Anonymity in Preference-Aware Location-based Services without Third Trusted-Party

Félix J. García Clemente
Dpto. Ingeniería y Tecnología de Computadores
University of Murcia
Murcia, Spain
fgarcia@um.es

ABSTRACT

Mobile devices equipped with indoor positioning capabilities can access a broad range of different Location-Based Services (LBS). There are advanced LBS applications that use the user's location and preferences in order to give the most precise answer to location-dependent queries. To protect privacy, the user's location and preferences must not be disclosed. Existing solutions utilize a trusted anonymizer between the users and the LBS. This approach has the drawback that this component may not always be available, and it may itself present security problems. We propose a novel framework to support private location-dependent queries, based on the concept of k-anonymity to protect the user's identity, which does not require a trusted third-party, since privacy is achieved via grid-maps and entropy-based techniques.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection; H.2.8 [**Database Management**]: Database Applications—*spatial databases and GIS*

Keywords

Anonymity, privacy, location-based services

1. INTRODUCTION AND MOTIVATION

Nowadays the availability of the communication devices (e.g., mobile phones, tablets, wearables, and others) and the deployment of the indoor positioning systems allow users to easily access a broad range of Location-Based Services (LBS) from basic applications based on the user's location (e.g. "find the nearest restaurant") to more advanced ones like recommendators that include the user's profile (e.g. "find the nearest restaurant offering quality Spanish and Chinese food"). However, mobile applications hardly ever incorporate security mechanisms that avoid to disclose sensitive information online to potentially malicious LBS, including real-time location tracking data or even lifestyle preferences, that may result in unsolicited advertisement (i.e., spam) or worse situations (e.g. stalking or mugging).

To address the privacy issue, most existing solutions adopt the concept of k-anonymity [2] and rely on a third trusted-party (TTP) that acts as an anonymizer between the users and the LBS. In general, this mechanism tries to find a set of k users that are indistinguishable from each other such that an attacker cannot identify a single user out of the set. This set of k user is received by the LBS that sends back a set of k answers. Malicious LBS cannot identify the user with probability larger than $1/k$. Note, however, that in order to guarantee a non-negligible level of privacy, the $k-1$ users cannot be selected naively.

Existing k-anonymity-based location privacy techniques exhibit three significant limitations. First, some require a TTP anonymizer that maintains all user locations. Such a component may not always be available, and it may itself present security problems. Second, a large number of cooperating, trustworthy users is needed. And third limitation, the underlying k-anonymity considers the notion of distance between identities, but it misses the distance between locations.

In order to achieve the user's privacy in non-trusted systems, we propose a framework for preference-aware location-based queries implementing the concept of k-anonymity to protect the user's identity. Our solution incorporates new features that distinguish it from other approaches, our contributions are:

- We propose a framework where the anonymizer is a software component of the user's mobile, so no TTP is used to achieve the user's privacy.
- We present an anonymizer that generates a set of $k-1$ fake users with dummy locations and preferences by carefully selecting algorithm which considers entropy and distance between the k users.
- We design a selecting algorithm based on grid-maps that can be easily supervised and even managed by users in order to get trustworthy results.

There are several works presenting the state-of-the-art techniques to protect location privacy [1, 3], and they distinguish the following principles: position dummies, spatial obfuscation, encryption, and k-anonymity. Different from existing works, our proposal is based on k-anonymity, generates fake users, and does not require a TTP.

.

