COP 5711: Parallel and Distributed Database systems

Project Report: Privacy issues in Location Based Services.

By: Anmol Sureshkumar Panchal; UID: 4446829 Fall 2017

The research problem

There are different kinds of Application on mobile platforms for user which needs access to their location to function. They install and use those application and allow access to their location to be updated at location servers when applications demand to do so. This generates the privacy concerns for the users as most of the applications have untrusted location servers and their sensitive information involved in the submitted queries may be released to third parties and their privacy can get exploited.

LBSs poses a serious threat to users' privacy. By collecting the location information embedded in the LBS queries, an adversary who has compromised the LBS server can infer sensitive privacy information about service recipients. To address these privacy issues, many approaches have been proposed over the past decade based on two categories 1- Location Privacy and 2- Query Privacy. Location privacy and query privacy, which are usually preserved through policy based solutions, cryptography primitive-based approaches, location perturbation and obfuscation techniques & spatial and temporal cloaking techniques. Most of the approaches protect either location privacy, query privacy or both, by employing trusted third-party server, such as centralized location anonymizer. Their privacy degrees can be quantified by some well-used privacy metrics such as k-anonymity and l-diversity. However, employing trusted third-party server may cause serious privacy problems such as single point of failure.

Solutions

Method:1- Time Obfuscation-Based Privacy-Preserving Scheme.

- They have proposed a design based on time obfuscation-based scheme for privacy aware mobile users in LBSs, termed TOP-privacy. Its, different from existing approaches, TOP-privacy generates and sends some carefully generated dummy queries at leisure time, and sends the real query when LBS is needed.
- They designed a dummy query generation algorithm, which contains two modules, dummy location selection module and classified POI pool construction module. The former module aims to generate solid dummy locations, which cannot be easily filtered out by the adversaries with some side information. The latter module guarantees the effectiveness of the generated dummy point of interests to user's query privacy.
- They present a WIFI access point based solution to implement our TOP-privacy. Analytical and simulation results indicate the effectiveness and efficiency of our scheme.
- Dummy location should be determined for each dummy query. Specifically, based on the user's current location li and the defined location offset d, the mobile user constructs a virtual circle with the center li of the radius d, which limits the maximum movement of the mobile user, and finds out a set of candidate locations within this virtual circle by comparing the location distribution of the real user and the other cells.
- Any locations have similar location distributions can be put into the candidate locations set. Otherwise, the selected dummy locations might be easily filtered out by the adversaries with background information. Note that, the location offset d cannot be too big since the user may submit his own query at any dummy locations in the following timestamps.

- Let’s see the example in Fig. 1, if the dummy location d3 is very far away (out of the virtual circle) from l3, then, it is hard for the user to move to location l4 within the t4 − t3, thus the real query will be easily distinguished.

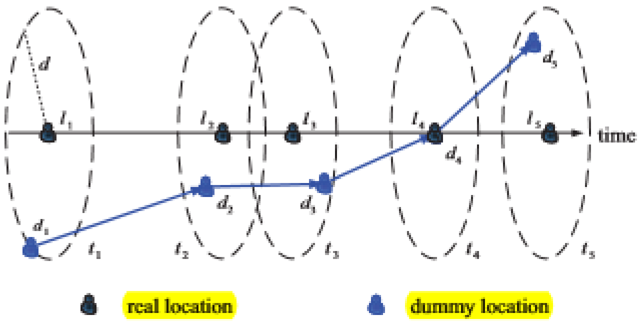


Fig. 2. Our basic idea

- When the dummy query is needed, the mobile user can easily find the corresponding query distribution level based on his real poi, then he randomly chooses several pois from the chosen level of the classified POI pool.
- For each chosen poi, he computes the query probability of this poi on each location in the obtained set of candidate locations C.
- Finally, the optimal combination (a location and a poi), which guarantees the minimum query probability, can be produced.
- The time interval between any two continuous queries can be randomly defined, as a result, adversaries have less useful information to predict or infer the real location and real poi of a user.

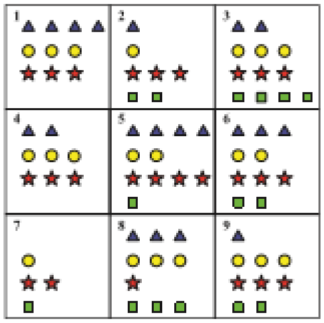
Method:2- A Personalized Two-Tier Cloaking Scheme for Privacy-Aware Location-Based Services.

