

Toward an Easy Configuration of Location Privacy Protection Mechanisms

Sophie Cerf, Bogdan Robu, Nicolas Marchand, Antoine Boutet, Vincent Primault, Sonia Ben Mokhtar, Sara Bouchenak

▶ **To cite this version:**

Sophie Cerf, Bogdan Robu, Nicolas Marchand, Antoine Boutet, Vincent Primault, et al.. Toward an Easy Configuration of Location Privacy Protection Mechanisms. ACM/IFIP/USENIX Middleware conference, Dec 2016, Trente, Italy. hal-01376640

HAL Id: hal-01376640
https://hal.science/hal-01376640

Submitted on 3 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Poster Abstract: Toward an Easy Configuration of Location Privacy Protection Mechanisms

Sophie Cerf, Bogdan Robu, Nicolas Marchand
Univ. Grenoble-Alpes, CNRS, Gipsa-lab
{sophie.cerf,bogdan.robu, nicolas.marchand}@gipsa-lab.fr

Antoine Boutet, Vincent Primault, Sonia Ben Mokhtar, Sara Bouchenak
Université de Lyon, INSA-Lyon, LIRIS, CNRS
{antoine.boutet,vincent.primault,sonia.benmokhtar, sara.bouchenak}@insa-lyon.fr

ABSTRACT

The widespread adoption of Location-Based Services (LBSs) has come with controversy about privacy. While leveraging location information leads to improving services through geo-contextualization, it rises privacy concerns as new knowledge can be inferred from location records, such as home/work places, habits or religious beliefs. To overcome this problem, several Location Privacy Protection Mechanisms (LPPMs) have been proposed in the literature these last years. However, every mechanism comes with its own configuration parameters that directly impact the privacy guarantees and the resulting utility of protected data. In this context, it can be difficult for a non-expert system designer to choose appropriate configuration parameters to use according to the expected privacy and utility.

In this paper, we present a framework enabling the easy configuration of LPPMs. To achieve that, our framework performs an offline, in-depth automated analysis of LPPMs to provide the formal relationship between their configuration parameters and both privacy and the utility metrics. This framework is modular: by using different metrics, a system designer is able to fine-tune her LPPM according to her expected privacy and utility guarantees (i.e., the guarantee itself and the level of this guarantee). To illustrate the capability of our framework, we analyse Geo-Indistinguishability (a well known differentially private LPPM) and we provide the formal relationship between its ϵ configuration parameter and two privacy and utility metrics.

CCS Concepts

•**Security and privacy** → **Pseudonymity, anonymity and untraceability; Privacy-preserving protocols;** *File system security;*

1. INTRODUCTION

Location Based Services (LBSs) such as navigation or recommendation applications have been widely adopted by people using mobile devices. Although location-aware systems have greatly improved the quality of many services by introducing geo-contextualization, such systems rise important concerns about privacy.

Indeed, a collection of mobility traces (a set of timestamped locations reflecting the user's moving activity) can reveal many sensitive information about its user such as home and work places, favorite restaurants, religious leaning, or the individuals she met. To overcome this privacy issue, many Location Privacy Protection Mechanisms (LPPMs) have been proposed in the last decade. However, the effectiveness of these mechanisms usually rely on the tuning of a set of configuration parameters, often with a large range of possible values. In addition, this tuning depends on the expected privacy guarantee and utility of the resulting protected data, which form two conflicting assessment dimensions. Consequently, this parametrization task can be difficult for a non-expert system designer.

Few works have been done to automatically assess LPPMs in terms of privacy and utility as well as to assist a system designer in their configuration. Most of the existing LPPMs are not automated and focus on privacy guarantees [1, 3]. To the best of our knowledge, ALP [4] is the only solution enabeling to address this problem. Specifically, ALP uses a greedy solution to possibly make the configuration parameters converge to values which aim to maximize or minimize given privacy or utility metrics. As far as we know, no solution provides the formal relationship between configuration parameters and privacy and utility metrics as we propose.

In this paper, we present a framework which aims to help in the fine-tuning of LPPMs according to a set of expectations in term of privacy and utility defined by a system designer. To achieve that, our solution adopts an automated approach which performs an in-depth analysis of LPPMs, and provides the formal relationship between their configuration parameters and both privacy and utility metrics. This framework is modular: by using different metrics it is possible to adapt the provided model to specific privacy and utility guarantees.

Through a real example, we illustrate the capability of our framework to configure Geo-Indistinguishability [2], a well known LPPM based on differential privacy, according to privacy and utility metrics.

2. PROBLEM STATEMENT

To illustrate the problem related to the configuration of a LPPM and the capability of our framework to solve it, we consider the case of using Geo-indistinguishability [2] (GEO-

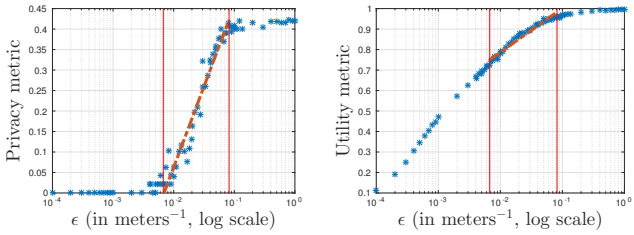
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Middleware Posters and Demos '16 December 12-16 2016, Trento, Italy

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4666-5/16/12.

