Optimal Sporadic Location Privacy Preserving Systems in Presence of Bandwidth Constraints

Michael Herrmann
KU Leuven ESAT/COSIC, iMinds
Leuven, Belgium
michael.herrmann@esat.kuleuven.be

Carmela Troncoso
Gradiant
Vigo, Spain
ctroncoso@gradiant.org

Claudia Diaz
KU Leuven ESAT/COSIC, iMinds
Leuven, Belgium
claudia.diaz@esat.kuleuven.be

Bart Preneel
KU Leuven ESAT/COSIC, iMinds
Leuven, Belgium
bart.preneel@esat.kuleuven.be

ABSTRACT

Various Location Privacy-Preserving Mechanisms (LPPMs) have been proposed in the literature to address the privacy risks derived from the exposure of user locations through the use of Location Based Services (LBSs). LPPMs obfuscate the locations disclosed to the LBS provider using a variety of strategies, which come at a cost either in terms of quality of service, or of resource consumption, or both. Shokri *et al.* propose an LPPM design framework that outputs optimal LPPM parameters considering a strategic adversary that knows the algorithm implemented by the LPPM, and has prior knowledge on the users' mobility profiles [23]. The framework allows users to set a constraint on the tolerable loss quality of service due to perturbations in the locations exposed by the LPPM. We observe that this constraint does not capture the fact that some LPPMs rely on techniques that augment the level of privacy by increasing resource consumption.

In this work we extend Shokri *et al.*'s framework to account for constraints on bandwidth consumption. This allows us to evaluate and compare LPPMs that generate dummies queries or that decrease the precision of the disclosed locations. We study the trilateral trade-off between privacy, quality of service, and bandwidth, using real mobility data. Our results show that dummy-based LPPMs offer the best protection for a given combination of quality and bandwidth constraints, and that, as soon as communication overhead is permitted, both dummy-based and precision-based LPPMs outperform LPPMs that only perturb the exposed locations. We also observe that the maximum value of privacy a user can enjoy can be reached by either sufficiently relaxing the quality loss or the bandwidth constraints, or by choosing an adequate combination of both constraints. Our results contribute to a better understanding of the effectiveness of location privacy protection strategies, and to the design of LPPMs with constrained resource consumption.

WPES'13, November 4, 2013, Berlin, Germany.

ACM 978-1-4503-2485-4/13/11 ...$15.00.
http://dx.doi.org/10.1145/2517840.2517853.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*Security and protection*; K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*

Keywords

Design; Security; Privacy

1. INTRODUCTION

Location Based Services (LBSs) enable users to, among others, let their friends know where they are, find nearby points of interest, or obtain contextual information about their surroundings. The typical LBS implementation is such that user locations are by default disclosed to the LBS provider. This raises privacy concerns, as location information is known to reveal potentially sensitive private information (e.g., visiting the mosque, church, or temple reveals religious beliefs). A variety of Location Privacy-Preserving Mechanisms (LPPMs), e.g., [6, 7, 17], have been proposed in prior research to mitigate these concerns. To do so, these mechanisms obfuscate user locations before sending them to the LBS provider.

The great majority of LPPMs in the literature are designed considering a non-strategic adversary. This assumes that the adversary is unaware of the LPPM obfuscation algorithm, and that he has no prior knowledge on the users' mobility profiles. However, both the LPPM's internal algorithm and the user mobility patterns leak information that can be exploited by the adversary to reduce her estimation error when locating users [21]. Hence, designs and evaluations that neglect such information overestimate the level of privacy protection offered by the LPPM.

Shokri *et al.* [23] proposed a framework to design LPPM with optimal parameters considering an adversary that has (and exploits) information on: i) the LPPM algorithm implemented; and ii) the mobility profile of the user. This framework facilitates the design of LPPMs that maximize the location estimation error of strategic adversaries. Furthermore, the framework allows users to establish a maximum tolerated quality of service loss stemming from the use

of the LPPM. The framework is suitable to model LBSs in which users only reveal their location *sporadically*, i.e. subsequent location exposures of the same user are assumed to be sufficiently apart in time that it is not possible to link them as related to the same individual. Examples of applications in which location revelations are sporadic include check-in services [1], or services for finding nearby points of interest [2].

The problem statement in Shokri's framework [23] does not consider constraints on resources utilization (e.g., bandwidth, battery consumption). These are however likely to be a concern for users in reality, since LBSs are mostly accessed from resource-constrained mobile devices. Our first contribution is to extend the framework to account for resource limitations.

Prior research has only applied the framework to the design of perturbation-based mechanisms, i.e., LPPMs that modify the location that is disclosed to the LBS provider. As second contribution, we model two other popular privacy-preserving strategies in the context of the framework. Both types of mechanisms increase the adversary's uncertainty on the user's actual position by raising the number of locations from which the user could have issued a query. In dummy-based mechanisms [14, 16, 26] the LPPM sends fake locations to the LBS server along with the actual user requests. In precision-based mechanisms [9, 11, 25] the LLPM decreases the precision of the disclosed location sent to the LBS provider, so that there is a bigger geographical region in which they user might be located.

Contrary to the perturbation-based LPPMs considered by Shokri *et al.* [23], dummy-based and precision-based LPPMs may consume more resources (e.g., bandwidth and battery) in order to conceal the user's location. Our third contribution is a study of the trilateral trade-off between quality of service, bandwidth consumption, and privacy using these LPPMs as case study. We find that for the considered LPPMs both quality loss and bandwidth constraints can be traded for privacy. In fact, the maximum achievable level of privacy can be reached either when the quality loss constraint is sufficiently loose (as in [23]), when sufficient bandwidth is allowed, or when an adequate combination of both is allowed. Our simulations show that, for given bandwidth and quality constraints, dummy-based LPPMs offer better protection than precision-based LPPMs. This is because dummy-based LPPMs have more degrees of freedom than precision-based LPPMs in choosing the cover locations, and hence can better exploit the available resources. Finally, both dummy-based and precision-based offer a better privacy level than just perturbation for the same quality of service, provided that the system can tolerate the introduction of a communication overhead.

The rest of this paper is organized as follows: the next section gives an overview of the state of the art in location privacy-preserving systems design. Section 3 describes the system and adversarial models, as well as the constraints imposed on the design; and Section 4 revisits Shokri *et al.*'s framework. We describe the linear programs to compute different classes of resource-consuming LPPMs in Section 5, and validate them against real data in Section 6. Finally, we conclude in Section 7.