MOBIQUITOUS 2015, July 22-24, Coimbra, Portugal

DOI 10.4108/eai.22-7-2015.2260301

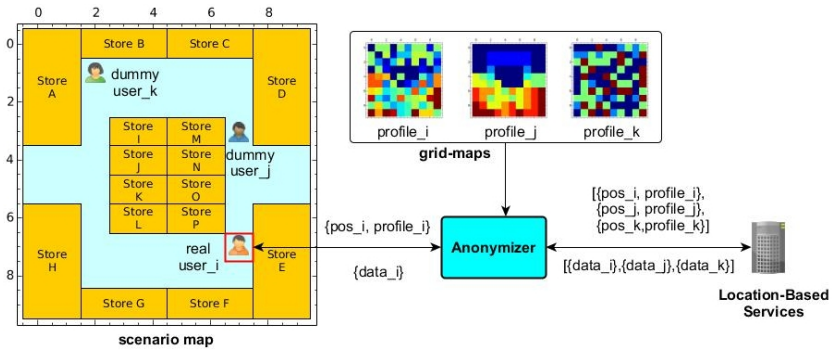

Figure 1: Example scenario with k-anonymity and grid-maps

2. PROPOSAL

The main component of our proposal is an anonymizer that is a trusted middleware located in the user's mobile device. It acts as an intermediary and privacy shield between the user and the LBS. When the user wants to make use of a LBS, her mobile device sends its position information and the user's preferences together with the location-dependent query to the anonymizer, which manages and protects the user's information making the user indistinguishable from other $k-1$ dummy users.

To increase the privacy protection, we use entropy to measure the privacy level, i.e. the uncertainty to identify a user from the candidate set. To compute the entropy, each possible user has a probability of querying, denoted by p_i, and the sum of all probabilities p_i is 1. Then, the entropy H of identifying a user in the candidate set is defined as

$$H = -\sum_{i=1}^{k} p_i \cdot \log_2(p_i) \tag{1}$$

The anonymizer determines a maximum value of H to guarantee a set of k users indistinguishable. The maximum entropy is achieved when all the k possible users have the same probability, i.e. $1/k$, where the maximum entropy will be $log2(k)$. The user's probability p_i is linked to the user's location by a grid-map that is a statistical map that shows where a user is likely to be found under his preferences. The grid-maps could be built by crowdsourcing techniques and supervised by users.

The selected fake users and real user must be separated as far as possible from each other while entropy maintains a minimum acceptable value. A fake user closed to the real user discloses the position, so this dummy must be avoided by the anonymizer. In this sense, a fake user with a grid-map much more different than the grid-map of the real user must be also avoided by the anonymizer, because it decreases the entropy between grid-maps.

The selection algorithm selects the most different dummy user's profiles, but as likely as the real user's profile. Given a degree of anonymity k, besides the real user's profile, we need to determine the other $k-1$ user's profiles. The following show how the algorithm addresses this problem:

1. The candidate user's profiles must have a minimal *diversity*. That is to ensure there are at least l distinct values for the each feature in the set of k selected profiles. The algorithm creates all the sets of profiles that fulfill this restriction.
2. The candidate user's profiles must have a minimal *closeness* to the real user's profile. The algorithm uses the Earth Mover's distance that is based on the minimal amount of work needed to transform one distribution to another by moving distribution mass between each other. Let P and Q be grid-maps divided into $m \times m$ cells, and d_{ij} be the ground distance between position i of P and position j of Q. We want to find a flow $F = [f_{ij}]$ where f_{ij} is the flow of mass from position i of P to position j of Q that minimizes the overall work:

$$WORK(P,Q,F) = \sum_{i=1}^{m^2}\sum_{j=1}^{m^2} d_{ij} f_{ij} \tag{2}$$

3. Based on the previous steps, the algorithm selects the best $k-1$ candidate user's profiles. After that, the algorithm calculates the entropy considering each position of the grid-maps and selects the set of dummy positions which maximize the relation between entropy and the distance between locations.

The Fig. 1 shows an example scenario (i.e. shopping center) and the anonymizer process. In order to preserve the privacy, the user utilizes an anonymizer installed in its mobile device to send her position and profile to the LBS server. The anonymizer will include two fake users that will be generated using the selecting algorithm.

3. ACKNOWLEDGMENTS

This work was supported by the Spanish MINECO, as well as European Commission FEDER funds, under grant TIN2012-38341-C04-03.

4. REFERENCES

[1] J. Krumm. A survey of computational location privacy. *Personal and Ubiquitous Computing*, 13(6):391–399, 2009.

[2] L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[3] M. Wernke, P. Skvortsov, F. Dürr, and K. Rothermel. A classification of location privacy attacks and approaches. *Personal and Ubiquitous Computing*, 18(1):163–175, 2014.