PERTURBED ANONYMIZATION: TWO LEVEL SMART PRIVACY FOR LBS MOBILE USERS

Ruchika Gupta, Udai Pratap Rao and Manish Kumar

Department of Computer Engineering, S. V. National Institute of Technology, Surat, Gujarat, India

ABSTRACT

The use of smart mobile devices like tablets, smart phones and navigational gadgets provide most promising communication and better services to mobile users. Location Based Services (LBS) have become very common in recent years. Mobile users submit their location dependent queries to the untrusted LBS server to acquire a particular service. Ideally, user's personal information such as location data is supposed to be protected while communicating to LBS and at the same time quality of service must be maintained. Therefore, there is a need to have a balanced trade-off between privacy and quality of service. To fulfil such trade-off, this paper proposes a solution that first forms the cloaking region at mobile device, perform perturbation to handle the problem of trusted third party and the anonymizer further anonymizes the location to remove the problem of enough users required to form the cloaking region. The proposed approach protects the location privacy of the user and also maintains the quality of service by selecting appropriate service to the particular user. The proposed algorithm provides two-level location protection to the user, and thus ensures smart mobility of the LBS user.

KEYWORDS

Location Privacy, Location Based Services, Perturbation, Anonymity, Point of Interest, Smart Mobility

1. INTRODUCTION

As mobile commerce is growing at a fast rate, there is a huge demand of location services by the mobile users. As a result, the user's personal information is vulnerable to the privacy breach. Almost all location services demand exact location of the user to provide accurate services in return. Location based services (LBS) continue to grow with the maturity of the positioning technology like global positioning system (GPS) [1]. LBS can be initiated when a predefined event occurs, for e.g. occurrence of an event when a user approaches or leaves the point of interest (POI). Due to the frequent exchange of private information, effective mechanisms are needed for positioning management and protection of the location data. It is agreed that users of LBS demand to have complete control over their location data due to its high vulnerability to location privacy attacks [2]. Ideally, LBS provider must provide the features so that users can manage their location information and can decide with whom and under what conditions their private information can be disclosed. The user shares her location with location server to gain personalized services in return. These services include the discovery of POIs, beforehand traffic information, route assistance and the like. The shared private information can be misused if heard by an adversary. For instance, a user requests a list of highly specialized medical care providers informing about her medical condition to the service provider who may indirectly reveals user's

David C. Wyld et al. (Eds) : AIAPP, CRIS, CoSIT, SIGL, SEC - 2018
pp. 69–79, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.80406

medical condition and may misuse them later for some personal gains. The server processes the user's query and returns a set of POIs back. Sometimes, there can be the case when an LBS server promotes a business and in return of few candidate results, it also offers coupons to the user against multiple queries sent as shown in Fig 1. The main drawback of such approach is that the server returns too much unnecessary POIs that leak our important information of the database.

Fig 1: Query Processing Example

Trusted Third Party (TTP) based LBS solution has a major disadvantage of single point of failure and can also be viewed as a single promising point of attack by an adversary. Privacy is at stake if an adversary somehow takes control of the third party. On the other hand, mobile based mechanisms face the problem of enough users to form cloaking region and introduces unnecessary delay till the required number of users are found in the region. This paper proposes a two-level privacy preserving approach that effectively deals with the mentioned issues of existing techniques and provides protected services to the user with optimum level of quality of service. We suggest that before sending the exact location, perturbation is applied to the nearby cluster region at client side in order to handle the trust issues of third party. In case of single point of failure i.e. when anonymizer fails; LBS server would not be able to identify the exact location of the user due to the perturbation applied at mobile device. The proposed algorithm provides two-level location privacy protection to the user, and thus ensures smart mobility of the LBS user so that the user can freely move anywhere without any privacy breach apprehensions.

