A privacy-protected k-NN query processing algorithm based on Weighted Adjacency Graph for Location-based services

Hyunjo Lee, Taehoon Kim and Jae-Woo Chang

Department of Computer Engineering Chonbuk National University Jeonju, South Korea
{o2near, taehun3718, jwchang}@jbnu.ac.kr

Abstract: Due to the advances of mobile devices with GPS (Global Positioning System), a user' privacy threat is increased in location based services (LBSs). So, in LBSs, it is important to process given queries efficiently while protecting users' privacy with a low bandwidth usage. For this, a 2PASS was proposed to search k-nearest POIs by generating cloaking regions which can hide the actual user location and reduce the bandwidth consumption. However, this method suffers from privacy attack. Therefore, we, in this paper, propose a privacy-protected k-NN query processing algorithm based on Weighted Adjacency Graph for Location-based services. Our privacy-protected k-NN query processing algorithm can reduce bandwidth usages and efficiently support k-nearest neighbor queries without revealing the private information of the query issuer. With performance evaluation, we show that our scheme outperforms the existing one.

[Hyunjo Lee, Taehoon Kim and Jae-Woo Chang. **A privacy-protected k-NN query processing algorithm based on Weighted Adjacency Graph for Location-based services.** *Life Sci J* 2014;11(7):626-631]. (ISSN:1097-8135). http://www.lifesciencesite.com. 87

Keywords: Privacy threat; location-based services (LBS); location privacy; bandwidth; k-anonymity; k nearest neighbor query processing algorithm; weighted adjacency graph

1. Introduction

Location-based services(LBSs) allow users to connect with others based on their current locations. In most cases, people use their positioning devices (i.e., iPhone, Android, Blackberry) to find out his/her location like restaurants, bars and stores that they visit. However, frequent and continuous accesses to the services expose users to privacy risk. A LBS provider might be able to collect private and delicate information about a user's choices and habits from the user's location. For example, for a k-nearest neighbor query (k-NN), a user requests to LBS about the nearest dental clinic from his/her location and based on his/her location, a LBS server returns the nearest dental clinic name to the user. For this, the LBS server can infer user health conditions that might be harmful for a user's privacy. Due to an increasing awareness of privacy risks, users might desist from accessing LBSs, which would prevent the proliferation of these services [1, 2].

Current research aligns on developing techniques to elaborate on k-anonymity [3-16] that process a user's query with protecting privacy during the access of LBSs. A k-anonymity based scheme[18] tries to blur a user's exact location by generating a cloaking region which includes a query issuer and k-1 other users. One of the recent cloaking-based approach is 2PASS (2-Phase Asynchronous Search) [19]. 2PASS is based on a notion of voronoi cells and each cell contains one object that is the nearest neighbor of any point in its cell. The user can fix the cloaked area to the voronoi cell of the nearest neighbor object, if he/she knows the voronoi cells in advance. In 2PASS, a k-NN query consists of two steps. First, the user requests the voronoi cell information corresponding to the query. Secondly, it selects objects to request. However, 2PASS suffers from privacy attack.

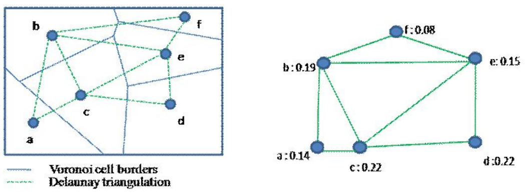
To solve the problem, in this paper, we propose a privacy-protected k-NN query processing algorithm based on weighted adjacency graph for location-based services. We follow user-cloaking-server model [3-11] where the trusted third party (location cloaker) performs location cloaking for the user. Our algorithm computes a k-NN query in three phases. First, the user requests the location cloaker (LC) and LC requests the WAG information corresponding to the query. Secondly, LC selects objects to request. Finally, LC returns the actual result to the user. We also include k-anonymity property for enhancing users' privacy while accessing LBSs.

The rest of this paper is organized as follows. In Section 2, we discuss related cloaking-based methods. We discuss our detailed system architecture and propose a privacy-protected k-NN query processing algorithm based on weighted adjacency graph in Section 3. Section 4 is devoted to experimental results. Finally, we conclude our work with future direction in Section 5.

