A Comparative Study of Spatial Cloaking Area Representations for Location Based Queries

Priti Jagwani
Dept. of Computer Science, RLA (E) College, University of Delhi

Abstract

Location-based services (LBS), provides access to the service at the point of need for the users having location-aware devices. By using LBS users are able to make queries about their surroundings anywhere and at any time. While this ubiquitous computing paradigm brings great convenience for information access, it also raises concerns about user location privacy. Privacy is an important issue for the users of location based services. Among the range of practices available to prevent location privacy spatial cloaking is most prevalent one. In spatial cloaking a user's location is blurred in a cloaking area and that area is sent to location based server along with query instead of exact location. In this paper various representations of cloaking areas (in terms of shape) are considered. Their generation techniques as well as their performance based on some factors like area, result set size, and communication cost etc is analyzed.

1. Introduction

Most of the people in the world are carrying a smart phone or a phone with location determination capability. Widespread adoption of these smart devices brought the ubiquitous computing on the finger tips of users. With this tremendous growth of Internet and mobile phones the term “Location based services” has become a buzz word now days. A Location Based Services (LBS) are information, alerts and entertainment services, accessible with the computers and mobile devices through Over the Air (OTA) network. LBS make use of the geographical position/location of the mobile device for various services. The LBS refers to the services in which the user location information is used in order to add value to the service as a whole. Since LBS is provided for users based on their exact location information, a major that about the user's location privacy has been raised. Technically speaking, Location based services are the blend of three technologies namely GIS, Internet and mobile telephony. On the basis of components used, Location based systems can be categorized as

1. Systems with trusted third party. (TTP)
2. Systems without trusted third party.

In TTP free architectures, users themselves are responsible for hiding their location and thereby maintaining their privacy. This can be achieved through collaborative methods (among users) or Private information retrieval (PIR) techniques. In TTP based architecture trust of users is simply diverted to the party existing in between the client and server. Here trusted third party is responsible for protecting privacy of user.

1.1 Threat Model

In trusted third party architecture there are basically three components in a location based system. Client, middleware (TTP) and location based server. Location based server is responsible for providing service and is considered as untrusted while middleware is the trusted component. Location is a necessary component to be supplied to location server in order to get the service but on the other hand it is required to hide the location form location base server as it is being considered as an adversary. So its very crucial to hide the location and (or) identity of user from location based server.

The request from the client is sent to middleware. Now its middleware's responsibility to remove the user's exact identity and generate the cloaking region based on user's privacy requirement. The most common form of privacy requirement is k anonymity (details will follow). So the middleware needs to generate a cloaking area covering atleast k users. Now this area along with the service request is sent to location based server (untrusted party). Location based server knows the whereabouts of cloaking area but now the exact location of user in that area. Results from locations server are sent to middleware. The whole process is shown in Figure 1.

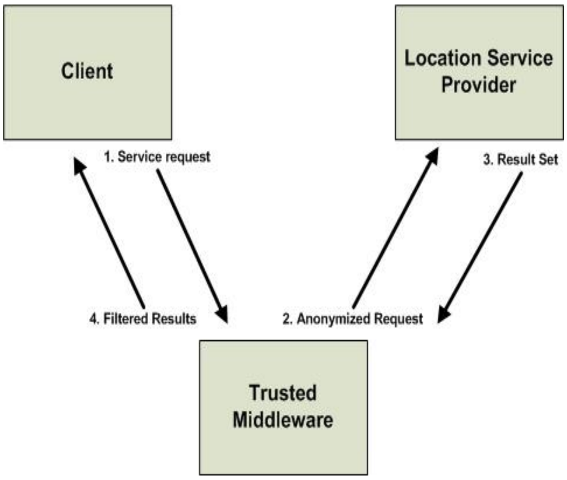

Figure 1. Transaction in LBS

As exact location of client is known to middleware, it filters out the correct results according to the true location of client and send the results to client. This approach although blurs the user location but on the other hand saves privacy to a desirable extent. Mostly location based queries are either range searches (for e.g. give me the address of restaurants with in 5 km of my current location) or nearest neighbor searches (for eg which is the nearest gas station).