2. RELATED WORK

It is widely accepted that the disclosure of location data entails a privacy risk: Hoh et al. show that car driving traces enable the inference of the drivers' home addresses [13]; this information by itself, or together with the driver's work place, can be used to re-identify anonymous traces [15, 10]. Furthermore, Freudiger *et al.* point out that a people's mobility patterns are persistent and unique [8]. Therefore, users are identifiable by the LBS even if they only share their location during a short period of time. Once location data is identifiable, it may reveal a detailed picture of the person's habits, lifestyle, and preferences [3]. To counter this threat various obfuscation-based Location Privacy Preserving Mechanisms (LPPMs) been proposed in the literature. These mechanisms obfuscate the revealed locations and thus prevent (or at least limit) the possible inferences that could be made from the data.

Following the categorization proposed by Shokri *et al.* [21] we briefly introduce the existing obfuscation strategies and refer the reader to [20] for a more detailed review. *Perturbation-based* LPPMs [12, 17] modify a user's reported location such that at least two users might be associated to a location. *Pseudonymization-based* LPPMs regularly change the identity with which users identify themselves to the LBS provider, in order to prevent the linkage of two subsequent user locations, thus preventing the adversary from reconstructing the trajectories followed by the users of the system. These LPPMs can be combined with *hiding-based* LPPMs, which allow users to sometimes hide their location [5], further decreasing the adversary's capability to link location exposures. *Precision-based* LPPMs [4, 9, 11, 25] reduce the granularity of the location data revealed to the provider, so that it is not possible to pinpoint the exact location of a user within a geographical region. Finally, *dummy-based* LPPMs [14, 16, 26] automatically generate queries with fake position data that are indistinguishable from the users' real queries. Here the adversary is unable to determine whether the location associated with a query corresponds to the user's actual position, or is a decoy.

Shokri *et al.* have proposed methods to quantify and systematically evaluate the level of privacy provided by obfuscation-based LPPMs [21, 22]. They formalize the obfuscation process performed by the LPPM, as well as the attack strategies that an adversary can use to invert the location transformations made by the LPPM. They measure privacy as the expected error of a strategic adversary when estimating the actual location of a user. This quantitative approach is a cornerstone of their LPPM design framework, where they propose a systematic method to design LPPMs that are optimal with respect to strategic adversaries, who are aware of the LPPM's internal operation and the users' mobility profiles [23].

This framework allows users to indicate the maximum quality loss (derived from the use of the LPPM) that they are willing to tolerate. The design framework then outputs a set of parameters for the LPPM that maximize the error of the adversary when attempting to locate users. Our work builds on this framework and extends it to account for not only quality loss, but also for limitations on bandwidth consumption.

Finally, we note that there are other approaches to building location privacy systems that are not based on obfuscation strategies and are thus out of the scope of this paper.

This includes cryptographic approaches such as those based on Private Information Retrieval protocols [18].

3. SYSTEM MODEL

In this paper, we extend the framework by Shokri *et al.* [23] to account for bandwidth constraints in Location Privacy Preserving Mechanisms (LPPMs). Therefore, we follow the framework's system model and definitions and augment them when needed to account for bandwidth constraints. The focus of the framework is on user-centric mechanisms, in which the configuration of the LPPM is decided on independently by each user, without knowledge about other users in the system. Thus, without loss of generality, we restrict our model and analysis to a single user. We note that cloaking mechanisms, in which the geographical region disclosed is chosen taking into account the positioning of a set of users [11], can also be modeled as user-centric mechanisms because their privacy guarantees depend only on the size of the region [24].

User model: Similarly to prior work [23] we consider that the user moves around in a finite geographical area that is divided into M discrete regions $\mathcal{R} = \{r_1, r_2, \ldots, r_M\}$. Users only expose their location $r \in \mathcal{R}$ *sporadically* to an LBS provider in order to obtain a service. A user's LBS usage pattern is described by her *mobility profile* $\psi(r)$, $\sum_r \psi(r) = 1$, a probability distribution describing her likelihood of being at location r when querying the LBS. We make no particular assumption on the users' mobility patterns, i.e., we impose no restrictions on the profiles $\psi(r)$. As usage is sporadic, the locations from which the user accesses the service at different time instants are independent from each other. Therefore, the mobility profiles only describe the frequency with which users' visit locations, and does not contain information about transitions between regions.

Location privacy-preserving mechanism: The user runs in her personal device an LPPM that transforms her real location $r \in \mathcal{R}$ into a pseudo-location $r' \in \mathcal{R}'$. This transformation is made according to a probability distribution $f(r'|r) = \Pr(r'|r)$. The pseudo-location r' is exposed to the LBS provider instead of her actual location r. Shokri *et al.* [23] consider that $\mathcal{R}' = \mathcal{R}$. In this work we extend $\mathcal{R}'$ to be the powerset of $\mathcal{R}$ except the empty set; i.e., $\mathcal{R}' = \mathcal{P}(\mathcal{R}) - \{\emptyset\}$. Hence, i) pseudo-locations r' may or may not contain the real location r; and ii) differently from prior work [23], in which pseudo-locations r' are formed by one region in $\mathcal{R}$, here r' may be formed by one or more regions r_i in $\mathcal{R}$.

Adversary model: We consider that the user wants to protect her real location towards a passive adversary that has access to the locations exposed to the LBS. This adversary could be the LBS provider, an eavesdropper of the user-provider communication, or other LBS subscribers with which exposed locations are shared. We assume that the adversary knows the users' profiles $\psi(r)$, which can be inferred, for instance, using existing learning techniques [21].

Following prior work [23] we model the adversary's strategy as a probability distribution $h(\hat{r}|r') = \Pr(\hat{r}|r')$. This distribution describes the probability that, given an exposed location r', the estimated location $\hat{r}$ corresponds to the user's real position r. We measure the privacy loss as the adversary's expected error in this estimation $\hat{r}$ given that the real location is r. We model the adversarial error as a function $d_p(\hat{r}, r)$ that depends on both the user's privacy criteria and on the semantics of the location [23]; as well as on the transformation function $f(r'|r)$ implemented by the LPPM. (We provide examples of functions $d_p(\cdot)$ that are adequate for particular LPPMs in Section 5.)

