distance instead of Euclidean distance.

The rest of the paper is organized as follows. Section II specifies related work. Section III defines system architecture of our technique. Section IV elaborates the technique. Section V shows the experimental result. Section VI concludes the paper.

II. Related Works

There are many existing methods for query processing in road networks. Section A specifies the basic notation of road network. Section B specifies related query processing techniques.

A. Basic Notation of Road Network

Consider a road as a graph which consists of nodes and edges. Road network is modeled as a weighted graph which contains nodes (N) and edges (E). The graph is represented as graph G= (N, E) in Fig.1 .The network weight is assigned to each edge. Network distance defines the traveling time between two nodes. We can also calculate it based on the pixel value relative to each node. The partial weight is assigned relative to the user location and the nearest node. In case of small network shortest path algorithm Dijkstra's algorithm is used to calculate nearest location. But it is complex for large network [1].

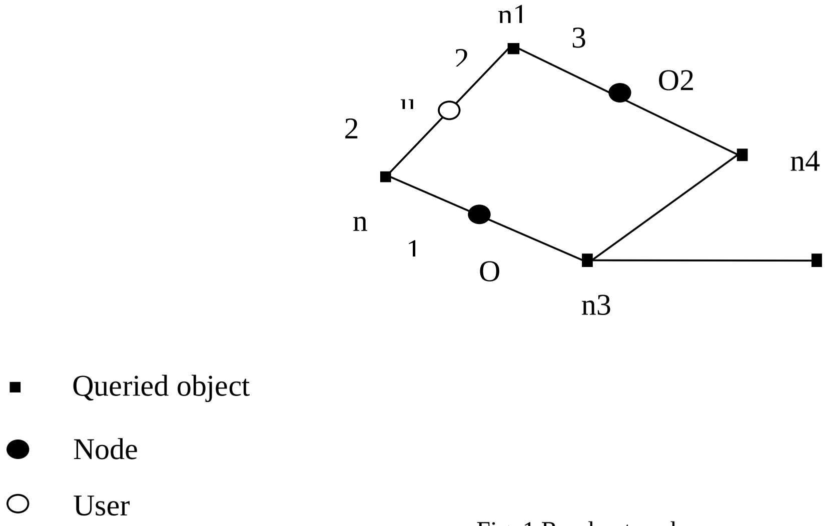


Fig. 1 Road network

The fig 1 represents the road network with nodes and edges. Spatial database stores the spatial networks in real time applications.

B. Query Processing Techniques

The location of the user should be hidden from the location services to provide privacy. This is the cloaking mechanism. The anonymizer is used to hide the actual location of the user using anonymity techniques. It can be achieved by suppression and generalization. Both these process leads to information loss by hiding the variables or replacing the variable respectively. The l-diversity is also the technique which removes the sensitive attribute or by providing the sensitive attributes. The recent technique is the K-anonymity model for data privacy. The main aim of K-anonymity is to provide records which are not distinguishable from other records. It should be anonymized if each record is indistinguishable from at least k-1 other users. Thus the probability to find the user location should be greater than 1/k. In our approach we use K-anonymity to process the query.

Most of the spatial database considers the Euclidean distance. It uses range query, k-nearest neighbor query and closest pair to process the query. These queries retrieve the needed location of the user from the location server.

In range query, the LBS receives the query range which is either an axis-parallel rectangle or a circle .In kNN queries, k-nearest-neighbor classifier is commonly based on the Euclidean distance between a test sample and the specified training samples. The following equation specifies the Euclidean distance calculation.

$$d(x_i, x_l) = \sqrt{(x_{i1} - x_{\ln\square})^2 + (x_{i2} - x_{l2})^2 + .. + (x_{ip} - x_{lp})^2} \qquad \text{… (1.1)}$$

Let x_i be an input sample with P features, $(x_{i1},x_{i2},\ldots,x_{ip})$ ^{n} be the total number of input samples (i=1,2,…n) and p the total number of features (j=1,2,..p).Euclidean distance between sample xi and xl (l=1,2,3..n) defined as in equation (1.1).

There are two expansion technique [2] used to expand the road network. The network expansion is mostly used compared to Euclidean restriction. It expands the network and discovers the objects by traversing through the nodes. Euclidean restriction is in applicable in road network because it is necessary to draw a perpendicular bisector to derive objects which cannot transferable in road network. Hence we can use network expansion technique in our model.

Casper [3] is the first work on Anonymizer implementation which constructs Anonymous Spatial Region. It provides privacy by differentiating user query data. The user query is send to the anonymizer which converts user location into spatial region.

