# PRIVACY-PRESERVING PROCESSING IN LOCATION BASED QUERIES

**Neethu Asokan,Prof Ms.K.Sasi kala Rani(Guide)**
**Student(M.E C.S.E), Dept of Computer Science & Engg**
**Hindusthan College of Engineering & Technology, Othakkalmandapam, Pollachi Main Road**
**Coimbatore – 641 032, Tamil Nadu, India**

**Abstract:** A solution for the location-based query problems is presented. Here a major enhancement introduced into two stage approach, where the first stage is based on Oblivious Transfer and the second stage is based on Private Information Retrieval (PIR), to achieve a secure solution for both parties. Here an algorithm used in PIR, that is not better than Private Circular Query Protocol (PCQP).So Here introduce a successful privacy-preserving Location Based Services (LBS) must be secure and provide accurate query results. PCQP is deal with the privacy and the accuracy issues of privacy-preserving LBS. The protocol consists of a space filling curve and a public-key homomorphism cryptosystem. The protocol can resist correlation attack and support a multiuser scenario as long as the predescribed secret circular shift is performed before each query; in other words, the robustness of the proposed protocol is the same as that of a one-time pad encryption scheme. As a result, the security level of the proposed protocol is close to perfect secrecy without the aid of a trusted third party.

*Keywords: information retrieval, location based service, index generation.*

## INTRODUCTION

THE number of smart phones or mobile devices is rapidly increasing nowadays. Because of the popularity of mobile network and the soaring trends of cloud computing, people can enjoy the convenient life experiences offered by the mobile devices and remote servers. One of the popular services is LBS (e.g., Google Latitude), in which users can utilize the geographical information for gaining entertainment services. To incorporate Context information revealed by user mobility, we also take into account the visited physical locations of users in the data[6]. Since this information can be conveniently obtained by GPS devices, it is hence referred to as GPS locations. GPS locations play an important role in mobile web search. For example, if the user, who is searching for hotel information, is currently located in “Shinjuku, Tokyo,” his/her position can be used to personalize the search results to favor information about nearby hotels. Here, we can see that the GPS locations Help reinforcing the user’s location preferences derived from a user’s search activities to provide the most relevant results.

Our proposed framework is capable of combining a user’s GPS locations and location preferences into the personalization process. To the best of our knowledge[5], our paper is the first to propose a personalization framework that utilizes a user’s content preferences and location preferences as well as the GPS locations in personalizing search results. It was found that a significant number of queries were location queries focusing on location information[11]. In order to handle the queries that focus on location information, a number of Location-based search systems designed for location queries have been proposed. Yokoji [2] proposed a location-based search system for web documents. Location information was extracted from the web documents, which was converted into latitude-longitude pairs.

When a user submits a query together with a latitude-longitude pair, the system creates a search circle centered at the specified latitude-longitude pair and retrieves documents containing location information within the search circle. Later on, Chen et al. [8] studied the problem of efficient query processing in

location-based search systems[23]. A query is assigned with a query footprint that specifies the geographical area of interest to the user. Several algorithms are employed to rank the search results as a combination of a textual and a geographic score. The proposed secret circular shift is performed before each query and the amount of shiftiness determined only by the querying user, which can be regarded as a one-time pad encryption scheme, and therefore[14], providing high security. Servers cannot infer any knowledge about the user's location from the query history and the user's profiles, since the amount of shift has been scrambled by user and the POI information has also been encrypted.

Under such circumstance, the Correlation Attack and Background Knowledge Attack made by the server cannot succeed. For building privacy-preserving LBS, there are two major challenges: security and accuracy (in *K-NN* search)[3]. There are two major types of research works dealing with the predescribed challenges in the *K-NN* search of LBS which can be classified into 3-tier and 2-tier LBS architectures. The 3-tier architecture hides user's location with the aid of a trusted third party (TTP) [1]–[7]. There are some drawbacks when we rely the privacy-preserving LBS upon a TTP. First, in these approaches, a TTP is a must for hiding the location of user. The TTP knows too much sensitive information about the user and becomes a single point to be attacked.