Quality of service: Users expect to obtain relevant information from their queries to the LBS. Because the response of the LBS to a query depends on the observed location r', and not on the real location r, the information contained in the response may be of less utility to the user than that contained in a response to a query in which r is exposed. Given an LPPM $f(\cdot)$, the *expected quality loss* suffered by the user can be computed as:

$$\mathrm{E}[\mathrm{Q}_{\mathrm{loss}}(\psi, f, d_q)] = \sum_{r,r'} \psi(r) f(r|r') d_q(r', r)\,. \qquad (1)$$

In this formula $\psi(r)$ represents the prior probability of the user accessing the LBS from location r (i.e., according to her mobility profile); $f(r'|r)$ represents the probability of exposing r' given that the user is at r; and the function $d_q(r', r)$ represents the quality loss resulting from exposing r' instead of r to the LBS provider. (We provide examples of $d_q(\cdot)$ functions adequate for particular LPPMs in Section 5.) In layman words, $\mathrm{E}[\mathrm{Q}_{\mathrm{loss}}(\psi, f, d_q)]$ reflects the average discontent experienced by users when utilizing an LPPM.

We assume that the user imposes a maximum tolerable service quality loss $\mathrm{Q}_{\mathrm{loss}}^{\max}$. The LPPM output must satisfy the constraint $\mathrm{E}[\mathrm{Q}_{\mathrm{loss}}(\psi, f, d_q)] < \mathrm{Q}_{\mathrm{loss}}^{\max}$.

Bandwidth constraints: The fact that Shokri *et al.* consider $\mathcal{R}' = \mathcal{R}$ implies that the LPPM never incurs in communication overhead when sending r' instead of r. Since we have set $\mathcal{R}' = \mathcal{P}(\mathcal{R}) - \{\emptyset\}$, sending r' may require more bandwidth than sending r (e.g., if r' is composed by several regions in $\mathcal{R}$). LBSs are mostly accessed from mobile devices which in general have restricted connectivity and limited resources, and hence users may want to limit the overhead introduced by the LPPM. We extend the existing model [23] to account for this fact by defining the *expected bandwidth overhead* incurred by LPPM $f(\cdot)$ as:

$$\mathrm{B}_{\mathrm{cost}}(\psi, f, d_b) = \sum_{r,r'} \psi(r) f(r|r') d_b(r', r)\,, \qquad (2)$$

In this formula $\psi(r)$ and $f(r'|r)$ have the same role as in Eq. (1). The function $d_b(r', r)$ represents the additional cost in terms of bandwidth derived from exposing r' instead of r. (We provide examples of $d_b(\cdot)$ functions adequate for particular LPPMs in Section 5.)

We assume that the user imposes a maximum tolerable bandwidth $\mathrm{B}_{\mathrm{cost}}^{\max}$. As with quality loss constraints, the LPPM must satisfy $\mathrm{B}_{\mathrm{cost}} < \mathrm{B}_{\mathrm{cost}}^{\max}$.

We note that, although we only consider limitations on communication overhead, the function $d_b(\cdot)$ can model other constraints related to resource consumption resulting from exposed pseudo-locations that may be formed by several regions, e.g., the increase in battery consumption needed to send more packets, or to process more responses.

Privacy: The level of privacy enjoyed by users depends on the attack strategy deployed by the adversary. Following the definition by Shokri *et al.* [21, 23] we measure the *expected privacy* of the user as:

$$\mathrm{Privacy}(\psi, f, h, d_p) = \sum_{r,r',\hat{r}} \psi(r) f(r'|r) h(\hat{r}|r') d_p(\hat{r}, r)\,. \quad (3)$$

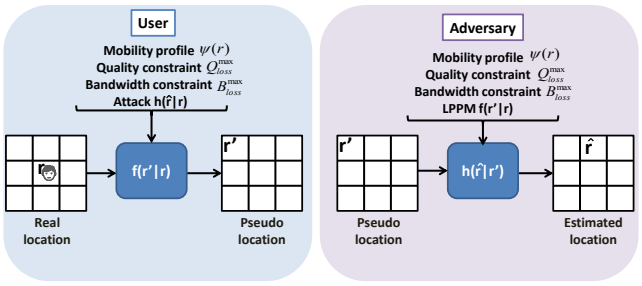


Figure 1: System model.

Each summand in this equation represents the probability that the user obtains a privacy level $d_p(\hat{r}, r)$, when she accesses the LBS from location location r, exposes pseudo-location r', and the adversary estimates $\hat{r}$ given the observation.

Figure 1 illustrates the relationships between the different elements of this model. Note that we consider that the defense (resp., the attack) takes into account the attack (resp., the defense) implemented by the adversary (resp., the user), as well as the user's mobility profile and her constraints in terms of bandwidth and quality of service.

4. A GAME THEORETIC APPROACH TO LOCATION PRIVACY

In this section we revisit the design methodology proposed by Shokri *et al.* in prior work [23]. This method allows the user to choose optimal parameters for the LPPM $f(\cdot)$, given an adversary that implements the optimal attack $h(\cdot)$ against this defense. Given a user mobility profile $\psi(r)$ and quality of service constraint $\mathrm{Q}_{\mathrm{loss}}^{\max}$, the method models the design of the optimal LPPM as an instance of a zero-sum Bayesian Stackelberg game.

The Stackelberg competition in the context of location privacy is stated as follows: a *leader* (the user), commits first to an LPPM $f(\cdot)$ that satisfies the quality constraint $\mathrm{Q}_{\mathrm{loss}}^{\max}$. For this purpose the LPPM takes the user's actual location r as input, and outputs a pseudo-location r'. Upon observing the exposed location, a *follower* (the adversary), estimates the real location through the attack $h(\cdot)$, taking into account both the user's profile $\psi(r)$ and the LPPM $f(\cdot)$ chosen by the user. The adversary 'pays' an amount $d_p(\hat{r}, r)$ to the user that represents the estimation error from the adversary's perspective, and the location privacy gain from the user's perspective.

Both players aim at maximizing their payoffs: the adversary tries to minimize the amount to pay (i.e., minimize her estimation error), while the user tries to maximize this amount (i.e., maximize her location privacy). The game is zero-sum, as the adversary's information gain equals the privacy lost by the user, and vice-versa. It is also a Bayesian game since the adversary only has access to probabilistic data about the user's real location; i.e., her information on the user is incomplete.

