Using Uncertainty to Provide Privacy-Preserving and High-Quality Location-Based Services

Reynold Cheng and Sunil Prabhakar

Department of Computer Science, Purdue University
{ckcheng,sunil}@cs.purdue.edu

Abstract. Location-based services, such as finding the nearest gas station, require users to supply their location information. However, a user's location can be tracked without her consent or knowledge. Lowering the spatial and temporal resolution of location data sent to the server has been proposed as a solution. Although this technique is effective in protecting privacy, it may be overkill and the quality of desired services can be severely affected. In this paper, we investigate the relationship between uncertainty, privacy, and quality of services. We propose using probabilistic queries to precisely control location uncertainty. We also suggest a framework where uncertainty can be controlled to provide high quality, privacy-preserving services.

1 Introduction

Positioning technologies such as GPS, GSM, RF-ID and WiFi(802.11) have undergone rapid developments in recent years [7, 8, 4]. These new technologies allow locations of users to be determined accurately, and enable a new class of applications known as Location-Based Services (LBS). An important LBS is the E-911 application mandated by the U.S. (correspondingly E-112 in Europe), which requires cell phone companies to provide an accurate (within a few hunderd feet) location of a cell phone user that call for emergency help [4]. Another example is the use of RF-ID tags on items such as razors in large departmental stores for inventory management [8].

Although LBS applications hold the promise of safety, convenience, and new business opportunities, the ability to locate users and items accurately also raises a new concern – intrusion of *location privacy*. According to [1], location privacy is defined as the "ability to prevent other parties from learning one's current or past location". For example, a service provider can track the whereabouts of a user and discover her personal habits. Preventing location privacy from being invaded is thus of utmost importance.

Of course, if the user provides little location information to the service provider, the risk of her privacy being compromised will be significantly reduced. However, this may prevent an LBS from providing the best service to the user. Alternatively, a user can *cloak* her information before sending it to the LBS, by providing her location at a coarse granularity in terms of time and space [4, 1]. This reduces the degree of accuracy, or quality of the service provided by the LBS, but provides better protection for the user's privacy. According to [6], "there is an inherent tradeoff between the utility that databases can offer and the privacy they afford their constituents...to understand the relationship between privacy and utility, and thereby find a comfortable position between

the extremes of fully disclosed and completely withheld data". Indeed, we identify a tradeoff among: (1) How fuzzy or uncertain the location information sent by a user to the LBS can be, (2) the quality of service provided by the LBS, and (3) the location privacy of the user. We investigate the interaction of these three components. We first discuss related works in location privacy and uncertainty management in Section 2. We then propose our model that captures data uncertainty, privacy and quality of service and outline interesting research problems in Section 3. We conclude the paper in Section 4.

2 Location Cloaking and Uncertainty Management

An approach for protecting user privacy is to *sanitize* or *anonymize* user information before it is dispatched to service providers. In general, location information can be anonymized by reducing temporal and spatial resolutions of location information (called *location cloaking* in [4]). These techniques can be found in anonymous applications (i.e., those that work with location information only and do not require user's identities) and pseudonymous applications (i.e, those that need to know the identity of a user, but the user's pseudonym can be used in place of her true name). They assume a central, trusted server between users and the LBS.

Anonymous applications. Gruteser and Grunwald [4] suggested that users "cloak" the location data before sending them to the LBS. Specifically, let (x,y,t) be the location (x,y) at time t sent by the user to the anonymity server. Using their *cloaking algorithm*, the anonymity server outputs a *cloaking tuple* $([x_1,x_2],[y_1,y_2],[t_1,t_2])$ to a LBS, where $([x_1,x_2],[y_1,y_2])$ is the rectangular area within which (x,y) is found, between the time interval $[t_1,t_2]$. In essence, the LBS receives location information of a coarse granularity from the anonymity server, so that location privacy of the user is protected. To measure the effectiveness of the cloaking tuple in protecting location privacy, the authors introduced a metric known as k-anonymity. This metric measures between time interval $[t_1,t_2]$, the number of users, k, at the same spatial vicinity $([x_1,x_2],[y_1,y_2])$. A larger value of k indicates more difficulty in linking a location to a particular user. The value of k is specified by the user as a parameter to control her desired level of privacy.

Pseudonymous Applications. In this class, a location value is associated with a fake user identity (called pseudonym). When the user is observed to stay at a certain place for a long time, and that place corresponds to that user's office or home, her identity can be easily revealed even though she is using a pseudonym. Beresford et al. [1] proposed location privacy to be protected through a trusted middleware that renames pseudonyms frequently, so that a user's identity cannot be traced. Moreover, this renaming is done while there are at least k users in the same zone at the same time period. This metric, similar to k-anonymity, provides privacy protection to a certain degree.

