Computer Communications 64 (2015) 21–32

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

A uniformity-based approach to location privacy

Pericle Perazzo *, Gianluca Dini

Department of Information Engineering, University of Pisa, Largo Lucio Lazzarino 1, 56122 Pisa, Italy

ARTICLE INFO

Article history:
Received 10 April 2014
Received in revised form 3 December 2014
Accepted 28 February 2015
Available online 31 March 2015

Keywords:
Privacy
Location-based services
Obfuscation techniques
Uniform obfuscation

ABSTRACT

As location-based services emerge, many people feel exposed to high privacy threats. Privacy protection is a major challenge for such services and related applications. A simple approach is *perturbation*, which adds an artificial noise to positions and returns an obfuscated measurement to the requester. Our main finding is that, unless the noise is chosen properly, these methods do not withstand attacks based on statistical analysis. In this paper, we propose UniLO, an obfuscation operator which offers high assurances on obfuscation uniformity, even in case of imprecise location measurement. We also deal with service differentiation by proposing three UniLO-based obfuscation algorithms that offer multiple contemporaneous levels of privacy. Finally, we experimentally prove the superiority of the proposed algorithms compared to the state-of-the-art solutions, both in terms of utility and resistance against inference attacks.

1. Introduction

Recent years have seen the widespread diffusion of cheap localization technologies. The most known is GPS, but there are many other examples, like cellular positioning and ultra-wide band positioning. [1–3]. The emergence of such technologies has brought to the development of *location-based services* (*LBS*) [4–6], which rely on the knowledge of location of people or things. The retrieval of people's location raises several privacy concerns, as it is personal, often sensitive, information. The indiscriminate disclosure of such data could have highly negative effects, from undesired location-based advertising to personal safety attempts.

A classic approach to the problem is to introduce strict access-control policies in the system [7,8]. This approach has a main drawback: if the entity does not need complete (or exact) information, granting the access to it is a useless exposure of personal data. The "permit-or-deny" outcome of access control is often too rigid.

Samarati and Sweeney [9,10] proposed the concept of *k-anonymity*: a system offers a k-anonymity to a user if his identity is undistinguishable from at least $k-1$ other users. k-anonymity concepts have been applied to location privacy [11–13] by obfuscating the user's position in such a way to confuse it with the positions of other $k-1$ users. Location k-anonymity offers high levels of privacy, because it protects the user's identity. However, since k-anonymity does not permit the identification of the user, it is not applicable in services in which the user authenticates, e.g. payable services or location-based social networks. In addition, they require the presence of $k-1$ users in the proximity, that could be missing, and a central anonymizer, that could not be fully trusted by the users.

A different and promising approach is *data obfuscation* [14,15]. The aim is not to reach anonymity, but rather to artificially reduce the precision of location data before disclosing it. In this way, the service can still be delivered, but an adversary cannot infer other sensitive information. We focus on obfuscation through *noise perturbation* [16,17]. An underrated problem in the literature is how to choose a suitable noise to effectively perturb data. We found that, if noise is not chosen properly, perturbation will not resist to attacks based on statistical inference. In particular, an obfuscation operator must offer a spacial *uniformity* of probability. Such a requirement is often postulated, rather than fulfilled, by state-of-the-art perturbation methods.

We propose UniLO, a location obfuscation operator able to guarantee uniformity even in the presence of imprecise location measurements. UniLO does not require a centralized and trusted obfuscator. We deal with service differentiation by proposing and comparing three UniLO-based obfuscation algorithms offering multiple contemporaneous levels of privacy. Finally, we experimentally prove that UniLO outperforms state-of-the-art perturbation algorithms both in terms of utility and resistance against inference attacks. This paper extends our previous work [18] with multiple levels of privacy and an in-depth analysis of the utility and the resistance against inference attacks. All the simulations scripts of the present paper can be downloaded from [19].

* Corresponding author.
E-mail addresses: pericle.perazzo@iet.unipi.it (P. Perazzo), gianluca.dini@iet.unipi.it (G. Dini).

http://dx.doi.org/10.1016/j.comcom.2015.02.014
0140-3664/© 2015 Published by Elsevier B.V.

The rest of the paper is organized as follows. Section 2 analyzes some related works and the differences with UniLO techniques. Section 3 introduces some basic concepts concerning the system model and the terminology. Section 4 formally describes the agnostic adversary model, the concept of uniformity, and a way to quantify it. Section 5 presents the basic UniLO operator and show its properties in terms of uniformity. Section 6 presents the problem of offering multiple levels of privacy and three algorithms to adapt UniLO in this sense. Section 7 evaluates UniLO algorithms in terms of utility on an example location-based service. Section 8 evaluates UniLO algorithms in terms of resistance against inference attacks. Finally, the paper is concluded in Section 9.

2. Related works

Approaches for location privacy can be roughly divided in *identity protection* and *data protection*. Identity protection avoids the re-identification of anonymous users. Data protection avoids the disclosure of precise locations.