4.1 Perturbation-based LPPM

Shokri *et al.* validate their framework by applying it to the design of perturbation-based strategies. In this scenario $\mathcal{R}' = \mathcal{R}$, and hence the pseudo-locations r' output by the

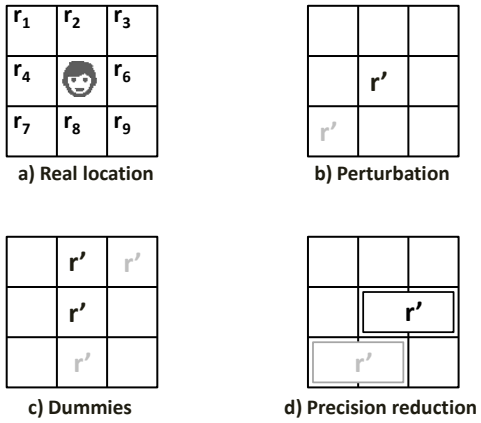


Figure 2: Toy example ($\mathcal{R} = \{r_1, \ldots, r_9\}$): a) Real user location; b) Perturbation-based LPPM $\mathcal{R}' = \mathcal{R}$; c) Dummy-based LPPM; d) Reducing precision-based LPPM.

LPPM are formed by one region $r_i \in \mathcal{R}$, which may or may not be equal to the real location r. Let us consider the toy example in Fig. 2a, in which the area $\mathcal{R}$ is formed by 9 regions, and where the user queries the LBS provider from location r_5. Two possible pseudo-locations r' are shown in Fig. 2b, depicted in black and gray. Note that the black **r'** coincides with the real user location $r = r_5$, while the grey pseudo-location r'$= r_7$ does not.

Solution: We now present the linear programs developed in prior work [23] to compute the optimal perturbation and attack strategies $f(\cdot)$ and $h(\cdot)$. These linear programs compute the theoretic equilibrium of the game described above.

The user runs the following linear program to find the optimal parameters for her perturbation-based LPPM:

$$\textbf{Choose } f(r'|r), x_{r'}, \forall r, r' \text{ that}$$

$$\textbf{maximize} \sum_{r'} x_{r'} \tag{4}$$

$$\textbf{subject to}$$

$$x_{r'} \leq \sum_{r} \psi(r) f(r'|r) d_p(\hat{r}, r), \ \forall \hat{r}, r' \tag{5}$$

$$\sum_{r} \psi(r) \sum_{r'} f(r'|r) d_q(r', r) \leq \mathrm{Q}_{\mathrm{loss}}^{\max} \tag{6}$$

$$\sum_{r'} f(r'|r) = 1, \ \forall r \tag{7}$$

$$f(r'|r) \geq 0, \ \forall r, r' \tag{8}$$

The decision variable $f(r'|r)$ represents the LPPM algorithm, while $x_{r'}$ represents the expected privacy of the user (see Appendix A). The inequalities defined by Eq. (5) express the privacy constraint, ensuring that $f(r'|r)$ is chosen to maximize $x_{r'}$; while the inequalities defined by Eq. (6) express the quality constraint, ensuring that the expected quality of service loss is at most $\mathrm{Q}_{\mathrm{loss}}^{\max}$. Finally Eq. (7) and (8) ensure that $f(\cdot)$ is a proper probability distribution.

On the other hand, the adversary runs the following linear program to obtain the optimal attack function $h(\hat{r}|r')$, which

minimizes privacy when the user implements a perturbation-based LPPM $f(r'|r)$:

$$\textbf{Choose } h(\hat{r}|r'), y_r, \forall r, r', \hat{r}, \text{and} z_q \in [0, \infty) \text{ that}$$

$$\textbf{minimize } \sum_r \psi(r)\, y_r + z_q \mathrm{Q}_{\text{loss}}^{\max} \tag{9}$$

subject to

$$y_r \geq \sum_{\hat{r}} h(\hat{r}|r') d_p(\hat{r}, r) + z_q d_q(r', r), \forall r, r' \tag{10}$$

$$\sum_{\hat{r}} h(\hat{r}|r') = 1, \forall r' \tag{11}$$

$$h(\hat{r}|r') \geq 0, \forall r', \hat{r} \tag{12}$$

$$z_q \geq 0 \tag{13}$$

The decision variable $h(\hat{r}|r')$ represents the adversary's attack strategy on the LPPM algorithm, and y_r the expected privacy of the user (see Appendix A). The variable z_q acts as *shadow price* for the quality. It expresses the loss (gain) in privacy when the maximum tolerated expected quality loss $\mathrm{Q}_{\text{loss}}^{\max}$ decreases (increases) by one unit. We refer the reader to Shokri's prior work for more details on the meaning of this variable [23]. The inequalities defined by Eq. (10) represent constraints on privacy, ensuring that $h(\hat{r}|r')$ is chosen to minimize privacy given the quality constraints; and Eqs (11) and (12) ensure that $h(\cdot)$ is a proper probability distribution. Finally Eq. (13) ensures that the trade-off between quality and privacy expressed by z_q is non-negative.

Quality, Bandwidth, and privacy constraints: Perturbation-based LPPMs output one-sized regions $r' \in \mathcal{R}' = \mathcal{R}$. This determines the functions used to model the constraints imposed by the user. Since pseudo-locations and real locations have the same size, there is no communication overhead in the model. Therefore, the bandwidth constraint $\mathrm{B}_{\text{cost}}^{\max}$ does not affect the optimization and does not appear in the linear programs.

Furthermore, in this setting both the quality and the privacy constraints can be expressed in terms of the distance between the exposed location r' (resp., the inferred location $\hat{r}$) and the actual user location r [23]. For the sake of simplicity, in our experiments for perturbation-based LPPMs we model both $d_q(r', r)$ and $d_p(\hat{r}, r)$ as the Manhattan distance between the two locations (e.g., $d_p(\hat{r}, r) = \|\hat{r} - r\|_1$).