2. Related Work

Recently, considerable research interest has focused on preventing identity inference in location-based services. The main concern is to allow the mobile user to request services without compromising his/her privacy. In cloaking-based technique, it generates blur area (circular or rectangular) that encloses an actual query issuer with other users based on his/her privacy requirement (e.g., k-anonymity, granularity metric etc). By this, the user can hide his/her location from adversary or LBS server.

One of the recent cloaking-based scheme is 2-Phase Asynchronous Search (2PASS), which was proposed by H. Hu and J. Xu [19]. For generating a cloaking region, 2PASS constructs two main parts; (i) weighted adjacency graph (WAG) that stores voronoi cell information and (ii) WAG-tree indexing for computing the objects to request from the server. Figure 1(a) shows an example of voronoi diagram with 6 objects such as a, b, c, d, e and f. The solid lines show the borders of the voronoi cells, and the dotted lines, which is called Delaunay triangulation of the space, connect the adjacent cells' objects. If the user knows the voronoi cells in advance, he/she can set the cloaked region to the voronoi cell of the nearest neighbor objects.



(a) Voronoi Diagram (b) WAG

Figure 1. Voronoi Diagram and its WAG

To access the voronoi cell information, they develop a weighted adjacency graph (WAG). WAG is a weighted undirected graph that stores the voronoi diagram and Delaunay triangulation. For example, in Figure 1(b), each vertex in this graph denotes an object, and each edge denotes a line in the Delaunay triangulation. Each vertex is also assigned a nonnegative weight. The specialty of this graph is to notify that the WAG vertices are weighted based on voronoi cell area size. User can compute out the objects to request from the server by using WAG-tree index. The criteria of object selection are a combination of the following: a) the sum of the areas of Voronoi cells from the selected objects must exceed τ; b) the genuine nearest neighbor o* must be selected; and c) these Voronoi cells must be connected, i.e., no cell is isolated from the rest of the cells.

In 2PASS, a k-NN query processing algorithm consists of two steps. First, the user requests the voronoi cell information corresponding to the query. Secondly, it selects objects to request and sends the result to the query issuer. In order to reduce computational overhead, 2PASS proposes to partition the entire WAG into WAG snippets of reasonable size so that the user receives only the snippets surrounding the query location. For example, in Figure 2, the four snippets are obtained by partitioning the space into four sub-spaces A, B, C and D of equal widths and heights and computing their WAG's, respectively. To improve the query processing performance in terms of the time, 2PASS manages this WAG snippets using WAG-tree as shown in Figure 3. In WAG-tree, each node maintains objects whose voronoi cells in the whole space overlap this sub-space. With WAG snippets and WAG-tree, 2PASS is able to save bandwidth usage compared with others by returning less number of non-result objects.

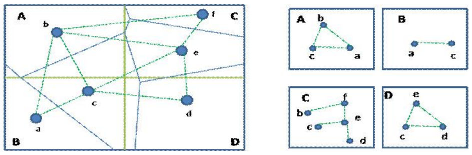


Figure 2. WAG Snippets

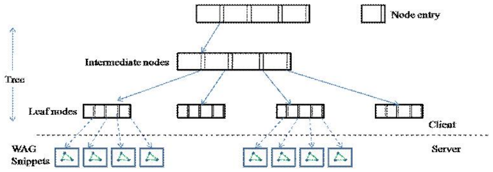


Figure 3.WAG-Tree

However, because 2PASS doesn't consider the number of users in a voronoi cell, it cannot guarantee the k-anonymity property. For example, if there is no other users except the query issuer in a cell, the location of the user can be obtained by an attacker.

3. A privacy-protected k-NN query processing algorithm based on Weighted Adjacency Graph for Location-based services

The forest herbs species in the oak and pine forests belongs to 21 families. The total number of species present in the oak forest and pine forest was 32 and 41, respectively.

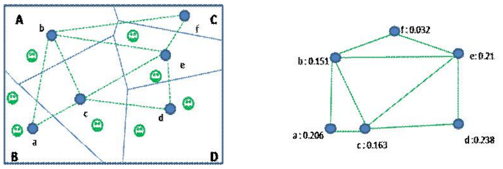
In this section, we propose a privacy-protected k-NN query processing algorithm based on Weighted Adjacency Graph for Location-based services in LBSs. We adopt trusted third party(location cloaker) which acts as a mediator between user and the server, and performs location cloaking. Also, in our scheme, we improve the WAG by setting the weight of WAG vertices based on both the size of voronoi cell area and the number of users on that cell, in order to improve the privacy of query issuer.