- Firstly, heavy reliance on the location anonymizer may cause the single point of failure on either system performance or user privacy.
- Secondly, previous work always ignore the relationship between location privacy and query privacy, and only focus on protecting them independently, which may not satisfy the increasing privacy concerns of mobile users.
- Thirdly, although dummy locations/queries can be used to provide anonymity, how to select these information is still a challenge.
- In this paper, they have proposed a personalized spatial cloaking scheme, called TTcloak, which considers the location privacy, query privacy and the desired size of cloaking region, simultaneously. By performing their proposed Dummy Query Determining (DQD) and Dummy Location Determining (DLD) algorithms, a user in TTcloak first determines several dummy queries with similar query probability as her location, and then selects some cells as the candidates.
- The basic idea behind TTcloak can be divided into several steps to provide different properties in our defined (k,l,Adesired)-indistinguishability.
- In the following fig 2,the local map is divided into 9 cells, marked from 1 to 9, respectively. Different shapes represent different queries sent from that cell. Totally, we get 4 types of query in the whole map, the number of each type of query means the query probability.
- Suppose the real user Alice is at the 1st cell and query for the red star (i.e., nearest hospital), the parameters k and l are set as k =3and l =2. As the first step, she counts the number of red star in current cell, and it is 3. To achieve 2-diversity on her query privacy, she needs to find another query, such as the yellow circle in this case whose number is 3 as well.

- Note that, green rectangle and blue triangle cannot be chosen since they can be easily filtered out. In the second step, she aims to get 3-anonymity on her location privacy by searching all the cells and filtering out part of them based on the query probability of the red star in her current cell.

- Specifically, since the determined two queries are red star and yellow circle, she first computes the number of red star in all the other cells, then filters out the cells with different query probability. Here, cells 5, 7 and 8 are gone and the remaining are cells 1, 2, 3, 4, 6 and 9. By this way for yellow circle, cells 2 and 6 are removed, the cells 1, 3, 4 and 9 are left.

- Besides her current cell, she then can assign other two dummy locations into cells 3, 4 and 9 freely, which can guarantee 3-anonymity on her location privacy and 2-diversity on her query privacy, simultaneously.



- **Dummy Query Determining Algorithm:**

  - Given a set of cells (c1,c2,⋯,cn) of the divided local map, we can easily obtain a cell (i.e., cu) where the real user Alice located.

  - Suppose there are m queries issued in the whole system, they are denoted as q1,q2,⋯,qm). For the jth query qj in the ith cell ci, the query probability can be denoted as pij, where 0 <i≤ n and 0 <j≤ m.

  - In this case, we define a query puv, which means that Alice issues the vth query from the uth cell. To achieve l-diversity on protecting her query privacy,our DQD algorithm aims to find a set of dummy queries which have the similar probabilities to be issued as the user's real query in cu.

  - Technical details are shown as follow. (i) In user's current cell cu, she searches all the query probabilities of other m−1 queries except her real query and computes the query difference set as D = _|puj −puv|_, where0 <j≤ m. (ii) She resets the order of elements in the query difference set D to D_ from small to large. (iii) To effectively achieve l-diversity while providing potential randomization, she chooses the former s elements from D' and adds the related s queries into the candidate query set Q, where|Q| = s and l ≤ s<m . By now, she obtains the candidate query set Q which can be presented as 'q1,q2,⋯,qs'.

- **Dummy Location Determining Algorithm:**

  - Alice should filter out some proper cells into a candidate cell set (C) to assign these dummy locations.

  - The query probabilities of different queries within the chosen cells should be like each other as well as the real user's query probability

  - Alice can achieve k-anonymity with freely assigning the dummy locations within the candidate cells in C

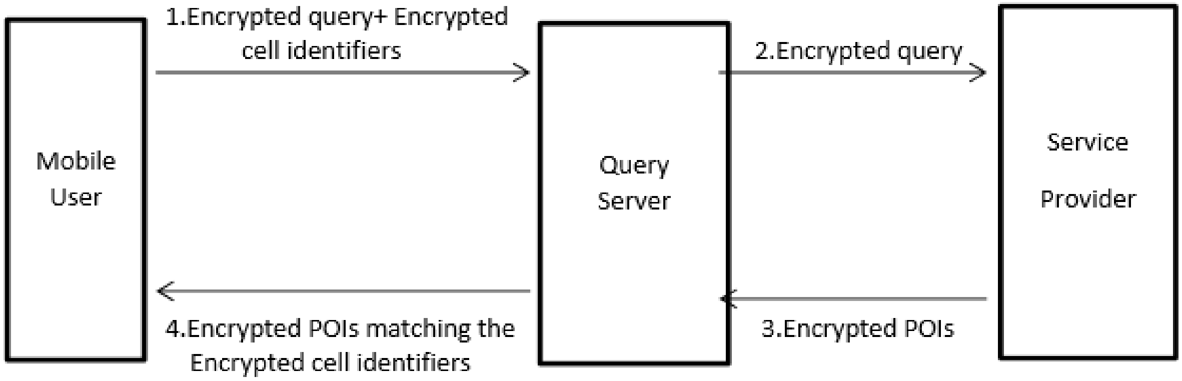
  - DLD algorithm can be executed through the following steps.

    (i) At the initialization phase, the set |C| = n contains all the divided cells within the map. Obviously, the computation complexity is huge when assigning dummy locations and we need to further reduce the size of C.