5. BANDWIDTH-CONSUMING LOCATION PRIVACY PRESERVING MECHANISMS

In this section we model two popular families of Location Privacy-Preserving Mechanisms (LPPMs) in the literature that consume extra bandwidth to increase users privacy: dummy-based LPPMs, and precision-based LPPMs. To model these strategies we extend the game-theoretic approach outlined in the previous section to also account for bandwidth constraints. We describe two linear programs that output the user's optimal LPPM $f(\cdot)$ and the adversary's optimal attack $h(\cdot)$, while respecting the quality and bandwidth constraints.

5.1 Dummy-based LPPM

Dummy-based LPPMs automatically generate dummy queries that are sent to the LBS provider along with the user's real queries [14, 16, 26]). The dummy queries contain fake locations and their goal is to increase the adversary's estimation error on the user's real location, since for the adversary all received locations are equally likely to correspond to the user's actual position.

A dummy-based LPPM $f(r'|r)$ outputs pseudo-locations r' from $\mathcal{R}' = \mathcal{P}(\mathcal{R}') - \{\emptyset\}$ formed by one or more *non-contiguous* regions $r_i \in \mathcal{R}$, which may or may not contain the real location r. In the toy example shown in Fig. 2c we can see two possible outputs r' when the user sends one dummy query formed by two regions. The black pseudo-location **r'**$= \{r_2, r_5\}$ contains the real location $r = r_5$, while the grey pseudolocation r'$= \{r_3, r_8\}$ does not. In the latter case the LPPM no only generates decoy locations, but also perturbs the user's position.

Solution: The linear program to compute the optimal dummy-based LPPM is similar to the perturbation-based case, with one important difference: it includes a set of inequalities to ensure that the expected communication overhead associated to the use of dummies does not exceed the maximum expected bandwidth consumption $\mathrm{B}_{\text{cost}}^{\max}$:

$$\textbf{Choose } f(r'|r), x_{r'}, \forall r, r' \text{ that}$$

$$\textbf{maximize } \sum_{r'} x_{r'} \tag{14}$$

subject to

$$x_{r'} \leq \sum_r \psi(r) f(r'|r) d_p(\hat{r}, r), \forall \hat{r}, r' \tag{15}$$

$$\sum_r \psi(r) \sum_{r'} f(r'|r) d_q(r', r) \leq \mathrm{Q}_{\text{loss}}^{\max} \tag{16}$$

$$\sum_r \psi(r) \sum_{r'} f(r'|r) d_b(r', r) \leq \mathrm{B}_{\text{cost}}^{\max} \tag{17}$$

$$\sum_{r'} f(r'|r) = 1, \forall r \tag{18}$$

$$f(r'|r) \geq 0, \forall r, r' \tag{19}$$

The new inequality (17) adds the bandwidth constraint, so that the expected bandwidth consumption does not exceed $\mathrm{B}_{\text{cost}}^{\max}$.

From the adversary's point of view, the linear program used to compute the optimal attack $h(\hat{r}|r)$ differs from the perturbation-based case in that we introduce a new shadow price z_b in Eq. (25). This new variable models the relation between privacy and bandwidth in the same manner as z_q models the relation between privacy and quality. We obtain:

$$\textbf{Choose } h(\hat{r}|r'), y_r, \forall r, r', \hat{r}, z_q \in [0, \infty), z_b \in [0, \infty) \text{ to}$$

$$\textbf{minimize } \sum_r \psi(r)\, y_r + z_q \mathrm{Q}_{\text{loss}}^{\max} + z_b \mathrm{B}_{\text{cost}}^{\max} \tag{20}$$

subject to

$$y_r \geq \sum_{\hat{r}} h(\hat{r}|r') d_p(\hat{r}, r) + z_q d_q(r', r) + z_b d_b(r', r), \forall r, r' \tag{21}$$

$$\sum_{\hat{r}} h(\hat{r}|r') = 1, \forall r' \tag{22}$$

$$h(\hat{r}|r') \geq 0, \forall r', \hat{r} \tag{23}$$

$$z_q \geq 0 \tag{24}$$

$$z_b \geq 0 \tag{25}$$

Quality, bandwidth, and privacy constraints: As the dummy-based LPPM transmits dummy locations to the LBS provider, the functions $d_q(r', r)$ and $d_b(r', r)$, which express the constraints on quality and bandwidth, need to take into account that pseudo-locations r' can be composed by several regions.

With respect to the quality of service function $d_q(r', r)$ we distinguish two cases. If the actual location r is among the regions contained in the pseudo-location r', then the quality loss is zero, as the user receives a response corresponding to her real location r. Formally, $d_q(r', r) = 0, \forall r' : r \in r'$. If on the other hand the real location is not within the exposed pseudo-location, we assume that the response for the nearest location will provide the most useful response to the user, and thus measure the quality loss as the minimum of the distances between the real location and each of the locations r_i contained in r'. For instance, considering the Manhattan distance, $d_q(r', r) = \min_{r_i \in r'} \|r_i - r\|_1, \forall r' : r \notin r'$.

The bandwidth function $d_b(r', r)$ takes into account that the system sends and receives more traffic when dummies are implemented. This extra bandwidth consumption may be due to an increase in the length of the query if all dummies are sent in one request; or to an increase in the number of queries if dummies are sent in separate requests. In this paper we consider that each dummy increases the bandwidth overhead by 2 units: one unit for uploading and one unit for downloading. Formally: $d_b(r', r) = (\sum_{r_i \in r'} 2) - 2$.

As in the perturbation-based case, the privacy function $d_p(\hat{r}, r)$ considers the locations $\hat{r}, r \in \mathcal{R}$ and hence this function does not need to be modified.

5.2 Precision-based LPPM

Precision-based LPPMs reduce the precision of the location exposed by disclosing a larger region [9, 11, 25]. This makes it hard for the adversary to pinpoint the exact location of the user. As in the previous case, the LPPM $f(r'|r)$ outputs pseudo-locations r' from $\mathcal{R}' = \mathcal{P}(\mathcal{R}') - \{\emptyset\}$, but in this case r' is formed by a set of one or more *contiguous* regions $r_i \in \mathcal{R}$ that may or may not contain the real location r. The locations contained in r' form the region that is sent to the LBS provider. In the toy example shown in Fig. 2d, we can see two possible outputs r' when the precision is halved by exposing two regions. The black pseudo-location **r'**$= \{r_5\} \bigcup \{r_6\}$ contains the real location $r = r_5$, while the grey pseudo-location r'$= \{r_7\} \bigcup \{r_8\}$ does not. In the latter case the LPPM no only exposes decoy locations, but also perturbs the user's position.