DOI: http://dx.doi.org/10.1145/3007592.3007599



(a) Privacy (b) Utility

Figure 1: GEO-I configuration parameter ϵ according to privacy and utility metrics

I for short) to protect a whole dataset containing mobility traces of taxi drivers around San Franciso, with objectives in terms of privacy and utility.

GEO-I [2] follows the differential privacy model and is specifically designed to protect mobility data. To protect raw data, this LPPM adds random noise drawn from a Laplace distribution to the actual user location. GEO-I is parametrized by an ϵ parameter (expressed in meters^{-1}), quantifying the amount of noise to add (the lower the ϵ, the higher the noise).

As privacy objective, we consider the retrieval in the protected data of at most 10% of the Points of interest (POIs) of users (meaningful locations where a user made a significant stop). This objective is evaluated through a privacy metric which quantifies the proportion of actual POIs retrieved from the protected data for each user.

As utility objective, we consider maintaining a similar location precision at the scale of a city block. More precisely, the difference between the area coverage of users in the actual mobility traces and their protected counterpart is expected to remain about the size of a city block block and no less accurate. This objective is quantified through a utility metric.

The challenge here for a system designer is to correctly configure the ϵ of GEO-I to ensure that the expected objectives are respected. Figure 1 shows the level of privacy and utility according to different values of ϵ (the vertical lines shows zones where metrics are not saturated). Results show that the privacy metric rapidly changes from 0 to 0.4 with an ϵ from 0.007 to 0.08, respectively (Figure 1a). The utility metric, in turn, evolves more slowly on a larger range from 0.2 to 1 with an ϵ from 10^{-4} to 1, respectively (Figure 1b). Consequently, to ensure the expected privacy and utility objectives, the system designer has to choose carefully its ϵ. In our case, a $\epsilon = 0.01$ minimizes the privacy while maximizing the utility. In other words, with $l\epsilon = 0.01$ we ensure to a user that no more than 10% of her POIs can be retrieved while ensuring that 80% of her requests will concern the city block where she is.

3. OVERVIEW OF THE FRAMEWORK

Our framework proposes an automated approach to correctly configure LPPMs' parameters for given objectives in terms of privacy and utility, using a mathematical model of the LPPM behavior. Our solution proceeds in three automated steps:

1. First, the system needs to be defined: (1) the objective metrics for privacy (Pr) and utility (Ut), (2) the LPPM configuration parameters p_i and their range of values, and (3) the properties of the dataset d_i that are likely to influence privacy and utility metrics (i.e., reflecting impactfully characteristics of users such as the uniqueness). All these properties p_i, d_i are soundly chosen using a principal component analysis. For instance, in our illustration based on GEO-I, Pr measures the proportion of POIs that can be retrieved, Ut quantifies the difference in the service area coverage, only the parameter ϵ configures the LPPM, and no dataset properties is considered.
2. Then comes the modeling phase: experiments are automatically run where parameters p_i and d_i vary in turn while evaluation metrics are measured. Based on this data, a mathematical relationship between privacy and utility metrics, configuration parameters, and dataset properties is computed as an invertible function:
$$\begin{pmatrix} Pr \\ Ut \end{pmatrix} = f(p_1, \ldots, p_n, d_1, \ldots, d_m) \quad (1)$$
3. Finally, the LPPM configuration (i.e. the value of p_i) is computed by inverting the f function, using the specified privacy and utility objectives and the evaluation of dataset properties d_i.

For instance, in our illustration based on GEO-I, Pr and Ut only depend on parameter ϵ. If we focus only on the interval where ϵ impacts the privacy and utility metrics (i.e. between the vertical lines in Figure 1) we can approximate the experimental curves with the following linear equations:

$$log(\epsilon) = \frac{Pr - a}{b} = \frac{Ut - \alpha}{\beta}, \quad (2)$$
$$with\ a = 0.84, b = 0.17, \alpha = 1.21\ and\ \beta = 0.09.$$

Hence, to ensure that her privacy and utility objectives are met, a system designer can easily leverage this model to determine the appropriate value of ϵ. Specifically, to guarantee 10% privacy, configuring $\epsilon = 0.01$ ensures 80% utility.

4. CONCLUSION

In this paper we proposed a framework that enables easy and automated configuration of LPPMs according to privacy and utility objectives. Our preliminary results on the LPPM GEO-I show promising results as for the control of ϵ configuration parameter. Our future work will focus in testing other LPPMs and datasets, we also plan to extend our framework with more metrics and parameters.

5. REFERENCES

[1] B. Agir, T. Papaioannou, R. Narendula, K. Aberer, and J.-P. Hubaux. User-side adaptive protection of location privacy in participatory sensing. *GeoInformatica*, 18(1):165–191, 2014.

[2] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: Differential Privacy for Location-based Systems. In *CCS*, pages 901–914, 2013.

[3] K. Chatzikokolakis, C. Palamidessi, and M. Stronati. Constructing elastic distinguishability metrics for location privacy. In *PETS*, volume 2015, pages 156–170, 2015.

[4] V. Primault, A. Boutet, S. Ben Mokhtar, and L. Brunie. Adaptive location privacy with alp. In *SRDS*, 2016.