Optimal Trajectory Partitioning for Enhanced Privacy and Utility in Continuous Location Based Services

Heechang Shin, Jaideep Vaidya, Vijayalakshmi Atluri, Sungyong Choi
MSIS Department, Rutgers University
{hshin,jsvaidya,atluri}@cimic.rutgers.edu, sungyong@pegasus.rutgers.edu

1. INTRODUCTION

To deal with privacy issues in location based services (LBS), the concept of location k-anonymity has been advanced [4]. Location k-anonymity is based on the well-established notion of k-anonymity [12]; a dataset is said to be k-anonymized if each record is indistinguishable from at least $k-1$ other records with respect to certain identifying attributes [5]. In the LBS environment, an LBS request is said to preserve location k-anonymity if an adversary cannot distinguish the actual query issuer from at least k-1 other users. In order to achieve location k-anonymity, given an LBS query, recent approaches by Gruteser et al. [4] and Mokbel et al. [7] remove the identifying information such as user id and transform the exact location into a bounding box, called generalized region (GR), containing at least k people within it by utilizing a trusted third party.

However, these approaches to location privacy are limited to point queries, where query answers are based on the current location of the issuer, and thus, they are not suitable to new types of LBS services such as continuous nearest neighbor searches [13], where the query answers are based on continuous points along the route of the issuer. Since the query results are not based on a single location, processing such a *continuous LBS* query assumes that the knowledge of the user trajectory is known to the (untrusted) LBS service provider. For example, a user may ask for the closest gas stations along his planned itinerary from Las Vegas, NV to Los Angeles, CA (say by taking US 95 and US 40). The result is not just one gas station, but rather constitutes a set of gas stations with ⟨ *location*, *time interval* ⟩ tuples, such that the *location* of gas station is the nearest neighbor to a point on the trajectory corresponding to a specific *time interval*. The time interval is based on the user's velocity, and instead could be replaced by trajectory portions (i.e., the gas station at Ludlow is the closest along the entire stretch of US 40). Such trajectory traceability by an untrusted entity obviously raises privacy concerns. Observe that given a future trajectory movement, the query answer can be pre-computed in advance in this type of query.

To address this, we extend the notion of location k-anonymity to *trajectory k-anonymity*. Under trajectory k-anonymity, a user trajectory is being anonymized by at least $k-1$ other trajectories: therefore, a user's trajectory remains indistinguishable from at least $k-1$ other trajectories. However, the proposed trajectory k-anonymity requirement can lead to considerable GR expansion and associated loss of accuracy, thus not satisfying minimum QoS thresholds. The situation is further aggravated if the anonymization is over a sparse area. Therefore, to enable both privacy and quality of results, we propose a novel *trajectory partitioning* scheme.

2. PROBLEM STATEMENT

In order to process continuous LBS requests, there are two main approaches: (1) an LBS request is submitted repeatedly for each time instance until it expires, thus requiring the system evaluate the results continuously, and (2) the query result is computed only once if the information on the future trajectory is provided. The first approach suffers from the drawback of sampling [14], i.e., if the sampling rate is too low, the results will be incorrect; otherwise, there is a significant computational overhead, implying that there is no guarantee about the query results. Therefore, in this project, we focus on the continuous LBS environment where the query results can be computed in advance. Existing privacy-preserving work for continuous LBS [2, 16] only consider the first case, thus, still having the issues of sampling and correctness of query results. We are the first to address the anonymity of users based on the revealed information of future trajectory.

2.1 System Architecture and Adversary Model

Figure 1 shows the system architecture of existing work [3, 4, 5, 7, 10, 11, 16]. Each mobile user has its own privacy level k which indicates that the user wants to be k-anonymous, i.e., indistinguishable among k users. A mobile user uses a mobile device such as PDA or cellular phone to send a LBS request (including future trajectory and k) to a trusted location server (LS) using wireless technologies through a secure channel, such as secure socket layer (step 1). LS maintains the (past as well as current) locations and velocity of users and uses those information to perform anonymization (step 2) based on users' privacy requirements (k). The LS first removes any identifying information from the original request and anonymizes it by replacing the trajectory with the GR (based on the given anonymization model). Thus, the anonymized request includes GR and is forwarded to the LBS providers (step 3). On receiving the anonymized request from LS, the LBS provider computes a candidate list of answers satisfying the request, and sends it back to the LS (step 4). Then, LS sends the actual result back to the user requesting the service (step 5). The LS is considered to be a trusted party and can be implemented as a globally distributed service to minimize single point of failure and being attractive to hacking attacks [17].