2. LITERATURE REVIEW

To handle the privacy concerns, two types of research exist in the literature. First, location cloaking based methods [3-8] and second, Private Information Retrieval (PIR) based techniques [9]. The information sent by the user (be it original or modified) should be under the control of the user who sent it [10]. Passive and Active are two different threats to the user privacy [10]. Schiller et al. [11] suggests the common architecture model of LBS with three different layers namely; positioning layer, a middleware layer and an application layer. Each layer has a dedicated responsibility in overall execution of the service.

Numerous techniques are discussed to make communication with server. One of the techniques by Dewri et al. [12] proposes a location based query in the presence of privacy supportive LBS provider. In this scheme, the user sends the query to the LBS server, even though the user uses her geographic location in a generalized way. Authors in [7] have categorized the then existing privacy preservation techniques in a hierarchical manner. Author [13] discussed the use of TTP (often called as anonymizer) as an intermediate entity that plays a key role in protecting the user's identity. Anonymizer's main aim is to hide the users' true real world identity by omitting (or modifying) the location information [14]. In the policy based scheme, the user sets the set of policies which is supposed to be followed by the service providers [13]. Due to dishonest behaviour of third party the methods that do not rely on trusted third parties are proposed [15-18]. In collaboration based method, the user do not discloses the exact location while sending a query.

The user makes changes in the location and broadcast it to her neighbours. In return, neighbours also send their modified location and centroid is calculated at user's end. This centroid value along with the user query is then sent to the LBS provider [15]. Obfuscation based methods is the process of degrading the information quality of the user's location. By using the imprecision method of obfuscation, one can easily degrade the information quality. In this method, the location space is modelled as a graph where vertices represent locations. The user sends the set of vertices instead of sending the single vertex of her own [16]. In Personal Information Retrieval (PIR) based method, the service provider cooperates with the user by following the PIR protocol, where the LBS provider answers the queries containing the location information [17]. Another relatively new approach called Privacy Enhancing Technologies (PETs) [18] restricts anonymity issues based on Trusted Computing Technologies results in a better privacy of user's personal information. Privacy Enhanced Trusted location based services (PE- TLBS) [18] focus to implement a simple protocol in which the user authenticate the server, while preserving anonymity and avoiding the possibility of their personal information leakage. The concept of dummy nodes proposes the use of dummy locations with the real location to protect the location privacy of the node [19]. The quality of requested service degrades when the number of dummy node increases.

Cloaking based approach [8][13][21] works well in protecting the user's location but vulnerable to untrusted third party i.e. Anonymizer [14], which cloaks the user's location and anonymize before sending to the service provider. Cloaking is the technique to blur the location of the node by including k-1 more nodes from the same location besides the target node [20]. In the concept of k-anonymity, locations of k users are cloaked together and all nodes in the cloak act as one of the possible sender of the query. Therefore, it becomes difficult for the third party to identify the actual user [21]. Research shows that constructing the cloak of user location does not ensure the absolute user privacy [22], however, [23] [24] propose another alternatives using TTP based and TTP free architectures.

3. PROBLEM DESCRIPTION

While cloaking the location, there is a problem of the number of users needed to form the cloak region and there may be circumstances when only single user is available for cloaking, which is surely not suitable to form the cloak. Our goal is to protect the user's private location information from the untrusted third party (or anonymizer) and at the same time mitigating the problem of minimum number of users by using the anonymizer that anonymize the location information before sending to location service provider.

3.1. Problem Formalization

We focus on the privacy of the user in a two-dimensional location based services where a trusted third party cloaks the query; Q: < location, query string >. The mechanism is protected to a great extent in the setup of dishonest third party with the problem of enough users to compute the cloak region.

4. PROPOSED ALGORITHM

4.1. Algorithm Design

In the proposed framework, cloaking region is used where the global cloaking region is split into local or sub cloak regions using clustering approach as shown in Fig 2.