    (ii) For each chosen query in Q, Alice searches all the candidate cells in C and filters out some cells based on their query probabilities and her real query probability. Any cell should be filtered out if the difference is bigger than a user-defined threshold (e.g., τ =0 .1). Then, the size of the candidate cell set C can be reduced to an acceptable complexity after s rounds of refinement.

Method:3- Dynamic Grid System for Mobile based Location Based Services.

- The system architecture for dynamic grid system (DGS) should be designed to provide privacy-preserving measures of LBS for mobile users.
- Our system consists of three main entities, service providers, query servers and mobile users. We will describe the main entities and their interactions, and then present the two spatial queries, i.e., range and k-nearest-neighbour (NN) queries, supported by our system.


1.Encrypted query+ Encrypted cell identifiers
2.Encrypted query
Mobile User
Query Server
Service Provider
4.Encrypted POIs matching the Encrypted cell identifiers
3.Encrypted POIs

System Architecture for DGS.

- **Service providers (SP):** Each SP is a spatial database association framework that stores the domain data of a specific kind of static POIs to store run data of a specific affiliation. The spatial database utilizes a current spatial summary to record POIs and answer grow ask for (i.e., recover the POIs organized in a specific zone).
- **Mobile users:** Every mobile user is furnished with a GPS-empowered gadget that decides the client's area in the frame. The client can get preview or consistent LBS from our framework by issuing a spatial inquiry to a specific SP through QS.
- **Query servers (QS):** Query Server acts an intermediate between the mobile user and Service Provider. Query Server just passes the encrypted query which is sent by the user. The user shares his current location to other user which is transmitted through query server. Here, Query server does not store any user's location information because everything is encrypted and sent to query server. Query server then sends this request to service provider which in turn service provider responds to query and sent to user through query server. Query Server does not store any information and it is a semi-trusted third party. This enriches user security.
- **Pros & Cons:**
  1. Time obfuscation has better privacy degree and ensures full time requirement of quality of service with less System overhead.
  2. The system overhead is quantified by counting the average number of queries sent to the LBS server. Therefore, the system overhead increases with the simulation time. We see that the performance of Spatial Cloaking-TTcloak [1][2] is the worst, due to the large number (k-anonymity and l-diversity) of queries submitted each time. As a result, the overhead increases almost linearly with the simulation time.
  3. Spatial Cloaking is resistant to colluding attacks and Inference attacks.
  4. How DGS stands in front of other techniques we have seen so far:

Parameters	Computation on cost	Privacy	Cloaking area	Communication cost	GPS connectivity
Other Techniques providing Location-Aware Location Privacy Protection for Mobile.	Medium	Medium	High	Medium	Yes
Dynamic Grid System	Low	High	High	Low	No

- **My Method:**

1. Working:

   - The time required to crack an encryption algorithm is directly related to the length of the key used to secure the communication.
   - Uses 128 bit key to encrypt first and then store the GPS coordinates to the server.
   - It stores location in encrypted form in real time.

2. Advantages of these systems over Other Techniques:

   - More Mathematically efficient
   - Elegant Cryptographic Algorithm
   - Symmetric Key cipher.
   - High Computation complexity for ANY BRUTE FORCE ATTACKS.
   - Fast processing.
   - Cracking a 128 bit AES key with a state-of-the-art supercomputer would take longer than the presumed age of the universe. And AES even uses 256 bit keys.
   - As of today, no practicable attack against AES exists. Therefore, AES remains the preferred encryption standard for governments, banks and high security systems around the world.

- **My final thoughts:**

1. The introduced randomization in both dummy location selection and classified POI pool construction modules in Time obfuscation guarantees their scheme to be resistant from colluding and inference attacks.

2. We saw a two-tier cloaking scheme for personalized users in privacy-aware LBSs, termed TTcloak. With carefully generated dummy users, TTcloak guaranteed user's location privacy and query privacy simultaneously within a user-defined CR against adversaries with side information but at large system overhead.

3. Then we saw DGS- the recommended system which can resolve the problems of other methods associated with computation and communication cost with high privacy levels and low computation and communication cost.

4. In my opinion, the AES encryption in real time processing of GPS coordinates will result in robust and tolerant privacy protection technique by providing less system overheads with dynamic processing capabilities.

References:

[1] Li, F., Wan, S., Niu, B., Li, H., & He, Y. (2016). Time obfuscation-based privacy-preserving scheme for Location-Based Services. 2016 IEEE Wireless Communications and Networking Conference Workshops(WCNCW).

[2] B. Niu, X. Zhu, W. Li, H. Li, Y. Wang, and Z. Lu, “A personalized two-tier cloaking scheme for privacy-aware location-based services,” in Proc. of IEEE ICNC 2015.

[3] Saravanan, G., Sundaramurthy, G., Sanjay, R., & Geetha, R. (2017). Providing location privacy protection using dynamic grid system for mobile location based services. 2017 International Conference on Information Communication and Embedded Systems (ICICES). doi:10.1109/icices.2017.8070791