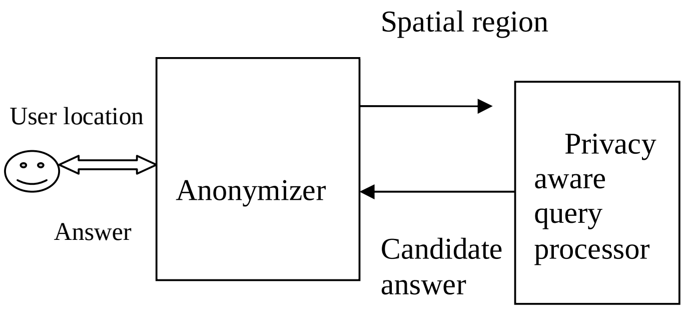


Fig.2. Casper model

The Fig 2 explains the casper model. It contains anonymizer and privacy aware query processor. The location anonymizer contains location of each user into spatial area and matches user privacy profile. According to the profile, the results are displayed. It uses Euclidean distance. The processing of kNN query is more complex on geometric properties in the Euclidean distance. Location server retrieves all objects located inside the ASR and k-nearest neighbor objects along the ASR sides. Location server eliminates duplicates and returns the candidate set. Here the user location is revealed by correlating ASRs of different users. Certain users have same anonymous region and can easily find out who posed the query. We use this model to process the query. The another framework called *Private Information Retrieval(PIR)* [4] which process the query without third party anonymizer. It is used to compute the queried object using cryptographic techniques. There are many cryptographic techniques which are theoretical. The protocol is designed that the client can retrieve data from the location server by sending the encrypted request q (i) to the server. The server responds to the client with the function of r(X, q (i)).the client computes Xi value. But the communication cost is more in this approach. This query processing is slow and in applicable to range query and nearest neighbor query. In [5][7], user forwards dummy locations in addition with his location to the location server. The obfuscation set of query points are formed. If the real location of the user is known, then the adversary eliminates the dummy location and find out the identity of user. The another approach is a location privacy method by khoshgozaran and Shahabi[6] . It uses the function which converts 2D location into 1D value. The user sends this location to the location server. It searches with this value but returns the object relative to the location. Hence it does not retrieve actual nearest object.

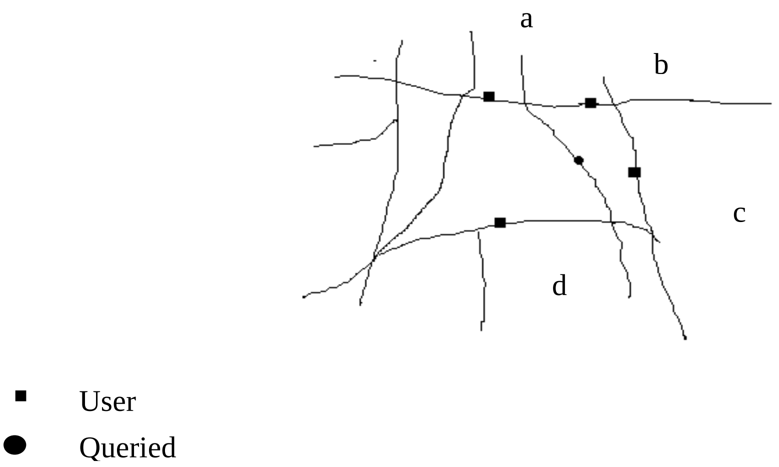


Fig.3. Road Network with Spatial Data

The framework for preventing location based identity inference [9] of users who posed spatial queries to location server. It mainly focuses on cloaking mechanism and algorithm in the location based server. Hilbert cloaking is the method to cloak the user. The two existing technique *HilbASR* and Casper are the spatial anonymity techniques which uses the Euclidean distance. The spatial region sent to the location server. The Euclidean distance does not retrieve the exact location of the object. It calculates the distance with respect to the spatial region. Consider Fig. 3 for an example to find the nearest location of a hospital.

The user q in the fig 3 queries about the nearest hospital. The locations a, b, c and d are the nearest hospital. In the existing method the Euclidean distance is calculated with respect to the user location q. it retrieves the object in the order of c, b, a and d. But the nearest location is a. To overcome this, we choose network distance by traversing the road network with respect to nodes and edges. Weights are assigned to edges with respect to the travelling time. In the Casper, users are divided as an ASR region. Consider the following Fig.4, user u1, u2, u3are divided as region.