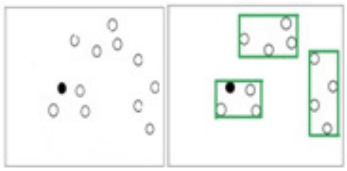

Fig 2: Cloak Region Formation: a. Global Cloak region, b. Local cloak region

The global cloak region with k users is shown in Fig 2a. The global cloak region splits into sub cloak regions as shown in Fig 2b. Dividing the global cloak region, say for k=12 users into local cloak regions for n=3; each sub local cloak region contains k'=k/n=4 users. Perturbation process is now applied to all regions before sending to the anonymizer that further anonymize the location information. Fig 3 presents the flow of proposed mechanism.

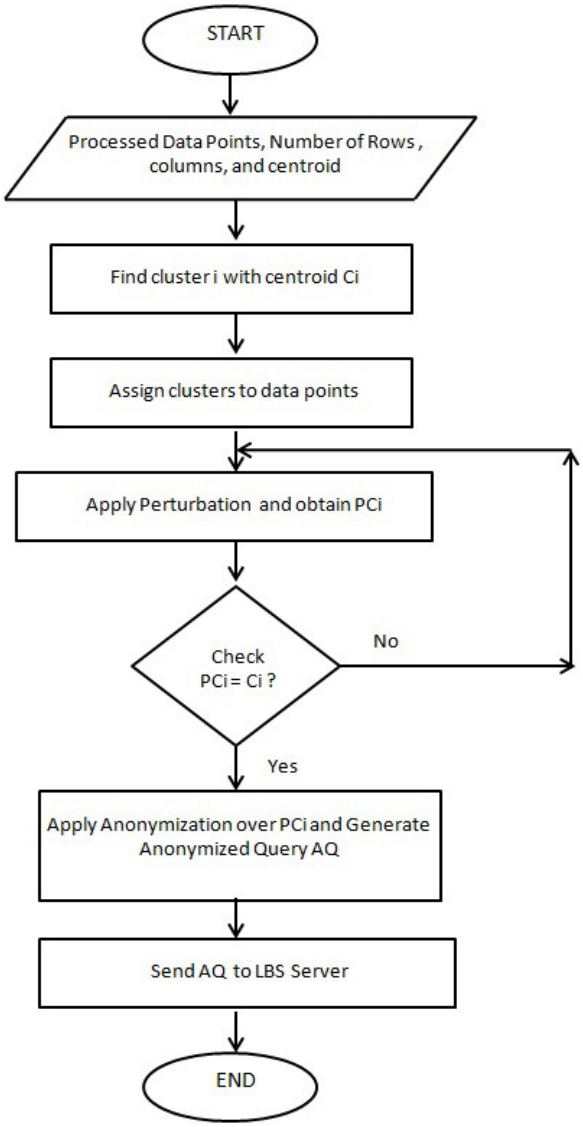

Fig 3: Flow of Proposed Mechanism

Algorithm 1:

Terminologies used:

- *NUMIT, BIGNUM* = Number of Iterations, Random Big number for minimum distance
- Mean, distort= to store coordinates sum, distorted distance
- *numRows, numCols* = Number of Rows in Input file, Number of Columns in Input file
- *numCent, newCent* = Number of Centroid in Centroid file, new centroid after Iterations
- *pCent*= Centroid after perturbation process

1. Process Input File and Centroid File
2. for(j=0 to *numRows*)
   for(k=o to *numCols*)
   mean[k]+=x[j][k] // ∑ x and ∑ y coordinates
3. for(k=0 to *numCols*)
   mean[k]/=*numRows* //calculate mean value
4. for (it=0 to *NUMIT*)
   set distort=0 ,count=0
   for(j=0 to *numRows*)
   set rmin=*BIGNUM*
   for(k=0 to *numCent*) set dist[j][k]=0
   for(e=0 to *numCols*)
   Calculate squared Euclidean distance
   END LOOP
   find Minimum Distance
   END LOOP
   distort += Minimum Distance, count++
   END LOOP
   Re-estimate new centroid points, *newCent*
   END LOOP
   Assign number of input points to respective cloak
   END LOOP
5. Call Perturbation(*newCent*);
6. Call Anonymization(*pCent*);
7. END

Algorithm 1 shows the overall functioning of the proposed algorithm. Perturbation process (step 5 of the Algorithm 1) and anonymization process (step 6 of the Algorithm 1) are described through Algorithm 2 and Algorithm 3 respectively.

