Protecting User Privacy Better with Query l-Diversity

Fuyu Liu, University of Central Florida, USA

Kien A. Hua, University of Central Florida, USA

ABSTRACT

This paper examines major privacy concerns in location-based services. Most user privacy techniques are based on cloaking, which achieves location k-anonymity. The key is to reduce location resolution by ensuring that each cloaking area reported to a service provider contains at least k mobile users. However, maintaining location k-anonymity alone is inadequate when the majority of the k mobile users are interested in the same query subject. In this paper, the authors address this problem by defining a novel concept called query l-diversity, which requires diversified queries submitted from the k users. The authors propose two techniques: Expand Cloak and Hilbert Cloak to achieve query l-diversity. To show the effectiveness of the proposed techniques, they compare the improved Interval Cloak technique through extensive simulation studies. The results show that these techniques better protect user privacy.

Keywords: Cloaking, Location-Based Services, Location k-anonymity, Privacy Protection, Query Processing

1 INTRODUCTION

With rapid advances in wireless communication and wide spread of location positioning systems, *Location-Based Services* (LBS) are becoming increasingly popular. Mobile users can send queries, such as "where is the nearest gas station?" or "where is the closest women clinic?", to service providers and get query results back. These queries are called location-based queries, because they typically consist of a location, usually the location of the query issuer, and a question. There are a lot of challenges to answer this type of query for mobile users. One of the challenges is how to protect user's privacy. As we can see, to get a query result, one has to provide a location as well as a query question to the service provider, which raises a privacy concern if the service provider is not trustworthy.

Many researches have been done to address this challenge (Gruteser and Grunwald, 2003; Gedik and Liu, 2005; Mokbel et al., 2006; Chow and Mokbel, 2007; Bamba et al., 2008; Xu and Cai, 2007, 2008). Most existing solutions assume a three tier architecture, in which mobile users first send the location and query information to a *trusted* anonymizer server, then the anonymizer server performs some cloaking procedure to enlarge the query's location into a region, finally forwards that region to a service provider. Typically, the goal of this cloaking

DOI: 10.4018/jisp.2010040101

procedure is to enforce location k -anonymity. That is, the cloaked region must contain at least $(k-1)$ other users, such that an adversary can only link the cloaked query to the actual query issuer with $\frac{1}{k}$ probability. To protect against a *query sampling attack* (Chow and Mokbel, 2007), techniques have been proposed to ensure that all users included in the same cloaked region must report this region as their cloaked region.

Enforcing *k*-anonymity alone is not sufficient to ensure privacy. Let us consider a scenario, in which all users from a cloaked region are interested in the same type of service such as the location of a special club. In this case, even an adversary cannot link an individual query back to a specific user, it is still known to the adversary that all the users in the cloaked region have inquired about that special club. While this example depicts an extreme case, in reality, it is not uncommon that users from the same cloaked region request only a limited number of services. Consequently, an adversary can still infer that some user has issued a query on a certain service with a high probability. This kind of attack is referred to as *query homogeneity attack* (Xiao et al., 2008), and renders the existing *k*-anonymity model vulnerable. To counter this kind of attack, we modify the *l*-diversity concept (Machanavajjhalaet al., 2006), originally proposed for the relational database domain, and apply it in LBS domain to protect query contents. The key idea is to ensure that for all queries sharing the same cloaked region, their query contents must be different enough, such that the probability of linking a query to its original issuer is less than some pre-defined threshold.

In this paper, we first formally define the problem, and then propose two cloaking techniques that can counter against query homogeneity attack. Both of these techniques first divide the whole terrain into grid cells. Their space partitioning schemes, however, are different. The first technique starts from the center cell, and gradually expands over the space in all directions in search for a good way to partition the space. In contrast, the second technique first maps the two-dimensional grid space into a one-dimensional line of grid cells using a space filling curve, and then sequentially scans these cells to find the best partitioning strategy. We will describe these techniques in details later, and give simulation results to show that they are significantly better than the improved Interval Cloak technique (Gruteser and Grunwald, 2003).

The contributions we make in this paper can be summarized as follows:

- To the best of our knowledge, we are the first to use the l -diversity concept to address the query homogeneity attack.
- We consider a new anonymization criteria: $\langle k, l \rangle$ -sharing region, and propose two cloaking techniques to partition the space using this new criteria.
- We conduct extensive simulation studies to evaluate the proposed techniques.

The remainder of this paper is organized as follows. We first discuss related work in Section 2. The preliminary and some definitions are then presented in Section 3 to facilitate further discussion. The two proposed cloaking techniques are introduced in Section 4, followed by the simulation study in Section 5. Finally, we conclude this paper in Section 6.

2. RELATED WORK

In this section, we discuss the two concepts: *k*-anonymity and *l*-diversity in relational databases and their applications in location-based services.

2.1 K-Anonymity and l-Diversity in Relational Databases

In a relational database, to publish data (such as censor or medical data) to support third-party data mining applications, it is important to prevent an adversary from linking the published data back to an individual. One obvious solution is to remove the *identifiers* such as a person's name and social security number for