## RELATED WORK

**Anonymous Usage of Location-Based Services:**

Advances in sensing and tracking technology enable location-based applications but they also create significant privacy risks. *Anonymity* can provide a high degree of privacy, save service users from dealing with service providers' privacy policies[9], and reduce the service providers' requirements for safeguarding private information. However, guaranteeing anonymous usage of location-based services requires that the precise location information transmitted by a user cannot be easily used to re-identify the subject. This paper presents middleware architecture and algorithms that can be used by a centralized location broker service[20]. The adaptive algorithms adjust the resolution of location information along spatial or temporal dimensions to meet specified anonymity constraints based on the entities that may be using location services within a given area. Using a model based on automotive traffic counts and cartographic material, we estimate the realistically expected spatial resolution for different anonymity constraints. The median resolution generated by our algorithms is 125 meters[10]. Thus, anonymous location-based requests for urban areas would have the same accuracy currently needed for E-911 services; this would provide sufficient resolution for way finding, automated bus routing services and similar location-dependent services.

**Public-Key Cryptosystems:**

This paper investigates a novel computational problem, namely the Composite Residuosity Class Problem, and its applications to public-key cryptography [2]. We propose a new trapdoor mechanism and derive from this technique three encryption schemes: a trapdoor permutation and two homomorphism probabilistic encryption schemes computationally comparable to RSA. Our cryptosystems [6], based on usual modular arithmetic's, are provably secure under appropriate assumptions in the standard model.

**Anonymous Spatial Queries:**

The increasing trend of embedding positioning capabilities (e.g., GPS) in mobile devices facilitates the widespread use of Location Based Services [1]. For such applications to succeed, privacy and confidentiality are essential. Existing privacy enhancing techniques rely on encryption to safeguard communication channels, and on pseudonyms to protect user identities. Nevertheless, the query contents may disclose the physical location of the user [7]. In this paper, we present a framework for preventing location based identity inference of users who issue spatial queries to Location Based Services [22]. We propose transformations based on the well-established *K*-anonymity concept to compute exact answers for range and nearest neighbor search, without revealing the query

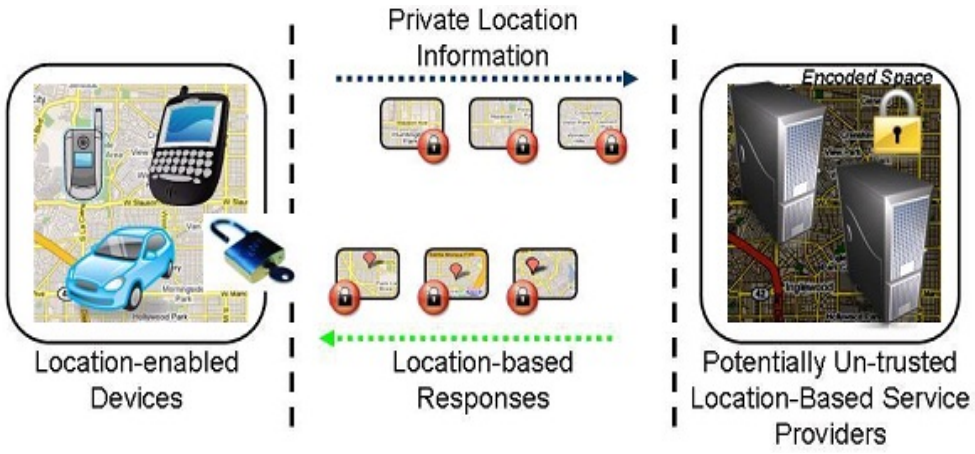


*Fig 1 System Architecture*

source. Our methods optimize the entire process of anonym zing the requests and processing the transformed spatial queries. Extensive experimental studies suggest that the proposed techniques are applicable to real-life scenarios with numerous mobile users.

## PRIVACY-PRESERVING REPUTATION TECHNIQUES

Existing works on personalization do not address the issues of privacy preservation. Location addresses this issue by controlling the amount of information in the client's user profile being exposed to the location based server using two privacy parameters, which can control privacy smoothly, while maintaining good ranking quality. Most existing location-based search systems, such as [22], require users to manually define their location preferences with latitude-longitude pairs or text form, or to manually prepare a set of location sensitive topics.

Location based profiles both of the user's content and location preferences in the ontology based user profiles, which are automatically learned from the click through and GPS data without requiring extra efforts from the user. We propose and implement a new and realistic design for location [9]. To train the user profiles quickly and efficiently, our design forwards user requests to the PMSE server to handle the training and re ranking processes.