Algorithm 2: Perturbation

Terminologies used:

- *numCent* = Number of centroid
- *pCent* = Perturbed centroid
- *rand( )* = Random Function

1. for(j=0 to *numCent*)
   Generate random centroid
   // *rand()%numCent*
   END LOOP
2. Assign perturbed centroid to original centroid
   pCent = *Cj;* where j Ɛ [1,j] and j is random
3. END

Algorithm 3: Anonymization

Terminologies used:

- *numRows* = Number of rows
- *pCent* = Perturbed centroid
- *temp* = holds temporary random value

1. for(j=0 to *numRows*)
   a. Generate random value and assigned to temp
   b. Anonymized point = *pCent* + temp
   END LOOP
2. Assign anonymized points to the original points
3. END

5. EXAMPLE

We have input file with coordinates {(0,2),(17,39),(10,57),(4,49),(82,31)} represented as P1, P2, P3, P4, and P5 respectively and a centroid file with coordinates {(90,10), (70,30), (50,50)} taken as C1, C2, C3 respectively. Let *numRows* be the number of rows in input file, *numCols* be the number of columns in input file and *numCent* be the number of centroid in centroid file. Here, *numRows*=5, *numCols*=2, and *numCent*=3.

Calculating mean of the given input points gives;

$\sum x = 0+17+10+4+82=113$, $\sum y = 2+39+57+49+31=178$

Therefore, mean coordinates are ($\sum x$ / numRows, $\sum y$ / numCols) i.e. (22.600000, 35.599998). Now, calculate distance matrix of each input points from each centroid will be given as,

Table 1: Distance Matrix between Centroid and Input Points

Input Points \ Centroid	C1	C2	C3
P1	8164	5684	4804
P2	6170	2890	1210
P3	8609	4329	1649
P4	8917	4717	2117
P5	505	145	1385

Dist [0][0] = $(0-90)^2 + (2-10)^2 = 8164$ {Squared Euclidean Distance}, Dist [0][1] = $(0-70)^2 + (2-30)^2 = 5684$, Dist [0][2] = $(0-50)^2 + (2-50)^2 = 4804$. Similarly, we find other distances from each input points forming the distance matrix as shown in Table 1. Now, assign input points to the nearest cluster with new centroids as; C1= (22.60, 35.60), C2= (82.00, 31.00), and C3= (7.75, 36.75). Distinctly, C2 is assigned for {(82,31)} and C3 is assigned for {(0,2), (17,39), (10,57), (4,49)}. After assigning clusters to each data point's perturbation is applied on each cluster centroid. The new perturbed centroid C2 is for input point {(0,2), (17,39), (10,57), (4,49)} and C3 is for {(82,31)}. Now, the perturbed centroid is further anonymized which resulted into the anonymized value used to contact to LBS server.

6. EMPIRICAL EVALUATION

6.1. Experimental Setup and Scenario

We implemented the algorithm on Dev-C++ version 4.9.9.2 on Intel core 2 duo 2.2 GHz machine with 4GB of RAM. We consider two entities; *McDonald's* and *Library* at five different locations. The selection of appropriate entity is based on the minimum distance between a particular Point of Interests (POIs) and entities. However, the result accuracy suffers as the number of user increases.

6.2. Results

The results of the proposed approach based on the number of iterations and number of POIs. The total computation time with respect to number of iterations and number of POIs are shown in Table 2 and Table 3 respectively.

Table 2: Time v/s Iterations, **Table 3:** Time v/s POIs

NUMIT (Number of Iterations)	Computation Time in ms
1	4.836
2	5.417
3	5.569
4	7.888
5	8.863

Number of POIs	Computation Time in ms
5	9.054
10	10.344
15	11.596
20	11.625
25	12.156
30	13.701

The above results show that as we increase the iterations, computation time of our approach increases. However, sometimes it depends on the number of processes running on the machine and may be different for different machines. It is observed that as number of POIs increases, the computation time also increases. Table 4 shows a brief comparison of our proposed approach with other existing approaches.