2.1. Identity-protection approaches

Gruteser and Grunwald [11] first applied *k-anonymity* approach in location-based services. The proposed solution involves the subdivision of the map in quadrants with different granularities. The user does not release his precise position, but a quadrant of the grid containing other $k-1$ users, in such a way his identity is confused with theirs. The k-anonymity approach is broadly used in many research works [12,13,20–22]. However, these methods require the presence of $k-1$ users in the proximity, that could be missing, and a central anonymizer, that could not be fully trusted by the users. In addition, they do not permit the identification of the user, so that they are not applicable in those cases in which the user authenticates himself, e.g. payable services or location-based social networks. Our approach aims at protecting the position, rather than the identity, and it is suitable also for authenticated users.

Xu and Caie [23] and Abul et al. [24] approach the problem of *trajectory k-anonymity*, offering methods to protect user's privacy in continuous tracking systems. Although it could be extended in that sense, the present work focuses on single-position queries, as they encompass a wide range of location-based applications.

A problem complementary to anonymity is *pseudonym unlinkability* in tracking systems, usually approached with the technique of *mix zones* [25–27]. Mix zones are areas of the map where users cannot be tracked and change their pseudonym. By carefully placing and dimensioning such mix zones it is possible to thwart the adversary from linking two consecutive pseudonyms of the same user.

2.2. Data-protection approaches

Location obfuscation aims at reducing the precision of location data before disclosing it. This can be done by adding noise [14] (*noise-based obfuscation*) as well as with other methods, for example by replacing the exact position with a quadrant of a grid [15]. Research on this topic has focused mainly on what kind of service can be delivered with imprecise positions [15,28–30]. The problem of generating such imprecise positions in a proper way is often underrated. In particular, the uniformity of the noise-based obfuscation is often postulated, rather than evaluated. As a result, the proposed solutions turn out to be poorly resistant against inference attacks. In this paper we focus entirely on noise-based obfuscation, so from now on we will omit the "noise-based" specification as implicit.

Ardagna et al. [14] proposed a set of obfuscation operators that perturb the location: radius enlargement, radius restriction, center shift. These operators transform a measurement area into an obfuscated one. Our approach guarantees both more private and more useful obfuscated areas. More private because UniLO noise significantly increases the uniformity of the resultant privacy areas. More useful because we always guarantee that the privacy areas contain the user's position. A service provider can thus rely on more powerful assumptions and offer more quality of service. In addition, in [14] the resistance against attacks relies on the fact that the adversary is unaware of the privacy preference of the user. This could be an optimistic assumption, which features a form of "security by obscurity" that should be avoided [31].

Krumm [16] surveyed many different obfuscation methods and applied them to real-life GPS traces. The objective was to prevent an attacker from inferring users' home positions. Krumm tried also a noise-based method, which involved noise with a Gaussian magnitude. He found that this method requires a high quantity of noise ($\sigma = 5$ km) in order to effectively prevent inference attacks. Our approach offers higher levels of uniformity, and reduces the amount of noise needed to resist to inference attacks.

Dürr et al. [17] proposed an obfuscation approach with multiple levels of privacy. They build different "shares" which are random vectors concatenated to the user's position. They store the shares in different servers to avoid a single point of trust. Each service provider reconstructs the position by "fusing" one or more shares from one or more servers. The privacy level is proportional to the number of shares the service provider is allowed to access. The authors generate the shares as random vectors having uniform magnitude. Our obfuscation operators guarantee more resistance against inference attacks.

Inspired by differential privacy [32], Andrés et al. [33] introduced the concept of *ϵ-geo-indistinguishability*. The idea is that the user obtains more privacy in the surroundings of his true position, and less farther. To achieve this, they perturb the true position with a 2-dimensional extension of the Laplacian noise. Such a noise is highly non-uniform. As a consequence, geo-indistinguishability offers far less resistance to inference attacks compared to UniLO.

Other notable obfuscation-based approaches are [28–30]. All these works postulate uniformity rather than providing for it. In contrast, our approach offers guarantees on the obfuscation uniformity, even in presence of imprecise location measurements.

Another research track [34–38] applies *private information retrieval* (PIR) techniques to protect user's location. The objective is to provide a location-based service without disclosing the user's location at all. While PIR approaches offer strong and provable security, they are quite resource-demanding at the server side. Actually, they require complex, computational intensive cryptographic operations or the employment of trusted hardware architectures. In contrast obfuscation techniques only provide for statistical guarantees in terms of privacy, but they are more affordable for the service provider.

3. System model

In our system, a *user* is someone whose location is measured by a *sensor*. A *service provider* is an entity that receives the user's location in order to provide for a *location-based service*. The user applies an *obfuscation operator* to location information prior to releasing it to the service provider. The obfuscation operator purposefully reduces the precision to guarantee a certain privacy level. Such a precision is defined by the user and reflects his requirements in terms of privacy. The more privacy the user requires, the less precision the obfuscation operator returns.