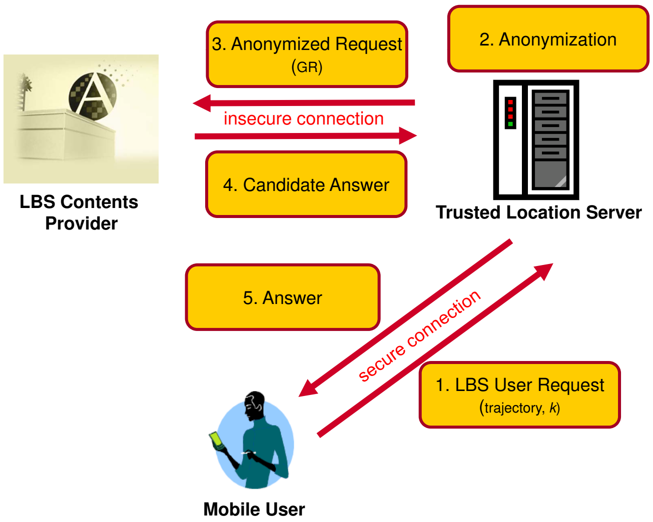

Figure 1: System Architecture

We assume that an adversary has the knowledge of (1) anonymized requests and (2) (past as well as current) user location and velocity from an external source. The first assumption states that (1) an adversary cannot gain access to original requests because the communication channel between a mobile user and the LS is secure, so that any entity eavesdropping on this channel still cannot recognize the contents of the messages, and (2) an adversary can be an LBS provider or the entity eavesdropping on an insecure communication channel between the LS and LBS providers. Therefore, any information submitted to the LBS provider is a potential threat to the privacy of mobile users if they are utilized to identify the query issuer. The second assumption states that the location and velocity of at least a few users within the vicinity of the targeted victim are revealed through triangulation, public databases, physical observation, and so on [5]. For example, traffic monitoring services such as Delcan technology can compute the location and speed of a vehicle by measuring the time of handoffs from cell to cell [8] in Maryland. Then, the movement direction can also be computed by utilizing an underlying road network. If an LBS provider can collude with traffic monitoring services, the location and direction of users can be revealed, and this information used to estimate the future trajectory by using a mobility model such as Gauss-Markov mobility model [6, 15]. Then, the identity of the mobile users can be possibly identified if the submitted trajectory information belongs to a particular user.

The objective of an adversary is to infer the identity of the query issuer in order to learn sensitive information about the user. This is because, the query itself unintentionally reveals sensitive information about the user. For example, assume Tom submits a continuous nearest neighbor query to find the nearest casinos along his path to the destination. If an adversary can identify Tom as a user who is likely to submit the query, his query information can be used to reveal his gambling habit.

2.2 Optimal k-Trajectory Partitioning

Given a user request, a trajectory is anonymized if GR includes at least $k-1$ trajectories of other users' requests. The following definition formalizes this notion.

DEFINITION 1 (TRAJECTORY k-ANONYMITY). *Given a user request $r \in R$ (R is a set of submitted user requests) and a spatiotemporal region GR, we say that trajectory k-anonymity is ensured for the query issuer of r if $\exists R' \subseteq R$, such that the number of users in R' is greater than or equal to r's specified k, and each trajectory of R is located completely within the GR in all time instance of r's query duration.*

In other words, LS ensures trajectory k-anonymity by creating a spatiotemporal region that includes trajectories of at least k-1 other requests. Thus, for any user request, the submitted trajectory of the user is transformed into a spatiotemporal region that would include at least k mobile users' trajectories after removing any identifying information (i.e. ID of a mobile user). However, the main issue of applying trajectory k-anonymity on the submitted trajectories is that it may not guarantee the target QoS level of the continuous LBS. It is well known that there exists an inverse relationship between the service quality and the level of privacy [3, 7]. This is because, better privacy is provided by increased generalization of a LBS region (i.e. larger number of k), which tends to create larger anonymized region. This may have adverse effect on the accuracy of the result. For example, for a continuous nearest neighbor LBS search with trajectory of 50 miles looking for nearest restaurants along the path, due to larger GR, the restaurants that are farther to the actual path will also be included.

Motivated by the limitations in the existing work, along with the inverse relationship between the service quality and the level of privacy, we aim to facilitate privacy of users while improving the QoS in a continuous LBS environment. To this end, we define the optimal k-trajectory partitioning problem which is defined as follows.

DEFINITION 2 (OPTIMAL k-TRAJECTORY PARTITIONING). *Given a set of submitted k future trajectories, our objective is to find a set of n-partitioning time points such that the sum of each partitioned regions is minimized while each partitioned region encloses at least k trajectories. We believe that the optimal trajectory partitioning can improve the privacy and QoS level at the same time.*

3. PROPOSED APPROACH

Proposing a general method that can find the optimal split points for arbitrary length trajectories is a complex task. Although we do not assume that the number of splits required for an optimal partitioning is known *a priori* in the optimal trajectory partitioning problem, this makes the problem significantly more difficult. In practice, we can assume that the maximum number of splits is given as input, and the algorithm simply has to find the best split time points.

Given n number of users in the system and a continuous LBS request with parameters k and m, we proceed in the steps outlined below.