Solution: The bandwidth consumed by a precision-based LPPM strongly depends on the type of information required by the LBS. Let us consider an LBS that returns nearby points of interest. When the user issues a request for a large pseudo-location r' (i.e., with reduced precision), the response contains more points than when the pseudo-location is small, requiring more bandwidth. This is similar to the dummy-based case but has different quality loss and communication overhead, as explained below. Hence, the optimal defense can be computed using the appropriate functions $d_q(\cdot)$ and $d_b(\cdot)$ in the linear program (Eqs (14)-(19)). We refer to this type of systems as *nearby precision-based LPPMs*.

Now consider an LBS in which the provider returns the value of interest (e.g., traffic congestion) for a representative location within r'. In this case the LBS response contains just one value independently of the size of the region, and hence diminishing the precision does not increase the bandwidth consumption. This is similar to the perturbation-based case, where there LPPM does not incur in a communication overhead, but has different quality loss as explained below. The optimal LPPM parameters can be computed using the appropriate function $d_q(\cdot)$ in the linear program (Eqs (4)-(8)). We denote these systems as *aggregated precision-based LPPMs*.

Quality, Bandwidth, and privacy constraints: The quality loss introduced by precision-based LPPMs depends on the type of system. For nearby precision-based LPPMs there is no quality loss when the user's actual location r is included in r', because the response includes the points of interest nearest to this location, and thus $d_q(r', r) = 0, \forall r' : r \in r'$. Otherwise, we measure the quality loss as the minimum distance between the user location r and the locations contained in r' ($d_q(r', r) = \min_{r_i \in r'} \|r_i - r\|_1, \forall r' : r \notin r'$). For aggregated precision-based LPPMs, in which the response is one representative value, larger regions r' reduce the expected quality of service. In our experiments we measure the quality loss as the average distance from the user location r to the regions $r_i \in \mathcal{R}$ in r', i.e., $d_q(r', r) = \sum_{r_i \in r'} \|r_i - r\|_1 / N$, being N the number of regions in r'.

The bandwidth consumption only increases for nearby precision-based LPPMs. We define the function $d_b(\cdot)$, that describes the communication cost, as $d_b(r', r) = (\sum_{r_i \in r'} 1) - 1$, and add one unit of bandwidth for each extra region r_i included in r'.

The estimation of the adversary is a location $\hat{r} \in \mathcal{R}$, and thus the privacy constraint does not need to be modified.

6. EVALUATION

The linear programs presented in the previous section output optimal LPPM parameters. In this section we evaluate the trade-off between location privacy, service quality, and communication overhead in different types of LPPMs. For this purpose we measure the expected Privacy(ψ, f, h, d_p) offered by an LPPM for a given mobility profile $\psi(r)$, using different combinations of maximum tolerable expected quality loss $\mathrm{Q}_{\mathrm{loss}}^{\mathrm{max}}$ and expected bandwidth consumption $\mathrm{B}_{\mathrm{cost}}^{\mathrm{max}}$. These constraints are modeled depending on the strategy followed by the LPPMs as described in Sections 4.1, 5.1, and 5.2. For precision-based LPPMs we distinguish between *nearby precision-based LPPMs*, which incur in communication overhead but no quality loss; and *aggregated precision-based LPPMs*, which do not consume extra bandwidth but reduce the quality of service.

Existing Dummy-based LPPMs [14, 16, 26]: In these schemes the LPPM algorithm selects a fixed number of requests b_u containing dummy locations. These dummy locations, which are sent to the LBS provider along with the real request, are chosen depending on the user's mobility profile. The real location may be perturbed or not. We model existing dummy-based LPPMs as follows: the user sets a value for the bandwidth consumption b_u that establishes the allowed communication overhead. Then r' is chosen according to the user's mobility profile from all possible pseudo-locations that contain b_u dummies. We note that, in some proposed systems, dummies are chosen also depending on previous exposures in order to resemble realistic movements. However, since we limit our analysis to sporadic LBSs, in which the locations from which the user makes subsequent requests

are not correlated, we do not consider past exposures when selecting dummy locations.

Existing Precision-based LPPMs [9, 11, 25]: In these schemes the user sets a parameter that defines the precision of the exposed location. The real location may be perturbed or not. We model existing precision-based LPPMs as follows: Given that the user chooses a maximal precision reduction s_u, the LPPM selects r' from all pseudo-locations containing s_u contiguous regions $r_i \in \mathcal{R}$, such that the following condition holds: $\forall r_i \in r' : \|r_i - r\|_1 \leq s_u$, considering the Manhattan distance as quality loss function.

Existing attacks: Similarly to prior work [23] we evaluate LPPMs with respect to Bayesian inference attacks [22]. This attack inverts the algorithm implemented by the LPPM using the posterior probability distribution over all locations given the user's profile.

Optimal attacks: We also evaluate the different LPPMs against optimal attacks. We test the performance of the optimal LPPM towards the optimal attack output by the framework; and the performance of existing defenses against the optimal attack against described in prior work which we repeat here for convenience [23]:

$$\textbf{Minimize} \sum_{\hat{r},r',r} \psi(r) f(r'|r) h(\hat{r}|r') d_p(\hat{r}, r) \quad (26)$$

$$\textbf{subject to} \sum_{\hat{r}} h(\hat{r}|r') = 1, \forall r', \text{and } h(\hat{r}|r') \geq 0, \forall \hat{r}, r' \quad (27)$$

6.1 Experimental setup

We use real mobility profiles obtained from the CRAWDAD dataset epfl/mobility [19] to evaluate the LPPMs' performance. This dataset contains GPS coordinates of approximately 500 taxis collected over 30 days in the San Francisco Bay Area.

The level of privacy offered by the LPPMs depends on the size of the area of interest, as well as on the number of regions M in which the area is divided. These parameters define the size of the regions r_i, and hence influence the accuracy with which the adversary estimates the user location. When the choice of parameters results in small regions r_i, the adversary can locate the user with more precision than when regions are big (e.g., a large region of interest divided in few regions). In the following we justify our choices for the size of the area of interest and the number of regions used in our experiments.