K anonymity. Anonymity describes a user's anonymous state regarding a specific action. Anonymity can also be expressed as unlinkability between a user and a specific action (e.g., sending a message). K anonymity is a term first introduced by Sweeney [5] in terms of releasing database records. Later on, it has become an important metric in the domain of location privacy. It's a way to measure level of location privacy. k anonymity guarantees that the client is indistinguishable among k users in a area. While using k anonymity as a metric for location privacy it is made sure that, each of the queries of user must be indistinguishable from that of at least k - 1 other users. For this, the pseudonyms of these k users are removed from their queries, and the location in their queries is obfuscated (distorted) to the same location-area (cloaking region) which is large enough to contain the users' actual locations.

2. Various representations of cloaking regions

In this section we will be focusing on various representations of cloaking areas and the techniques to generate them. The main goal of any cloaking scheme is to represent the cloaking region in terms of shape such that the result size of the query result set is minimized and it must contain the actual result. Other than result set size there can be various other metrics to measure the performance of cloaking area representation like area of the generated region, structure (regular or irregular), generation complexity in terms of time, communication cost etc. There is an inherent tradeoff between area size and privacy. More the area of the cloaked region more is the privacy guarantee. While communication and searching cost are also being affected by regularity and size of the area . Regular shaped cloaking regions incur small network overhead (when transmitted to the Location based server) and facilitate query processing.

2.1 Circular Cloaking Regions.

Circular cloaking regions can be generated by nearest neighbor K anonymizing spatial region mechanism [1]. For a user location u, the algorithm first determines a set S of k-nearest neighbors of u, including u. From S, the algorithm selects a random user u' and forms a new set S' that includes u' and the k -1 nearest neighbors of u'. Then, another new set S'' is formed by taking a union between S and S'. Finally, the required cloaked spatial region is the bounding circle which covers all the users of S''. The mechanism of generating circular cloaking area may suffer from *center of- ASR (anonymous spatial region)* attacks in which a user can be easily located at the center of cloaked circle. The above algorithm is known as *Nearest Neighbor Cloak* (*NNC*), and is not vulnerable to *center of- ASR* attacks [3]. It has been shown in [7] that, given a privacy requirement, representing the cloaking region with a circle generally leads to a smaller result superset than using other shapes. In general, a small result superset is preferred for saving the cost of data transmission and reducing the workload of the result refinement process (especially if this process is implemented on the mobile client). The average result size given by a circular cloaking region is less than that given by a square region of the same area. Also Evaluation of circular-region-based range queries is straightforward. Comparing different shapes for a cloaking region of area A, a circle gives the smallest value for both range and K nearest neighbor searches. It is well known that a circle has the shortest perimeter under a fixed area.. Considering the regularity aspect, one may easily conclude that circle is a well defined regular shape. Because of shape regularity time taken to search the results and communication cost is lesser.

2.2. Rectangular.

Rectangular/square shaped cloaking areas are the most prevalent versions of regular structured cloaking areas. A lot of techniques are available to generate rectangular cloaking regions. The earliest one of them is adaptive interval cloaking [2] **.** For each user location update, the spatial space is recursively divided in a KD-tree-like format until a minimum k anonymous subspace is found. The general idea behind the generation of rectangular cloaking area is to enclose the locations of k-1 users in a minimum bounding box (MBB). Size of a rectangular shaped cloaking area is more as compared to the circular cloaking area for the same query. Being a well structured shape it also saves on communication cost. The average result size for a rectangular shape can be more than the circular one. Other techniques generating rectangular cloaked regions are casper, clique cloak, Hilbert K anonymizing spatial region (hilbsr) [1].