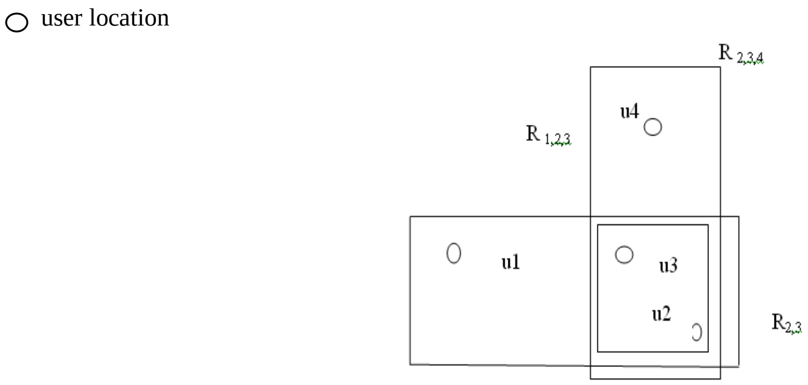


Fig.4 Spatial k- anonymity

The user u2 and u3 have the same 2-anonymous region $R_{2,3}$
whereas u_1 and u_4 have 2-anonymous regions R $_{1,2,3}$ and R $_{2,3,1}$ respectively as shown in figure 2.3. R $_{1,2,3}$ is the only ASR that contains u1.It is possible to deduce that u1 posed a query with ASR R $_{1,2,3}$. User u4 encounters the same problem when R $_{2,3,4}$ is used.

III. SYSTEM ARCHITECTURE

In this paper, we propose the anonymity technique based on the network distance. The location of the user is monitored whenever the user moves. User sends the query to anonymizer. The query is sent to the location server along with the k user's location from the anonymizer. Hence it achieves k-anonymity. The performance of the query processing is increased by storing the query details with respect to the distance. The distance between the user and nearest node is calculated. If any users within that distance rise same query it will returns the location of the queried object without computing again. Here the user details are not stored, hence it provides privacy. The edge and node information of the road network are stored in the anonymizer.

The following Fig. 4 explains the architecture diagram. Position of the user is tracked from the user query and movement of the user is updated frequently. When the user posed a query, location of the user is identified by the anonymizer with the help of selection server. Relative to the user's position, users are grouped together. Instead of spatial region here line segments or segments are considered. AEL is calculated and send to the location server. Query processing in location server results candidate set to the anonymizer and validated. Active result from the anonymizer is returned to user which contains nearest object and distance to that object from current location.

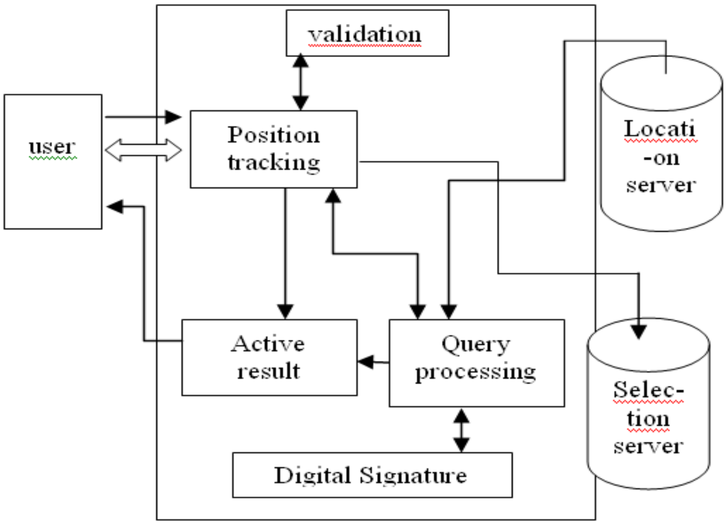


Fig.5 System architecture

The nearest node of the user is identified as in the fig.5. The location stored in the location server is searched based on the traversal based orderings. The breadth first and depth first algorithm are used to traverse. It is adaptable because each and every node is visited at least once. This makes the node to be visited multiple times. The user order is assigned based traversal. When the user moves, older user location is updated with the new location to identify the current edge.

IV. QUERY PROCESSING

The user query is processed in step by step process. Data Modeling is the first process in which the location server assigns the objects corresponding to the nodes and edges. It is the process of modeling the road network. The road is created with nodes and edges for a region. The users are made to move on the edge. The regions are divided into areas as grid. Hence the user's location is identified with the area. Whenever the user moves the location is updated. The user location is identified by using GPS system in the real time system. The location of the queried object is modeled in the road network. Anonymizer stores the nodes and edge information for identifying the line segments.