In our scheme, user sends a query with privacy requirement to location cloaker (LC) and LC requests the objects (including the genuine nearest neighbor (NN) together with other non result objects) based on the Voronoi cell information to satisfy the privacy requirement on the cloaked region. Our work is unique in that the LC controls what objects to request from the server so that their total number (i.e., the overall bandwidth) is minimized. To minimize the object number while still meeting the privacy threshold τ and k-anonymity requirement, the criteria of object selection are a combination of the following: (i) the sum of the areas of voronoi cells from the selected objects must exceed τ and the number of users ≥ k on that cell; (ii) the genuine nearest neighbor o* must be selected; and (iii) these voronoi cells must be connected, i.e., no cell is isolated from the rest of the cells. The last criterion guarantees that the cloaked region is a single region, which is a common assumption in all existing location cloaking approaches. Besides, the single-region assumption not only adapts to most location-based services which readily accept a single location as the input, but also alleviates some security problems.

For example, a single region is more resilient than isolated regions against background or domain knowledge attacks. With the iWAG, the object selection is equivalent to finding a sub graph that satisfies the following criteria: i) the sum of the weights of vertices in the sub graph must exceed τ and the number of user ≥ k on that cell; ii) o* must be in the sub graph; and iii) this sub graph must be a connected component. For this, we give the weight (w) for each object (V_w) based on voronoi cell area size (V_a) and number of user (U_n) on that voronoi cell. We set the priority for voronoi cell area size and the number of user in that cell. For example, if we consider the total priority, p = (α + β) = 1, then the preference of the number of user (β) is get priority than the preference of voronoi cell area size (α). Therefore, the following equation holds true.

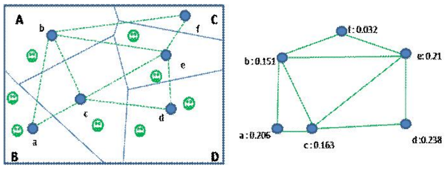
$$V_w = (V_a \times \alpha) + (\frac{U_n}{total\ U_n} \times \beta) \qquad -(1)$$

Figure 4(a) shows voronoi diagram with eight users. We calculate objects weight based on equation (1). For example, if we consider object a, then weight of a_w = 0.206. By this, we get all of objects' weight as shown in Figure 4(b).



(a) Voronoi Diagram with users (b) iWAG with users

Figure 4. Voronoi Diagram and iWAG with users



(a) Voronoi Diagram and (b) iWAG with user

Our scheme consists of three steps, i.e., iWAG generation step, iWAG-tree construction step, and k nearest neighbor query processing step.

In the iWAG generation step, first, our scheme calculates each voronoi cell area size. Then computes how many users exist on that cell. Secondly, it calculates vertex weight by using equation (1). Finally it sets all vertex weight. Actually, we consider the total vertex weight is 1. Our objective is to find out the valid weight connected component based on iWAG. For this, we follow approximate minimum valid weight connected component (MVWCC) algorithm [19]. Figure 5 shows the iWAG and its snippets.

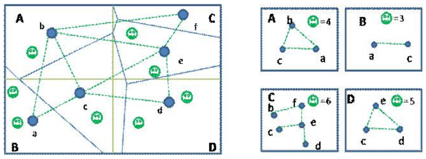


Figure 5. iWAG Snippets.

In the iWAG-tree construction step, first, our scheme checks both an area range and the number of object's (Point of Interest). Secondly, the scheme builds bounding areas (snippets) if satisfying certain conditions. Otherwise it partitions the whole space into four parts. Thirdly, it maintains objects whose voronoi cells in the whole space overlap this sub-space in order to support a k nearest neighbor (kNN) query efficiently. Finally, it recursively computes nearest neighbor of a child node until satisfying the certain criterion. Figure 6 shows the iWAG-tree and snippet pointed by it.