Table 4: Comparison with other Existing Approaches

Properties / Approaches	Accuracy	Single Point of Failure	Problem of enough users	Privacy
Mobile Device Based	✓	×	✓	✓
TTP Based	✓	✓	×	✓
Our Approach	✓	×	×	✓

Fig 4 and Fig 5 shows the computational graph of proposed mechanism with respect to Number of Iterations and Number of POIs respectively. We now present how accurate the service is provided to the requestor based on the same nearest service entity selection.

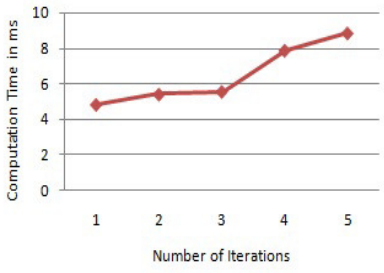

Fig 4: Time vs. Iterations

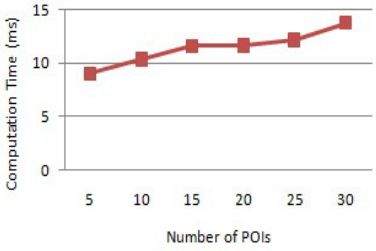

Fig 5: Time vs. POIs

Let's consider, two requested entities are McDonald's and Library with coordinates {(352.34, 534.3), (131,179.5), (192,245), (240,870), (132,564)} and {(625,387), (952,133), (287,235), (152,120), (367,755) respectively.

Now consider the input coordinates of five requesting entities i.e. P1, P2, P3, P4, and P5 according to the range in degrees as given below;

For 0-100: {(10, 12), (70, 98), (22, 87), (44, 49), (82, 31)}
For 100-200: {(110, 112), (170, 198), (122, 187), (144, 149), (182, 131)}
For 200-300: {(210, 212), (270, 298), (222, 287), (244, 249), (282, 231)}
For 300-400: {(310, 312), (370, 398), (322, 387), (344, 349), (382, 331)}
For 400-500: {(410, 412), (470, 498), (422, 487), (444, 449), (482, 431)}

Table 5: Nearest Entity Selection Table

Range	Co-ordinate Selection	
	McDonald's	Library
0-100	(131, 179.5)	(152,120)
100-200	(131, 179.5)	(152,120)
200-300	(192, 245)	(287, 235)
300-400	(352.34, 534.30)	(287, 235)
400-500	(352.34, 534.30)	(625, 387)

We also check if the above points assigned to the nearest service entity for both McDonald's and Library. Table 5 shows that which nearest service entity is selected for a particular input coordinate. For instance, P3 is assigned to both McDonald's and Library. Table 5 also shows how the same nearest service entity assigned to a particular requesting entity.

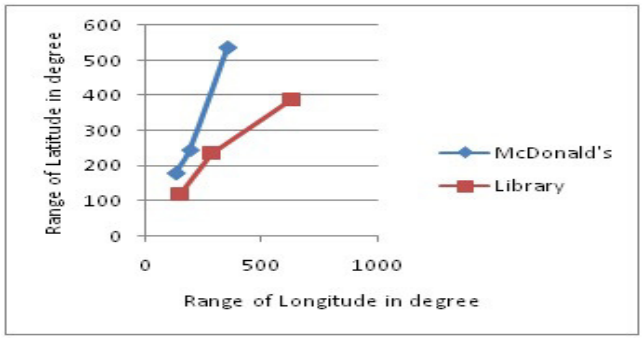

Fig 6: Nearest Entity Selection Comparison

The graph in Fig 6 shows that our approach selects same entity (McDonald's and Library) as it is selected by the original input coordinates which favours the accuracy of our approach and also provides better quality of services by selecting nearest entity in the range.