In the second step location cloaking, location of the user wherever he moves is updated in the anonymizer. The co-ordinates of the users are updated whenever the user moves. The location of the queried object is cloaked with respect to the signal received from the user GPS mobile. The user movement will be stopped for a specific location and the location of the current user position will identified. The edge location of the corresponding user location will be identified. It should be send along with other user information to the location server.

The next step is processing the query. The query given by the user is processed in the location cloaking with the knowledge of exact location of the user. The user location based on the edge information is compared with the queried object location and the nearest node. The nearest node location is identified and traversed to identify the nearest queried object. The processing of query is based on the edge and node information. It is not based on the spatial region. It results the nearest queried object location. The kNN query is used for processing. The depth first traversal is used to traverse the edge list. Single query processing is used to process the query. Single query is processed at a time and the candidate set is computed. kNN and range query are used to process the query. If a query type is kNN, it retrieves k nearest object based on the network distance. It processes the query with the node information. It retrieves any object that lies inside the AEL or on the edge away from the AEL. Another query type range query searches within the “r” distance. It also searches based on the edge information instead of spatial region.

Distance signature is used to improve the performance of the query processing. The distance between the user and the node nearest to the user are calculated. The signature is assigned to that distance. Whenever a new user comes within that distance and proposes a same query then signature is matched and returns required result. This reduces the processing time.

A. Candidate Set Computation Algorithm

Candidate set are computed based on the following algorithm.

```
/* q: query point, n: node, u: user set*/
result:=NULL
s: edge location of user, e: node list, c: candidate set
set b as array, i as zero , k as default value
Step 1: Order the user position $u_{1,}u_{2}….u_{m}$
Step 2: Identify the bucket value by dividing total number of user and k value
Step 3: Group the user in a bucket
Step 4: If the number of the user is less than bucket value go to the step 6
Step 5: Increment b value and continue step 3
Step 6: Identify the edge location corresponding to the user location
Step 7: Search in digital signature (DG) database using nodes and edges
Step 8 if DG is found, move step 14
Step 9: Perform range query or path kNN query
Step 10: For each query,check the flag value of the processed query
Step 11: Display the result without processing if the flag isset.
Step 12: Check the incident edges to find out the signature
Step 13: Calculate the distance between the user location with all nodes in the node table and the destination table
    Step 13.1: If the distance among the user and destination is less than the node distance, then assign the value to the result
    Step 13.2: Set the flag value to the processed query location
    Step 13.3: If the node distance is less than destination compared to the user location then traverse the node list and repeat step 7 till nearest destination reach.
Step 14: Sort the distance
Step 15: Remove the irrelevant objects
Step 16: compute the candidate set
Step 17: Digital Signature is set and stored in data base
Step 18: Display the nearest location of the object
```

Fig. 6 Algorithm for Candidate Set Computation

The algorithm in the fig. 6 explains how the query is sent and processed in our model. The user order is identified and grouped based on which user process the query. The user location is identified from the edge list. The user who posed the query is combined with other users. The node and edge information of the group of user is computed which is called AEL list. This AEL is sent to the location server. The distance between the user and node nearest to the user is calculated. A signature is assigned to the distance. With the node information, range query or kNN query are used to process the query. According to the AEL, the nodes are traversed. It first searches in the signature and the query. If it matches then candidate set is sent without processing the query. If it is not matched then the location server is traversed for the queried object and returns to the anonymizer. The candidate list is sorted and irrelevant objects are removed. If the queried object is not found within the range then it should return no such objects are located within the region.

V. Experimental Results

In our experiment, the road network is created as graph image with nodes and edges. The users are made to move on the road and destination locations are marked in the road network. Our algorithm is implemented in C# .Net as front end, and the node information are stored using SQL server. The mouse click event is used to stop the user movement. One of the users is stopped to raise a query. For example ,If the user wants to reach the nearest hospital for cancer, then the nearest cancer hospital location with the distance and the location area is displayed using our algorithm. Single query should be processed at a time.