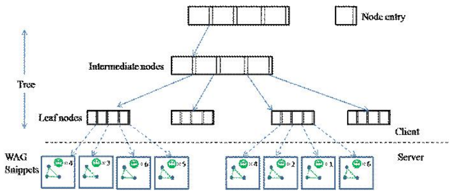


Figure 6. iWAG Tree

The k-NN query is processed as follows. First, the whole iWAG tree is sent to LC during the system initialization time. Based on the value k, the LC traverses the iWAG tree and finds out the snippet that contains the query point. Secondly, the LC matches the privacy requirements (k-anonymity, the area size (τ) etc) that are sent by the user. If the area of this snippet is still smaller than the user specified requirements, the user will locate the lowest-level child node of this snippet whose area-exceed privacy requirements. Finally, the user requests all snippets rooted at this node, called host snippets. Here, the LC adds the received host snippets into a single iWAG and calculates the minimum valid connected components by using MVWCC algorithm. In this process, LC does cloaking procedure instead of user and LC does not provide any location information or privacy requirements of user to the server.

4. Performance Analysis

In this section, we present the performance results of our location cloaker and k-NN query processing algorithm. Table 1 shows our experimental environment.

Table 1. Experimental environment

CPU	Intel® Xeon® CPU 2.00 GHz
Memory	2 GB
Simulator	Visual Studio 2010
OS	Windows XP

In our experiments, we use the real data set of Northern East America (NE) that contains 119,898 point of interest (POIs). We compare our work with the exiting approach 2PASS [19] in terms of response time and bandwidth size. The response time can be defined as how quickly the server returns the result set after receiving the query. The bandwidth can be defined as page size that encloses objects. The parameter settings are summarized in Table 2.

Table 2. Simulation parameter settings

Parameters	Range
Total User	239,668
Query Number	70,000(uniform distribution)
Granularity Threshold (τ)	0.000001, 0.00001, 0.0001, 0.001
Maximum Area of WAG Snippet	0.001
K-Anonymity	2, 4, 6, 8, 10
Average Number of User in each Cell	2

We vary the user specified threshold (τ) value from 0.000001 to 0.001 and the threshold value reflects the response time that means query processing time. As for query performance, the query processing time of 2PASS and our scheme are near to similar when τ is equivalent to small value. But, the query processing time of 2PASS is much higher than our scheme when τ becomes greater value. It is mainly due to the more number of objects it request. As a consequence, 2PASS also consumes more bandwidth than our scheme. Figure 7 demonstrates the query processing time with different τ value. Since a bigger threshold value usually contains more underlying network area, it takes longer to process. As shown in Figure 7, the increase of former metric is quite moderate until τ ≤ 0.0001. On the other hand, the latter metric linearly increases as τ grows.

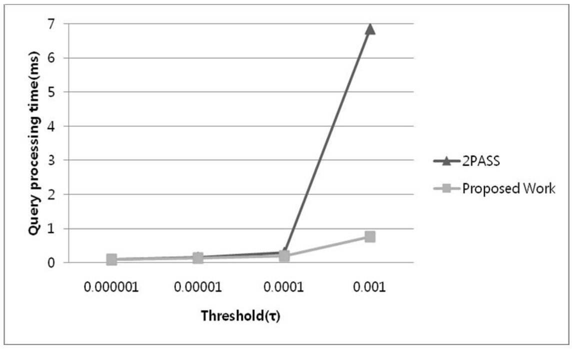


Figure 7. Query Processing Time according to τ

We measure the bandwidth size based on average number of objects returned by varying with different threshold value (τ). Figure 8 depicts the result set size with different τ. We observe that a bigger τ value generates a larger candidate result set. The reason is same that of query processing time.

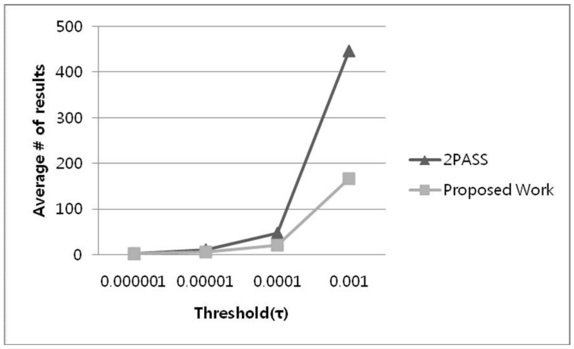


Figure 8. Average number of results according to τ

4. Discussions

The changes in topography, altitude, precipitation, temperature and soil conditions contribute to the diverse bioclimate that results in a mosaic of biotic communities at various spatial and organizational levels. Diversity represents the number of species, their relative abundance, composition, interaction among species and temporal and spatial variation in their properties. Where richness and evenness coincide, i.e., a high proportion of plant species in the vegetation are restricted, community of that area is supposed to have evolved through a long period of environmental stability.