**System Design**

In the location based client-server architecture, clients are responsible for storing the user click through and the query's derived from the location server. Simple tasks, such as updating click thoughts and query [4], creating feature vectors, and displaying reranked search results are handled by the search data clients with limited computational power. On the other hand, heavy tasks, such as RSVM training and reranking of search results, are handled by the location server [8]. Moreover, in order to minimize the data transmission between client and server, the query client would only need to submit a query together with the feature vectors to the server, and the server would automatically return a set of reranked search results according to the preferences stated in the feature vectors [19]. The data transmission cost is minimized, because only the essential data, query, feature vectors, and search results are transmitted between client and server during the personalization process. Location's client-server architecture, which meets important requirements.

Computation-intensive tasks, such as RSVM training, should be handled by the location server due to the limited computational power on mobile devices. Second, data transmission between clients

SYSTEM MODULES

**Create System model**

The system model consists of three types of entities the set of users1 who wish to access location data U, a mobile service provider SP, and a location server LS. From the point of view of a user [12], the SP and LS will compose a server, which will serve both functions. The user does not need to be concerned with the specific of the communication. The users in our model use some location-based service provided by the location server LS. For example, what is the nearest ATM or restaurant .The purpose of the mobile service provider SP is to establish and maintain the communication between the location server and the user.

The location server LS owns a set of POI records $r_i$ for $1 \leq r_i \leq \rho$. Each record describes a POI, giving GPS coordinates to its location $(x_{gps},y_{gps})$, and a description or name about what is at the location [18]. We reasonably assume that the mobile service provider SP is a passive entity and is not allowed to collude with the LS. We make this assumption because the SP can determine the whereabouts of a mobile device, which, if allowed to collude with the LS, completely subverts any method for privacy. There is simply no technological method for preventing this attack. As a consequence of this assumption, the user is able to either use GPS (Global Positioning System) or the mobile service provider to acquire his/her coordinates.

## Initialisation

A user *u* from the set of users *U* initiates the protocol process by deciding a suitable square cloaking region CR, which contains his/her location. All user queries will be with respect to this cloaking region. The user also decides on the accuracy of this cloaking region by how many cells are contained within it, whose size cannot be smaller than the minimum size defined by the location server [21]. Which is at least the minimum size defined by the server. This information is combined with the dimensions of the CR to form the public grid *P* and submitted to the location server, which partitions its records or superimposes it over pre-partitioned records.

This partition is denoted *Q* (note that the cells don't necessarily need to be the same size as the cells of *P*) [9]. Each cell in the partition *Q* must have the same number *rmax* of POI records. Any variation in this number could lead to the server identifying the user. If this constraint cannot be satisfied, then dummy records can be used to make sure each cell has the same amount of data [25]. We assume that the *LS* does not populate the private grid with misleading or incorrect data, since such action would result in the loss of business under a payment model.

## Transfer Phase:

The purpose of this protocol is for the user to obtain one and only one record from the cell in the public grid P [6]. The public grid P, known by both parties, has m columns and n rows. Each cell in P contains a symmetric key ki,j and a cell id in grid Q or (IDQi,j, ki,j), which can be represented by a stream of bits Xi,j . The user determines his/her i, j coordinates in the public grid which is used to acquire the data from the cell within the grid [20]. We remark that this key structure of this form is an enhancement from, as the client doesn't have access to the individual components of the key.

**Private Information Retrieval Phase:**

With the knowledge about which cells are contained in the private grid, and the knowledge of the key that encrypts the data in the cell, the user can initiate a private information retrieval protocol with the location server to acquire the encrypted POI data [23]. Assuming the server has initialised the integer *e*, the user *ui* and *LS* can engage in the following private information retrieval protocol using the *IDQi,j* , obtained from the execution of the previous protocol, as input. The *IDQi,j* allows the user to choose the

associated prime number power $\pi i$, which in turn allows the user to query the server.

## H-index generation:

H-Index versus H-Value: Observing the Hilbert curve one can find that the starting and the ending cells do not neighbor to each other. In this case, if the query position is near to the starting or ending cell of the curve, then the searching directions will be reduced from two to one, which is opposite to the starting or ending cell. Besides, one can also find that those H-values in DB are only used to string all the POIs together in a locality-preserving order and behave as a tool for addressing all POIs in DB [15]. As long as those H-values retain their numerical order, altering those H-values won't affect the result of query because only the order of H-values is of concern in retrieving k-NN search results. Since server is the constructor of the Moore curve and the H-table [13], it can identify the querying user's location based on the querying H-index.