2.3. Polygonal shape

Apart from the regular shaped versions of cloaking areas like rectangle and circle there are available other nonhomogeneous irregular polygonal shapes like voronoi diagram. Given below is the account of research work which opted voronoi as their representations of cloaking area and pros and cons of their usage.

Voronoi Diagram. It is a way of dividing space into a number of regions. A set of points (called seeds, sites, or generators) is specified beforehand and for each seed there will be a corresponding region consisting of all points closer to that seed than to any other.

Lets say there exist a finite set of points {p_1, …, p_n} in the euclidean space. Each site p_k is simply a point and its corresponding Voronoi cell R_k consisting of every point whose distance to p_k is less than or equal to its distance to any other site.

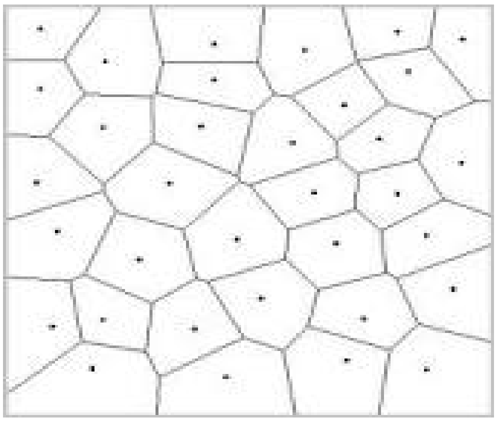

Figure 2. Voronoi Diagram

.

Kim H. [4] has shown that, Given a cloaked region including user location, finding the nearest POI to the user location cannot be achieved by range search with a fixed region. Figure 3 illustrates that the problem of range search with a fixed region. In this example, the nearest POI to the user location *u* is *p*1 but the reported nearest POI is P2. The conventional range search algorithms with a fixed region, that do not consider outer points of the region, cannot guarantee the nearest POI to user location. Therefore, outer points of a cloaked region should also be considered.

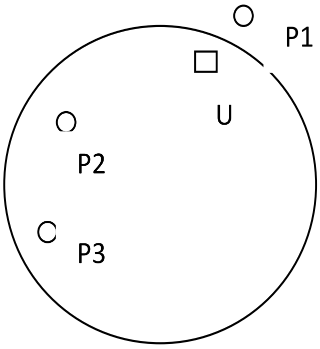

Figure 3: Conventional Range Search

To get rid of the above trouble author transformed the problem (service request) to find the optimal range. This is done by finding intersections of Voronoi cells for POIs with a cloaked region since the user position can be uniformly located at the cloaked region. But This also means that the minimum size of the candidate answer results in $\Omega(k)$ where *k* is the number of the Voronoi cells which intersect with a cloaked region.

When the locations of sites are dynamically changed or the server's storage is limited to maintain the overall Voronoi diagram, the precomputed Voronoi diagram cannot be used. Keeping in mind the objective to avoid computing the global Voronoi diagram for a large data set, which is forbiddingly costly in terms of CPU and memory. Authors in [4] design the online computation of the candidate neighbors using the local Voronoi diagram relevant to a cloaked region.

Contrary to common belief that cloaking approaches using range search incur expensive processing and communication cost, the experimental results of the above work show that the framework incurs reasonable processing and communication overhead even for large cloaked regions. By using voronoi diagrams (local/global) regularity of the shaped is although lost which results in increased communication cost but searching cost is minimized by exploiting the inherent properties of voronoi diagram.

K order voronoi diagram. The above mentioned approach uses local voronoi diagram instead of global to cut the processing cost. Authors in [6] have proposed a mechanism based on locality-sensitive hashing (LSH) to partition user locations into groups each containing at least K users (called spatial cloaks). The mechanism is shown to preserve both locality and K-anonymity. They devise an efficient algorithm to answer kNN queries for any point in the spatial cloaks of arbitrary polygonal shape. The arbitrary polygonal shape discussed above is generated in the form of k order voronoi diagram.