The observation in the present study showed that the oak forest was typically moister than the pine forest which is consistent with the study of Saxena and Singh (1982). Pine forest was about 25% more diverse (40 spp.) in comparison to the oak forest (32 spp.).

Asteraceae was the dominant family in pine forest because most of the species of the family are primary successionals and have different types of growth forms. This family showed basal as well as erect forms in which basal forms emerged near the ground-level with well-developed petioles and formed a short-umbrella (Mehrotra, 1998). They can tolerate cool temperatures to high irradiances with low density of herb cover. However, erect forms are less able to capitalize on the spring window of light than any other form. This showed that the different growth forms reflect a mixed type of forest response (harsh dry to mesic). Moreover, basal forms of Violaceae showed affinity to mesic and cold conditions under the oak forest. Few species are able to tolerate the entire spectrum of environment and range throughout the gradient (Brown, 2001).

Our study showed that perennials gained dominance over annuals in oak forest as well as pine forest (Figure 1). Perennial have ability to conserve soil and with their extensive root systems of perennial grasses they also add more organic matter to the soil than annuals which can be more favourable for plant growth. Singh and Singh (1987) observed that annuals colonize and dominate the early stages of succession. Annuals to perennials species ratio are higher at primary successional site than climax stage. Species richness generally increases during secondary succession when environmental and edaphic conditions are favourable with low fluctuations.

The above results indicate that the oak forest makes climax stage for succession. The evenness and β-diversity showed similar values in sub-sites of oak as well as pine forests. The high values of beta-diversity indicate that the species composition varied from one stand to another.

Equitability/evenness varied in pine forest with respect to sub-site from 27.3 (HB) to 31.4 (HT) (Table 3). This was because of the conditional presence or absence of functional relationship of species. Comparatively higher value of equitability in pine forest with respect to oak forest indicated that the individual herb species distribution is higher. This may perhaps due to intermediate level of disturbance.

The allocation of species in the Kumaun Central Himalaya is mainly governed by moisture and temperature gradients that incorporate the effect of many physical factors. Moustafa (1990) found that the association of community types is the result of the performance of the species in response to the environmental conditions that prevail in a particular forest type. Tewari (1982) assumed that the temperature gradient is the net product of elevation and aspect; while moisture gradient is a function of slope degree, soil texture and nature of soil surface.

In addition to that, hierarchical diversity concerns taxonomic differences at other than the species level. Pielou (1975) and Magurran (1998) suggested that hierarchical (taxonomic) diversity would be higher in an area in which the species are divided amongst many genera as opposed to one in which most species belong to the same genus, and still higher as these genera are divided amongst many families as opposed to few. The families, genera and species ratio was observed maximum in the pine forest as compared to the oak forest in the present study (Table 4), indicating diverse taxonomic vegetation in the pine forest.

Acknowledgements:

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(grant number 2013010099). Also this work was supported by the Brain Korea 21 PLUS Project, National Research Foundation of Korea.

Corresponding Author:
Prof. Jae-Woo Chang
Dept. of Computer Engineering
Chonbuk National University
Jeonju, Chonbuk 561-756, South Korea
E-mail: jwchang@chonbuk.ac.kr

References

1. Privacy concerns a major roadblock for location-based services say servay. http//www.Govtech.com/gt/ article-es/104064, 2007.
2. Muntz WR, Barclay T, Dozier J, Faloutsos C, Maceachren A, Martin J, Pancake C, Satyanarayanan M. IT Roadmap to a Geospatial Future. The National Academics Press, 2003.
3. Gruteser M, Grunwald D. Anonymous usage of Location-Based Services Through Spatial and Temporal Cloaking. Proc. of the First ACM/USENIX International Conference on Mobile Systems, Application and Services (MobiSys), San fransisco, USA, 2003; 31-42.
4. Gedik B, Liu L. Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms. IEEE Trans. Mobile Computing 2008; 7(1): 1-18.
5. Schilit BN, Hong JI, Gruteser M. Wireless Location Privacy Protection. IEEE Computer 2003; 36(12): 135-137.
6. Mokbel MF, Chow CY, Aref WG. The new casper: query processing for location services

without compromising privacy. Proc. of the International Conference Very Large Database (VLDB) 2006; 763-774.