1. *Find a candidate set of trajectories for anonymization:* Although it is possible to use any $k-1$ trajectories as candidates for anonymization, it is desirable to find them in such a way that they would lead to a small GR size due to an inverse relationship between the level of location privacy and

the level of QoS. Existing work can be categorized as (1) selecting k-nearest neighbors of trajectories and (2) minimum GR generation. The first approach is self-explanatory, and the second approach finds $k-1$ trajectories that would generate the minimum area of GR.

However, both approaches are susceptible to the privacy attack if the adversary has the additional knowledge of how the trajectory is being selected for anonymization. In order to deal with this, we employ a heuristic for selection process by utilizing TPR-tree [9]. Because the users selected by the heuristic have the same probability of submitting the request, it is not susceptible to known privacy attacks. Also, it generates smaller GR size because the objective function of the tree (i.e., generating the smallest sum of volumes) matches with our purpose.

2. *Partition the set of trajectories in an optimal manner:* The GR of those trajectories from the previous step is partitioned into m ($m \geq 1$) GRs. In order to minimize the cost (i.e., the sum of partitioned regions is minimized), we employ a solution based on a nonlinear programming method, called Karush-Kuhn-Tucker conditions [1]. This is a generalization of the method of Lagrange multipliers to have inequality constraints. After solving the optimization problem, fuzziness can be introduced in the length and time interval of the partition to minimize the risk of the adversary to reconstruct the trajectory from multiple paths.

Observe that since partitioning is performed on the set of trajectories that are used to form the anonymity set in the trajectory k-anonymity, we guarantee that each partitioned TMBR actually includes at least k different trajectories including that of the query issuer. Thus, one can see that the partitioned GRs contain the same set of trajectories as that of the non-partitioned GR. After partitioning, the LS submits all of the partitioned GRs to the LS, and the LS computes the candidate answers for each GR.

4. MAIN CONTRIBUTIONS

The main contributions of this project are summarized as follows:

- We present a trajectory k-anonymity model for protecting the privacy of users in a continuous LBS environment.
- We propose optimal trajectory partitioning methods that can achieve enhanced privacy and service quality.
- We can experimentally demonstrate that our partitioning approach is both efficient in practical situations and significantly outperforms existing trajectory partitioning approaches by using synthetic and real data sets.

5. REFERENCES

[1] Avriel, M. *Nonlinear programming: analysis and methods*. Dover Publications, 2003.

[2] Chow, C., and Mokbel, M. Enabling private continuous queries for revealed user locations. *Lecture Notes in Computer Science 4605* (2007), 258.

[3] Gedik, B., and Liu, L. Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms. *IEEE Transactions On Mobile Computing* (2008), 1–18.

[4] Gruteser, M., and Grunwald, D. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. *Proceedings of the 1st international conference on Mobile systems, applications and services* (2003), 31–42.

[5] Kalnis, P., Ghinita, G., Mouratidis, K., and Papadias, D. Preventing location-based identity inference in anonymous spatial queries. *IEEE transactions on knowledge and data engineering 19*, 12 (2007), 1719–1733.

[6] Liang, B., and Haas, Z. Predictive distance-based mobility management for PCS networks. In *IEEE INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings* (1999), vol. 3.

[7] Mokbel, M., Chow, C., and Aref, W. The new Casper: query processing for location services without compromising privacy. *Proceedings of the 32nd international conference on Very large data bases* (2006), 763–774.

[8] Ravi, N., Gruteser, M., and Iftode, L. Non-inference: An information flow control model for location-based services. In *Mobile and Ubiquitous Systems-Workshops, 2006. 3rd Annual International Conference on* (2006), pp. 1–10.

[9] Šaltenis, S., Jensen, C., Leutenegger, S., and Lopez, M. Indexing the positions of continuously moving objects. *ACM SIGMOD Record 29*, 2 (2000), 331–342.

[10] Shin, H., Atluri, V., and Vaidya, J. A Profile Anonymization Model for Privacy in a Personalized Location Based Service Environment. *The 9th International Conference on Mobile Data Management (MDM)* (2008).

[11] Shin, H., Vaidya, J., and Atluri, V. Anonymization models for directional location based service environments. *accepted for publication in Computers & Security* (2009).

[12] Sweeney, L. k-anonymity: A model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10*, 5 (2002), 557–570.

[13] Tao, Y., Papadias, D., and Shen, Q. Continuous nearest neighbor search. In *VLDB* (2002), pp. 287–298.

[14] Tao, Y., Papadias, D., and Sun, J. The TPR*-tree: an optimized spatio-temporal access method for predictive queries. *Proceedings of the 29th international conference on Very large data bases-Volume 29* (2003), 790–801.

[15] Tolety, V. Load reduction in ad hoc networks using mobile servers. *Master's thesis, Colorado School of Mines* (1999).

[16] Xu, T., and Cai, Y. Location anonymity in continuous location-based services. In *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems* (2007), ACM New York, NY, USA.

[17] Youssef, M., Atluri, V., and Adam, N. Preserving mobile customer privacy: an access control system for moving objects and customer profiles. *Proceedings of the 6th international conference on Mobile data management* (2005), 67–76.