Number of regions. The number of regions has a strong impact on the running time of the optimization because the number of possible real locations, pseudo-locations, and estimated locations define the number of inequalities involved in the linear programs. In our evaluation we need to run a large number of linear programs to test a significant sample of quality/bandwidth constraint combinations. Hence, we need to choose an appropriate number of regions in the area of interest to be able to run our experiments in reasonable time.

Let us consider that the area of interest is divided with a grid of $M = \alpha \times \beta$ regions, with no particular restriction on the regions' shape or size. In the strategies considered in this paper, the number of real and estimated locations (r and $\hat{r}$) is the same, and equal to the cardinality of $\mathcal{R}$, i.e., $M = \text{card}(\mathcal{R}) = \alpha \cdot \beta$. However, the number of possible pseudo-locations depends on the strategy implemented by the LPPM. The perturbation-based LPPM transforms real locations into one-region pseudo-locations, hence $\text{card}(\mathcal{R}') = \text{card}(\mathcal{R})$. The dummy-based strategy allows pseudo-locations to contain any combination of non-contiguous locations, and we can compute the number of possibilities for r' as $\text{card}(\mathcal{R}') = \sum_{i=1}^{M} \binom{M}{i}$. Finally, in the precision-based mechanisms pseudo-locations contain combinations of contiguous locations. For simplicity in our experiments for precision-based LPPMs we limit $\mathcal{R}'$ to rectangular pseudo-locations (this would make the pseudo-location $r' = \{r_4\} \bigcup \{r_7\} \bigcup \{r_8\}$ in Figure 2 ineligible). Therefore, the number of pseudo-locations is $\text{card}(\mathcal{R}') = \sum_{i=0}^{\alpha-1} \sum_{j=0}^{\beta-1} (\alpha - i)(\beta - j)$.

Table 1: Performance times for different grid sizes

Grid size	Mean	Std	% finished
Perturbation-based			
2x2	0.22s (0.00 h)	0.26s	100.00
3x3	0.28s (0.00 h)	0.36s	100.00
4x4	0.39s (0.00 h)	0.34s	100.00
5x5	2.30s (0.00 h)	0.64s	100.00
6x6	16.21s (0.00 h)	5.20s	100.00
7x7	211.42s (0.06 h)	128.48s	100.00
8x8	679.58s (0.19 h)	336.75s	100.00
9x9	3437.09s (0.95 h)	1450.49s	100.00
10x10	13199.39s (3.67 h)	6660.02s	100.00
Dummy-based			
Grid size	**Mean**	**Std**	**% finished**
2x2	0.22s (0.00h)	0.18s	100.00
3x3	0.82s (0.00h)	0.33s	100.00
4x4	6710.29s (1.86h)	32653.84s	78.82
Precision-based			
Grid size	**Mean**	**Std**	**% finished**
2x2	0.29s (0.00 h)	0.10s	100.00
3x3	0.26s (0.00 h)	0.18s	100.00
4x4	0.84s (0.00 h)	0.35s	100.00
5x5	6.47s (0.00 h)	2.26s	100.00
6x6	68.51s (0.02 h)	39.74s	100.00
7x7	470.37s (0.13 h)	292.18s	96.88
8x8	1772.80s (0.49 h)	546.09s	72.84
9x9	7056.62s (1.96 h)	1570.97s	68.00
10x10	26223.24s (7.28 h)	6080.76s	63.64

We run the linear programs on an HP ProLiant DL980 G7 server with 512 GB RAM and 8 processors Intel E7 2860 with 10 cores each (total 80 cores) using MATLAB's `linprog()` function, and MATLAB's parallel computing capabilities. Table 1 shows the amount of time needed to compute an LPPM function $f(r'|r)$ for different grid sizes αxβ, averaged over combinations of quality and bandwidth restrictions. As expected, the linear program running time grows slower for perturbation-based LPPMs than for precision-based LPPMs, and dummy-based LPPMs quickly become intractable (in fact, we could not compute any LPPM for a 5x5 grid).

While running the experiments we also noticed that when the size of the grid increases MATLAB's linear program solver could not find a solution for some of the optimization problems. The percentage of successful optimizations for each scenario is shown in the third column of Table 1. We note that other linear program solvers could improve

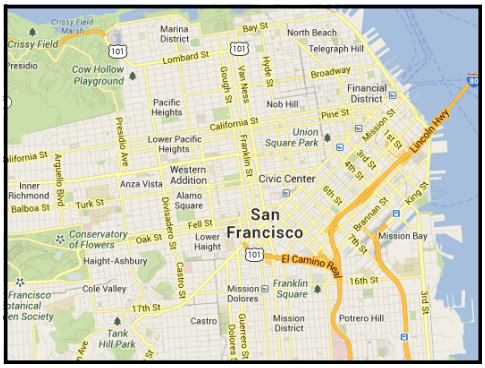


Figure 3: Considered area in San Francisco.

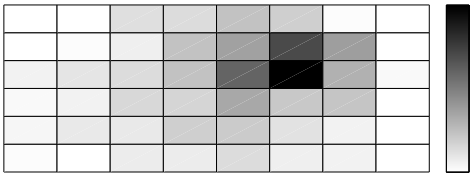


Figure 4: User profile. The darker the region the higher the probability that the user accesses the LBS from this location.

this percentage, as well as reduce the running time of the optimization.

For performance reasons, in our experiments we choose a grid size of 8x6 for perturbation-based and precision-based LPPMs, and 4x3 for dummy-based LPPMs. However, we must stress that a user only needs to run the linear program optimization *once* to compute her optimal protection strategy, and that the mobile device can outsource this operation to a trusted server via a adequately secured connection. Therefore, in reality a larger number of regions can be considered.

Area of interest size. Given a number of regions, the size of the area of interest defines the adversary's inference accuracy. Consider an area of 100 Km^2 divided by a 10x10 cartesian grid. The adversary can narrow his estimation of the users' location to at most 1 Km^2. If on the other hand the area is only 1 Km^2 the the adversary can tighten his estimation to 0.01 Km^2.

In order to make our experiments meaningful we select an area of 8×6 Km $= 48$ Km^2 in Downtown San Francisco which we show in Fig. 3. We divide the area in regions using a cartesian grid of 8x6 or 4x3, depending on the experiment. These grid sizes allow the adversary to infer (with more or less accuracy) the neighborhoods visited by the user. We note that in San Francisco frequent visits to a neighborhood may reveal sensitive information, such as sexual orientation (Castro district), financial status (Financial district), and cultural preferences (Haight-Ashbury).