7. Mokbel MF. Towards Privacy-Aware Location Based Database Servers. Proc. of the 22nd IEEE International Conference on Data Engineering Workshop(ICDEW), Atlanta, Georgia, USA, 2006; 93.
8. Kalnis P, Ghinita G, Mouratidis K, Papadias D. Preventing Location-Based Identity Inference in Anonymous Spatial Queries. IEEE Transactions on Knowledge and Data Engineering 2007; 19(20): 1719-1733.
9. Bamba B, Liu L, Pesti P, Wang T. Supporting anonymous location queries in mobile environments with privacy grid. Proc. of the International Conference World Wide Web 2008; 237-246.
10. Wang T, Liu L. Location privacy over road networks. GIT-CC Technical Report, 2009.
11. Hossain A, Hossain AA, Chang JW. Spatial Cloaking Method Based on Reciprocity Property for Users' Privacy in Road Network. IEEE 11th International Conference on Computer and Information technology (CIT) 2011; 487-490.
12. Chow CY, Mokbel MF, Liu X. A Peer-to-Peer Spatial Cloaking Algorithm for Anonymous Location-Based Service. Proc. of Ann. ACM Int'l Symp. Advances in Geographic Information Systems (GIS), 2006; 171 - 178.
13. Ghinita G, Kalnis P, Skiadopoulos S. PRIVE: Anonymous Location-Based Queries in Distributed Mobile Systems. Proc. of The International Conference World Wide Web (WWW) 2007; 371-380.
14. Monjur M, Ahamed SI, Chowdhury SH. ELALPS: A Framework to Eliminate Location Anonymizer from Location Privacy Systems. In 33rd Annual IEEE International Computer Software and Applications Conference(COMPSAC) 2009; 11-20.
15. Hashem T, Kulik L. Safeguarding Location Privacy in Wireless Ad-hoc Networks. Ubiquitous Computing (Ubicomp) 2007; 4717: 372-390.
16. Zhong G, Hengartner U. Toward a distributed k-anonymity protocol for location privacy. Proc. of the 7th ACM workshop on Privacy in the electronic society (WPES) 2008; 33-38.
17. Ghinita G, Kalnis P, Khoshgozaran A, Shahabi C, Tan KL. Private Queries in Location Based Services: anonymizers Are Not Necessary. Proc. ACM SIGMOD 2008; 121-132.
18. Sweeney L. k-anonimity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge Based Systems 2002; 10(5): 557-570.
19. Hu H, Xu J. 2PASS: Bandwidth-Optimized Location Cloaking for Anonymous Location-Based Services. IEEE Transactions and Parallel on Distributed Systems 2010; 21(10): 1458-1472.
20. Hu H, Lee D. Range Nearest Neighbor Query. IEEE Trans. Knowledge and Data Eng. 2006; 18(1): 78-91.
21. Anisi MH, Abdullah AH, Razak SA. Efficient Data Gathering in Mobile Wireless Sensor Networks. Life Science Journal 2012; 9(4): 2152-2157.
22. Beresford A, Stajano F. Location privacy in pervasive computing. IEEE Pervasive Computing 2003; 2(1): 46-55.
23. Kido H, Yanagisawa Y, Satoh T. An anonymous communication techniques using dummies for location based services. Proc. of 2nd ICPS 2005; 88-97.
24. Berg M, Kreveld M, Overmas M. Computational Geometry: Algorithm and Applications. Springer-Verlag, 1997.
25. Hameed S, Agha MH, Choudhary MA. The Role Of Data Protection Technologies: A Case Study. Life Science Journal 2012; 9(4): 1270-1279.
26. You T, Peng W, Lee W. Protect Moving Trajectories with Dummies. Proc. of the International Conference on Mobile Data Management (MDM) 2007; 278-282.
27. Suzuki A, Iwata M, Arase Y, Hara T, Xie X, Nisho S. A user location anonymization method for location based services in a real environment. Proc. of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems(GIS) 2010; 398-401.
28. Arun E, Rajeesh J, Krshnan TR. Privacy Preserving Mobile Data Cloud With Sandboxing. Life Science Journal 2013;10(7s): 1019-1023.

5/22/2014