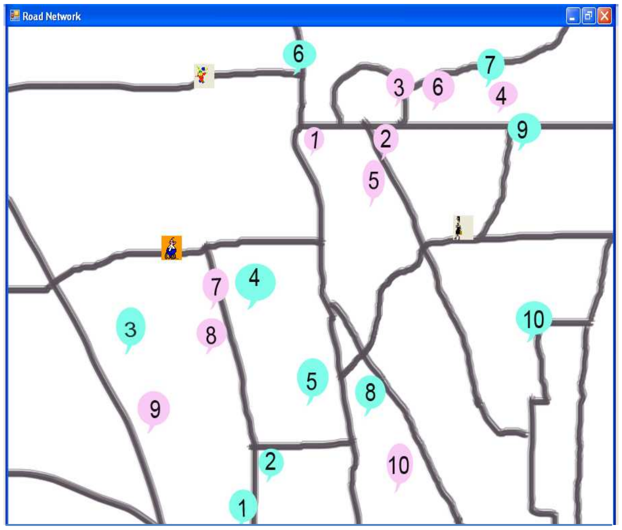
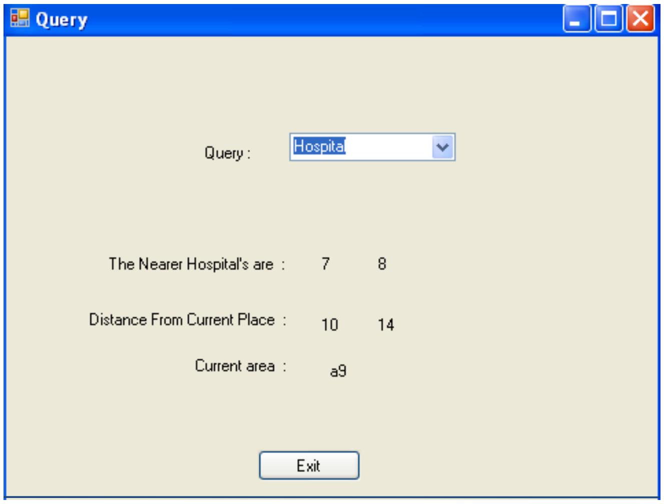


Fig. 7 Road Network

The fig 7 shows the user movement and the destination locations. The numbers and colors are used to differentiate the destination location.



Query : Hospital

The Nearer Hospital's are : 7 8

Distance From Current Place : 10 14

Current area : a9

Fig. 8 Query Result

The colored user in fig 7 raises a query for nearest hospital and the result is displayed in the fig 8. It shows the query result for the user with the distance from the current place and current area.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have studied the different anonymity technique for location privacy. The main idea is to conceal how the user identity is protected and the exact location of the nearest queried object. Our technique provides privacy by sending the edge and node information of the user who issues the query along with k other users' location to the location server. The nearest destination location of the user queried object can be provided with the additional information like directions and land marks to make the user to reach the destination easily.

REFERENCES

[1] E. W. Dijkstra, "A Note on Two Problems in Connection with Graphs," Numerische Math., vol. 1, pp. 269–271, 1959.

[2] D. Papadias, J. Zhang, N. Mamoulis, and Y.Tao, "Query Processing in Spatial Network Databases," Proc. Int'l conf. very Large Data Bases (VLDB), 2003.

[3] M. F. Mokbel, C.-Y. Chow and W. G. Aref. "The New Casper: Query Processing for Location Services without Compromising Privacy". Proc. Int'l conf. very Large Data Bases (VLDB), 2006.

[4] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan, "Private Queries in Location Based Services: Anonymizers Are Not Necessary," Proc. ACM SIGMOD, 2008.

[5] M. Duckham and L.Kulik, " A Formal Model of Obfuscation and Negotiation for Location Privacy," Proc. Int'l Conf. PervasiveComputing (PERVASIVE), 2005.

[6] A. Khoshgozaran and C. Shahabi,"Blind Evaluation of Nearest Neighbor Queries Using Space Transformation to Preserve Location Privacy," Proc. Int'l Symp. Spatial and Temporal Databases (SSTD), 2007.

[7] H. Kido, Y. Yanagisawaand T. Satoh,"An Anonymous Communication Technique using Dummies for Location-Based Services," Proc. Int'l Conf. Pervasive Services (ICPS), 2005.

[8] Kyriakos Mouratidis and Man Lung Yiu," Anonymous Query Processing in Road Networks," IEEE Trans. Knowledge and Data Eng., vol. 22,no. 1,Jan 2010.

[9] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias, "Preserving Location-Based Identity Inference in Anonymous Spatial Queries," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 12, pp. 1719-1733, Dec. 2007.

[10] T. Brinkhoff, "A Framework for Generating Network-Based Moving Objects," GeoInformatica, vol. 6, no. 2, pp.153-180, 2002.

[11] M. Gruteser and D. Grunwald, "Anonymous Usage of Location-Based Services through Spatial and Temporal Cloaking," Proc.MobiSys, 2003.

[12] Geng Zhao, Kefeng Xuan, and David Taniar "Path kNN Query Processing in Mobile Systems" VOL. 60, NO. 3, MARCH 2013