The issues of uncertainty in a moving-object database are studied in [3]. In such a system, object locations are constantly reported to the database. These location values are subsequently used to answer user queries. Due to continuous changes in locations, as well as limited resources (e.g., network bandwidth and battery power), it is infeasible for the database to keep track of the actual location of every object. Queries that use stale values in the database can produce incorrect answers. However, if the degree of uncertainty between the actual location value and the database value is limited, one can

place more confidence in answers to the queries in terms of probability values e.g., John and Peter are 70% and 30% likelihood respectively of being nearest to Mary. In [3], the issues of modeling uncertainty for a moving-object, as well as querying algorithms for evaluating uncertain locations and providing probabilistic answers, are studied. In [2], the issues of quantifying the quality of answers to probabilistic queries are quantified.

3 Uncertainty, Privacy and Quality of Service

In many applications, location values can be inaccurate to a certain degree. For example, to find a coffee shop nearby, it may not be necessary to supply a precise location to the LBS. In the 911 emergency services for wireless phone users (E-911), handset users must be located with an accuracy of 50 to 150 meters [5]. Thus, it is possible to artificially inject uncertainty to location data by, for example, using the location cloaking algorithm in [4]. The question is: to what degree can uncertainty be injected so that the uncertainty will not severely reduce the quality of service while preserving privacy?

In order to understand the relationship among location uncertainty, privacy and quality of service, we propose the preliminary framework shown in Figure 3. The user specifies her precise location, her service request and privacy requirements, to a *location cloaking engine*. The engine then injects uncertainty into location values i.e., produce a location value with a lower spatial and temporal resolution. The service provider executes the service request on the uncertain data and sends back probabilistic results to the user (e.g., John has a 90% chance of being closest to me). The quality of its service is also fed back to the location engine, in order to adjust the level of uncertain information and provide a better service quality if necessary. Using probabilistic queries, it is possible to quantify the quality of the results [2].

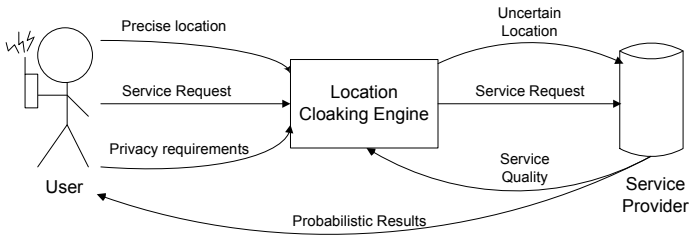

Fig. 1. A Framework Capturing Uncertainty, Privacy and Quality of Service

Essentially, the framework uses concepts similar to probabilistic queries to control precisely the amount of uncertainty injected into a location value. Probabilistic queries can also evaluate uncertain data and measure quality of services. We believe that the proposed framework can facilitate a more systematic understanding of the problem. Below we identify several interesting questions:

1. The location cloaking algorithms proposed in [4] and [1] only consider privacy, and thus may result in poor service quality. Moreover, they use k-anonymity to measure

privacy, which implies that at least k users have to be located in a certain region at a particular time.
2. k-anonymity does not consider the area of the cloaking tuple. It is also not clear how large the value of k should be.
3. How to provide location privacy for non-anonymous applications (i.e., applications that do not work without knowing a user's true identity)?
4. The location cloaking engine was proposed as a trusted middleware in [4, 1]. Can it be built on the user's system so that she does not need to trust any middleware?
5. How can a service provider manipulate uncertain location data and provide probabilistic guarantees of results in an efficient manner?
6. How can location uncertainty be controlled by the quality of probabilistic results?
7. How can a user specify her location privacy requirements for different applications? If the quality of the service provided is poor due to low location resolution, can a user conveniently reduce her privacy to get better service?

4 Conclusions

The idea of injecting uncertainty to anonymize location data and thereby protect lcoation privacy has been proposed recently. However, these anonymizers do not consider the quality and accuracy of services provided, and it is also not clear how such uncertain data (e.g., cloaking tuple) can be manipulated by service providers. We propose the use of *probabilistic queries* to solve this problem, which provide a precise control over data uncertainty, and allow quantitative measurements of service quality. We further suggested a simple framework that connects privacy, uncertainty and quality of service, in order to have a better understanding of these quantities. Our next goal is to develop a platform where an LBS can be delivered to users with high degree of quality, accuracy and privacy.

References

1. Alastair R. Beresford and Frank Stajano. Location Privacy in Pervasive Computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.
2. R. Cheng, D. Kalashnikov, and S. Prabhakar. Evaluating probabilistic queries over imprecise data. In *Proc. of the ACM SIGMOD Intl. Conf. on Management of Data*, June 2003.
3. R. Cheng, D. V. Kalashnikov, and S. Prabhakar. Querying imprecise data in moving object environments. *IEEE Transactions on Knowledge and Data Engineering (To appear)*, 2004.
4. Marco Gruteser and Dirk Grunwald. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In *Proceedings of the First International Conference on Mobile Systems, Applications, and Services*, May 2003.
5. Raddcomm Wireless Consulting Services L.L.C. E-911 overview. http://www.raddcomm.com/E911
6. Microsoft Research. Database privacy group. http://research.microsoft.com/research/sv/DatabasePrivacy/.
7. T. Robinson. Location is everything. *Internet week online*, September 12, 2000.
8. J. Warrior, E. McHenry, and K. McGee. They know where you are. *Spectrum*, 40(7):20– 25, July 2003.