### Secret circular shift:

Due to the characteristics of Moore curve, the POIs stored in H-table's first and last rows are very close to each other [2], geographically. That is, despite whatever the H-index distance between the first and the last row would be, the two POIs neighbour to each other in the 2-D space [11]. Following the same inference, the first and the last rows of H-table could be thought of as linking together just like an edge had been added to connect the two ending points of the corresponding Moore curve [18]. Let's define an entry (or a row) of H-table as the basic accessible unit; obviously, every entry (including both the first and the last one) has a neighboring relationship between its two adjacent entries. Now, if we circularly shift the POI-info column of H-table two units downward but keep the H-index column intact and then make a k-NN query at Q. In general, if we want to get the same k-NN query results after shifting the POI-info column units downward circularly, we just need to change our querying H-index, H-index (Q), to shifted querying H-index, shifted-H-index (Q), as

$$\text{shifted}-\text{H}-\text{index}(Q) = H-\text{index}(Q) + (d \times t)$$

and then send shifted-H-index (Q) to server as the new querying index [6]. Notice that, upward shifting the POI-info column is equivalent to set a negative integer to t.

## PERFORMANCE EVALUATION

After the users finished all of the five test queries in the test phase, the training phase begins. The clicked results from the test phase are treated as positive training samples Q in Location training. The click through data, the extracted content concepts, and the extracted location concepts are employed in RSVM training to obtain the personalized ranking. After the training phase, the evaluation phase is performed to decide if the personalized ranking function obtained in the training phase can indeed return more relevant results for the user. Each user was asked to provide relevance judgment on all of the top 100 results *R* for each query he/she has tested in the test phase by grading each result with one of the three levels of relevancy ("Relevant," "Fair," and "Irrelevant"). To this end, the user scans through the full-text of the results using the preview function provided by the prototype and then gives relevance ratings to all of the results returned by the search engine. Documents rated as "Relevant" are considered correct, while those rated as "Irrelevant" are considered incorrect o the user's needs. The ranking of the "Relevant" documents in R and R0 is used to compute the average relevant rank (i.e., ARR, the average rank of the relevant, for which a lower value indicates better ranking quality) and top N precisions of the baseline and personalized results.

### GPS Locations in Personalization:

In this section, we evaluate the impact of GPS locations, as defined in (14) and (16),location GPS employs only the location-based features which take into account both the location concepts and the GPS locations. The user's GPS locations and locations closely related to the GPS locations receive higher weights in the location weight vector as described in (14) and (16). Location GPS with different initial weights w GPS 0 for the decay function as described in (15). We observe that the lowest ARR is achieved when *w GPS= 0:1*. When *w GPS=0.1* increases beyond 0.1, the anking

quality degrades, because the ranking has a bias toward the GPS locations, while ignoring the location information extracted from the click through data.

**Experimental Results:**

When we compare this outcome with our previous result, we find that the protocol is still practical. For this comparison, we consider the performance of the client the most important, since we assume that a server is very powerful. Compared with the previous work, the first stage on the client side is 4-7 times faster, while in the second stage the client side is 2 times slower. We must keep in mind that the client side was implemented on a desktop machine in the previous work, and hence made the second stage slower. Also, we replaced the hash algorithm with an exponentiation operation that reduced the group space for *gRigCj* from 1024 to 160 bits. This security of this structure was protected by an outer group of 1024 bits. Because the client cannot directly access *gRigCj*, since the discrete logarithm is hard in the outer group, the client must operate in the outer group to remove the blinding factors. This contributed to faster execution in the first stage.

## CONCLUSION

In this paper gave a good performance of a privacy preserving processing in LBS by using the secret circular shift.

I expect the proposed framework not only can address the challenges of privacy preserving LBS, but also in-spire the research of secret computation with desired property to achieve privacy preserving information processing in the cloud computing era and LBS based on the GPS system, then cover all over the world.

## REFERENCE

[1] Appendix, http://www.cse.ust.hk/faculty/dlee/tkde-pmse/ appendix.pdf, 2012.
[2] Nat'l geospatial, http://earth-info.nga.mil/, 2012.
[3] svmlight, http://svmlight.joachims.org/, 2012.
[4] World gazetteer, http://www.world-gazetteer.com/, 2012.[5] E. Agichtein, E. Brill, and S. Dumais, "Improving Web Search Ranking by Incorporating User Behavior Information," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006.

[6] E. Agichtein, E. Brill, S. Dumais, and R. Ragno, "Learning User Interaction Models for Predicting Web Search Result Preferences," Proc. Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006.

[7] Y.-Y. Chen, T. Suel, and A. Markowetz, "Efficient Query Processing in Geographic Web Search Engines," Proc. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006.