7. ANALYSIS

7.1. Privacy

Our approach is free from the issues of single point of attack and enough number of users as first problem is removed by sending the perturbed location to the trusted third party and second problem is handled by using the anonymizer which further anonymizes the location information by adding random users.

7.2. QoS

It depends on the number of users forming the cloak region. Accuracy degrades as the number of users increase i.e. k users forming the cloak achieve relatively more accurate service as compared to *k+1* users.

7.3. Communication Cost

If K is the number of clusters or cloak regions, N represents number of users in each cluster, and M is the message size then,

- Communication rounds = 4(K+1) //in general and 2(K) //if anonymizer fails
- Message size = (K+M)
- Communication Message = 4(K(M)) //in general and 2(K(M)) //if anonymizer fails

7.3. Computation Cost

In our proposed approach, time complexity can be described as *O(i*m*c*n)*. For perturbation and anonymization processes, the complexity is *O(c)* and *O(m)* respectively. Therefore, overall complexity of the algorithm is given as *O(i*m*c*n).* Where m, n represents number of rows and columns in the input file of location coordinates, i represent number of iterations, and c is the number of centroid in centroid file.

8. CONCLUSIONS

Cloaking based approaches are successful in protecting the privacy of users to some extent but there is a trade-off between the privacy and retrieved information accuracy while accessing a particular service. In order to improve the trade-off, we proposed a two-level smart privacy cloaking mechanism in which a global cloak region is divided into local regions at mobile device. After location perturbation, the location is sent to the trusted third party that brings the removal of the problem of single promising point of attack. Now, the anonymizer further anonymizes the perturbed location to handle the problem of enough users required to form the cloaking region. The proposed algorithm provides two-level location protection to the user, and thus ensures smart mobility of the LBS user so that the user can freely move anywhere without any privacy breach apprehensions. The mechanism also works well to satisfy service accuracy need of the user. Our approach is applicable for two dimensional regions and can be extended further to three dimensions; hence can provide better accuracy by covering all the scattered point of interests along with z direction.

REFERENCES

[1] B. Hofmann-Wellenhof, H. Lichtenegger, and E. Wasle, GNSS–global navigation satellite systems: GPS, GLONASS, Galileo, and more. Springer Science & Business Media, 2007.

[2] Wernke, Marius, et al. "A classification of location privacy attacks and approaches." Personal and Ubiquitous Computing 18.1 (2014): 163-175.

[3] G. Ghinita, P. Kalnis and S. Skiadopoulos, “PRIVE: Anonymous Location-Based Queries in Distributed Mobile Systems,” in Proceeding of 16th international conference on World Wide Web, pp. 371-380, 2007.

[4] G. Ghinita, P. Kalnis and S. Skiadopoulos, “MobiHide: A Mobilea Peer-to-Peer System for Anonymous Location-Based Queries,” in Proceeding of 10th International Symposium on Spatial and Temporal Databases, pp. 221-238, 2007.

[5] P. Kalnis, G. Ghinita, K. Mouratidis and D. Papadias, “Preventing Location-Based Identity Inference in Anonymous Spatial Queries,” in Proceeding of Transactions on Knowledge and Data Engineering, pp. 1719-1733, 2007.

[6] W. Ku, Y. Chen and R. Zimmermann, “Privacy Protected Spatial Query Processing for Advanced LBS,”Wireless Personal Communications 2009 Volume 51, no. 1, 2009.

[7] M. Mokbel, C. Chow and W. Aref, “The New Casper:Query Processing for Location Services without Compromising Privacy,” in Proceeding of the International Conference on Very Large Data Bases, pp. 763–774, 2006.

[8] C. Y. Chow, M. F. Mokbel, and X. Liu. A, "Peer-to-Peer Spatial Cloaking Algorithm for Anonymous Location-based Services," in Proceeding of the ACM International Symposium on Advances in Geographic Information Systems, pp. 171–178, 2006.