The order-1 VD of a set of points, also known as sites, in the plane is a tessellation that divides the plane into nonoverlapping regions called *Voronoi cells* (or cells for short), each corresponding to a site. Similarly, if the plane is divide into regions, each being the locus of points closer to a set of k sites than to the others, we have an order-k VD as shown in figure 4. Here p7 (shaded area) is the 2 order voronoi cell corresponding to sites p6 and p7. However, forming the convex cloak of k users takes O(K logK) time, which only needs to be done one time as long as the user locations do not changes[6].

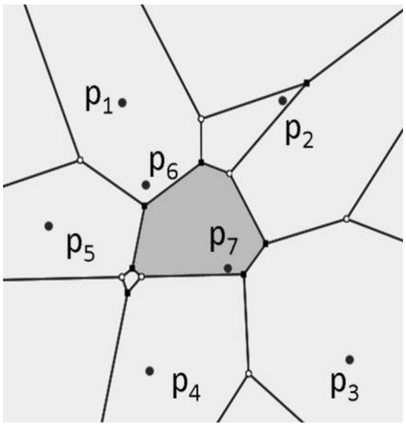

Figure 4 : Order 2 Voronoi Diagram

3. Discussions and Concluding Remarks

In range-based spatial cloaking, the most challenging issue is to minimize processing and communication overheads due to range search. There is an inherent trade-off between user privacy and service utility. A larger cloaked region implies higher guarantees for location privacy, but it also requires high computational and communication costs. Circular cloaked regions incur less communication cost but may result in higher searching costs. Although the result set is less than as compared to the rectangular region. Alternatively, the minimum bounding rectangle can be constructed in O(K) time to form the spatial cloaks. Because of regularity they also incur less communication costs.

Espousing voronoi diagram of any order is having its own advantages and disadvantages. Because of the inherent properties of voronoi diagram time taken to search result of K NN neighbor queries has reduced to many order of magnitudes as compared to the time taken in regular shapes like rectangle and circle. But communication cost (because of irregularity of shape) has increased. Also voronoi diagram incurs a time complexity of O (n log n) for generation. To avoid this problem researchers have used only local voronoi diagrams instead of the Voronoi of whole spatial space under consideration. In K order Voronoi diagram approach, the time for whole algorithm (result generation and forming of k order VD is linearly dependent on number of vertices in VD.

References

1. Chi-Yin Chow. “Spatial Cloaking Algorithms for Location privacy” In Shashi Shekhar and Hui Xiong (Eds.), Encyclopedia of Geographical Information Science, Springer, USA, ISBN 978-0-387-30858-6, 2008.
2. Gruteser M and D. Grunwald., “Anonymous usage of location-based services through spatial and temporal cloaking“. In *Proc. MobiSys ’03*, pages 31–42, 2003.
3. Kalnis, Panos, et al. "Preventing location-based identity inference in anonymous spatial queries." *Knowledge and Data Engineering, IEEE Transactions on* 19.12 (2007): 1719-1733.
4. Kim, Hyoungshick. "A spatial cloaking framework based on range search for nearest neighbor search." *Data Privacy Management and Autonomous Spontaneous Security*. Springer Berlin Heidelberg, 2010. 93-105.
5. Sweeney, L.: “k-Anonymity: A Model for Protecting Privacy”. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems 10(5), 557–570 (2002).
6. Vu, Khuong, Rong Zheng, and Jie Gao. "Efficient algorithms for K-anonymous location privacy in participatory sensing." *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012.
7. Xu, Jianliang, et al. "Privacy-conscious location-based queries in mobile environments." *Parallel and Distributed Systems, IEEE Transactions on* 21.3 (2010): 313-326.
8. Zhang, Chengyang, and Yan Huang. "Cloaking locations for anonymous location based services: a hybrid approach." *GeoInformatica* 13.2 (2009): 159-182.