[8] K.W. Church, W. Gale, P. Hanks, and D. Hindle, "Using Statistics in Lexical Analysis," Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon, Psychology Press, 1991.
[9] Q. Gan, J. Attenberg, A. Markowetz, and T. Suel, "Analysis of Geographic Queries in a Search Engine Log," Proc. First Int'l Workshop Location and the Web (LocWeb), 2008.
[10] T. Joachims, "Optimizing Search Engines Using Clickthrough Data," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 2002.
[11] C. Gentry and Z. Ramzan, "Single-database private information retrieval with constant communication rate," Proc. *Automata, Languages and Programming*, Lecture Notes in Computer Science, L. Caires, G. Italiano, L. Monteiro, C. Palamidessi, and M. Yung, Eds., 2005, vol. 3580, pp. 803 - 815.
[12] G. Ghinita, P. Kalnis, M. Kantarcioglu, and E. Bertino, "A hybrid technique for private location-based queries with database protection," Proc. *Advances in Spatial and Temporal Databases*, Lecture Notes in Computer Science, N. Mamoulis, T. Seidl, T. Pedersen, K. Torp, and I. Assent, Eds., 2009, vol. 5644, pp. 98 - 116.
[13] G. Ghinita, P. Kalnis, M. Kantarcioglu, and E. Bertino, "Approximate and exact hybrid algorithms for private nearest-neighbor

queries with database protection," *GeoInformatica*, pp. 1 - 28, 2010.
[14] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan, "Private queries in location based services: anonymizers are not necessary," Proc. *SIGMOD'08.*, 2008, pp. 121 - 132.
[15] G. Ghinita, C. R. Vicente, N. Shang, and E. Bertino, "Privacypreserving matching of spatial datasets with protection against background knowledge," Proc. *GIS '10*, 2010, pp. 3 - 12.
[16] M. Gruteser and D. Grunwald, "Anonymous usage of locationbased services through spatial and temporal cloaking," Proc. *1st international conference on Mobile systems, applications and services*, 2003, pp. 31 - 42.
[17] T. Hashem and L. Kulik, "Safeguarding location privacy in wireless ad-hoc networks," Proc. *UbiComp'07*, 2007, pp. 372 - 390.
[18] B. Hoh and M. Gruteser, "Protecting location privacy through path confusion," Proc. *SecureComm'05*, 2005, pp. 194 - 205.
[19] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias, "Preventing location-based identity inference in anonymous spatial queries," *IEEE T Knowledge and Data Engineering*, vol. 19, no. 12, pp. 1719 - 1733, 2007.
[20] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," Proc. *ICPS'05*, 2005, pp. 88 - 97.
[21] J. Krumm, "A survey of computational location privacy," *Personal and Ubiquitous Computing*, vol. 13, pp. 391 - 399, 2009.
[22] E. Kushilevitz and R. Ostrovsky, "Replication is not needed: single database, computationally-private information retrieval," Proc. *Foundations Computer Science*, 1997, pp. 364 - 373.
[23] L. Marconi, R. Pietro, B. Crispo, and M. Conti, "Time Warp: How Time Affects Privacy in LBSs," Proc. *ICICS'10*, 2010, pp. 325 - 339.
[24] S. Mascetti and C. Bettini, "A comparison of spatial generalization algorithms for lbs privacy preservation," Proc. *2007 International Conference on Mobile Data Management*, 2007, pp. 258 - 262.
[25] M. F. Mokbel, C.-Y. Chow, andW. G. Aref, "The new casper: query processing for location services without compromising privacy," Proc. *VLDB'06*, 2006, pp. 763 - 774.
[26] M. Naor and B. Pinkas, "Oblivious transfer with adaptive queries," Proc. *CRYPTO'99*, 1999, vol. 1666, pp. 791 - 791.
[27] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," Proc. *EUROCRYPT'99*, 1999, vol. 1592, pp. 223 - 238.
[28] R. Paulet, M. Golam Kaosar, X. Yi, E. Bertino, "Privacy-Preserving and Content-Protecting Location Based Queries," Proc. *ICDE'12*, 2012, pp. 44 - 53.
[29] B. Palanisamy and L. Liu, "Mobimix: Protecting location privacy with mix-zones over road networks," Proc. *ICDE'11*, 2011, pp. 494 - 505.
[30] S. Pohlig and M. Hellman, "An improved algorithm for computing logarithms over gf(p) and its cryptographic significance (corresp.)," *IEEE T Information Theory*, vol. 24, no. 1, pp. 106 - 110, 1978.