[9] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi and K. L. Tan, "Private Queries in Location Based Services: Anonymizers are notNecessary," in Proceeding of ACM SIGMOD international conference onManagement of data, pp. 121-132, 2008.

[10] F. Xu, J. He, M. Wright, J. Xu, "Privacy Protection in Location-Sharing Services,"in Proceeding of International Conference on Computer Application and System Modeling, ICCASM, pp. 488-491, 2010

[11] J. H. Schiller and A. Voisard, "Location-based Services: Protocol layers model," in IEEE transactions on mobile computing, pp. 29-31, 2004.

[12] R. Dewri and R. Thirumella, " Exploiting service similarity for privacy in Location based search queries," in IEEE transactions on parallel and distributed systems, vol.25, no. 2, pp. 374-383, February 2014.

[13] A. Solanas, J. Domingo-Ferrer and A. Martnez-Ballest, "Location privacy in location-Based services: Beyond TTP-based schemes," in Proceedings of the 1st International Workshop on Privacy in Location-Based Applications (PILBA), pp.12-23, 2008.

[14] L. Sweeney, "Achieving k-anonymity privacy protection using generalization and suppression, "International Journal of Uncertainty, Fuzziness and Knowledge Based Systems, pp. 571-588, 2002.

[15] J. Domingo-Ferrer, "Microaggregation for Database and Location Privacy," in Next Generation Information Technologies and Systems, Vol. 4032, pp.106-116, 2006.

[16] M. Duckham, L. Kulik, "Location Privacy and Location-Aware Computing," in Dynamic and Mobile GIS: Investigating Changes in Space and Time, CRC Press, pp. 35–52, 2007.

[17] G. Ghinita, P. Kalnis et al., "Private queries in location based services: Anonymizers are not necessary," in Proceedingof the 2008 ACM SIGMOD international conference on Management of data, pp.121-132, 2008.

[18] S. Yan, T. F. La Porta, and P. Kermani, "A Flexible Privacy-Enhanced Location-Based Services System Framework and Practice," in Mobile Computing, IEEE Transactions on, vol. 8, no. 3, pp. 304-321, 2009.

[19] R. Kato et al., "A Dummy based anonymization method based on user trajectory with pauses," in Proceeding of 20th international conference on advances in geographic information system, pp.249-258, 2012.

[20] J M.Mano, Y. Ishikawa, "Anonymizing user location and profile information for privacy aware mobile service," in Proceeding of2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, pp. 68-75, 2010.

[21] M. Gruteser and D. Grunwald, " Anonymous usage of location based services through spatial and temporal cloaking," in Proceeding of the 1st international conference in Mobile systems, applications and services, pp. 31-42, 2003.

[22] R. Gupta and U. P. Rao, "An exploration to location based service and its privacy preserving techniques: A survey," Wireles Personal Communications, vol. 96, issue 2, pp.1973–2007, 2017.

[23] R. Gupta and U. P. Rao, "Achieving location privacy through CAST in location based services," Journal of Communications and Networks, vol. 19, no. 3, pp. 239–249, 2017.

[24] R. Gupta and U. P. Rao, "A hybrid location privacy solution for mobile LBS," Mobile Information Systems, vol. 2017, pp. 1–11, 2017.

AUTHORS

Ruchika Gupta is a Ph.D. research scholar in Computer Engineering Department, National Institute of Technology, Surat, India. Her research interests include Information Privacy, Data Security, Mobile Computing, Peer to Peer communication, and Location Privacy.

Dr. Udai Pratap Rao is currently an Assistant Professor in Computer Engineering Department at S. V. National Institute of Technology, Surat, Gujarat, INDIA. He obtained his Ph.D. degree in Computer Engineering in 2014. His research interests include Information Security and Privacy, Location Based Privacy, and Big Data Analytics.

Mr. Manish Kumar is an alumni of Computer Engineering Department, National Institute of Technology, Surat, India. His research interests include Security and Privacy in LBS.