In [23] Shokri *et al.* demonstrate that the trade-offs between privacy and quality constraints have the same tendency for different users, and that the maximum level of privacy achievable by the LPPM depends on the user's mobility profile. We have run experiments for many individuals in the dataset and confirmed these results. Therefore, without loss of generality, we only show results for one user. We choose as target user the one for which more data is available in the dataset, to have a good estimation of the user's mobility profile. The target user's mobility profile, computed using 36 295 location exposures inside Downtown San Francisco, is shown in Figure 4.

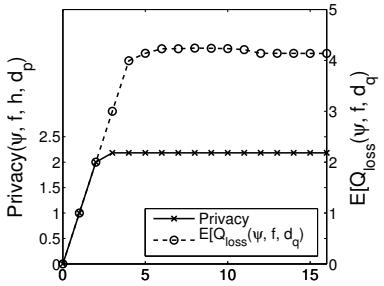


Figure 5: Perturbation-based LPPM: privacy level against the optimal attack; and average expected quality loss.

6.2 Results

We separate our evaluation in three steps. First, we show that the optimal dummy-based and precision-based LPPMs designed using the framework are superior to state of the art LPPMs. Second, we evaluate the impact of quality loss and bandwidth overhead constraints on the privacy provided by optimal LPPMs. Finally, we compare the optimal dummy LPPM with the nearby precision based LPPM in terms of privacy, bandwidth consumption and quality loss.

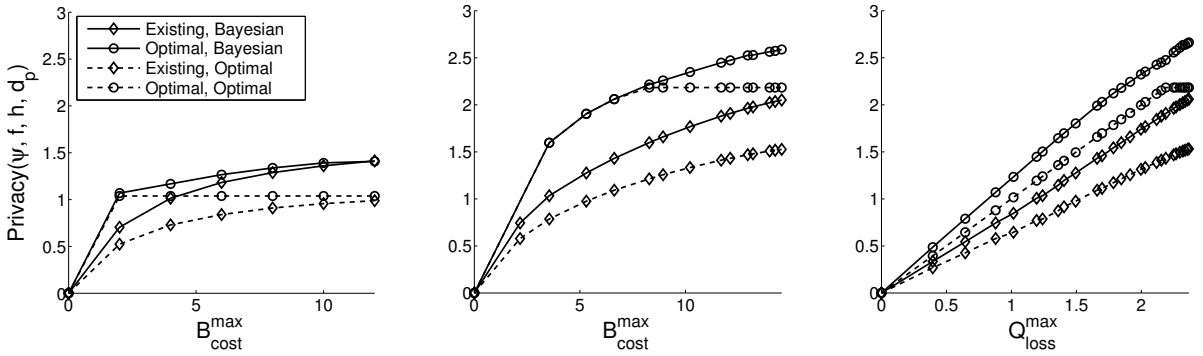
We note that few points are missing in the figures. This is because MATLAB's optimization algorithm was not able to find the solution for these particular combinations of constraints.

6.2.1 *Perturbation-based LPPM*

For the sake of completeness we make a performance analysis of the perturbation-based LPPM used in prior work using our dataset [23]. The results are shown in Fig. 5, where we compare the privacy offered by the optimal perturbation-based LPPM towards the optimal attack output by the linear program, for different expected quality loss constraints. Confirming previous results [23], we observe that when the service quality constraint is loosened sufficiently the level of privacy provided by the LPPM maxes out. This is because these loose constraints allow the LPPM to choose pseudo-locations that do not leak information that is useful for the attack. Therefore the best estimation of the adversary is only dependent on his prior knowledge, i.e., the user's mobility profile. Once quality constraints are sufficiently loosened, the linear program does can output parameters that do not fulfill tightly the quality constraint. As a consequence the average expected quality loss grows slowly and stabilizes around an optimal value that can be much smaller than the maximum tolerated expected quality loss $\text{Q}_{\text{loss}}^{\max}$.

6.2.2 *Bandwith-consuming Optimal LPPMs vs. Existing LPPMs*

Let us consider a case in which the quality loss allows the LPPMs to perturb the real location; i.e., $\text{Q}_{\text{loss}}^{\max} > 0$, and thus r' does not necessarily contain r. Given the considered grid sizes, we observe that as soon as some communication overhead is allowed both optimal and existing LPPMs reach the maximum level of privacy achievable.



(a) Dummy-based LPPMs (4x3 grid). $Q^{max}_{loss} = 0$; no perturbation.

(b) Nearby precision-based LPPMs (8x6 grid). $Q^{max}_{loss} = 0$; no perturbation.

(c) Aggregated precision-based LPPMs (8x6 grid). $B^{max}_{cost} = 0$; no communication overhead.

Figure 6: Comparison of Optimal and existing LPPMs and attacks.

Hence, our analysis focuses on the case where the quality constraint does not allow for perturbation, i.e., $Q^{max}_{loss} = 0$. In order to fairly compare optimal and existing algorithms for every possible user constraint b_u (resp., s_u), we construct an existing dummy-based LPPM (resp., precision-based) as described above, and evaluate its quality loss and bandwidth overhead. These values are used as constraints in the linear programs described in Section 5, which output optimal LPPM parameters that meet the same requirements than their corresponding existing counterparts.

Figure 6 shows the results of the comparison depending on the bandwidth constraint B^{max}_{cost}. We observe that both the optimal defense and attack perform better than their existing counterparts. Like with quality loss, if the bandwidth constraint is sufficiently loosened the level of privacy maxes out. Note that due to the running time of the algorithms the dummy-based strategy is tested on a smaller grid, and hence the maximum privacy achievable, given by the mobility profile, is lower than in the precision-based case. Finally, the aggregate precision-based LPPM does not impose any bandwidth overhead (see Section 5) and therefore the evaluation in in Fig. 6c considers different values for the quality constraint Q^{max}_{loss}.

6.2.3 Trilateral privacy, quality, bandwidth trade-off

We now study the trade-off between privacy, quality, and bandwidth consumption for dummy- and nearby precision-based LPPMs. We note that the aggregate precision-based LPPM does not impose a bandwidth overhead, and hence its performance is similar to that of the perturbation-based mechanism shown in Figure 5, with a slight difference in the expected quality of service loss.

Figures 7a and 8a show the impact of quality loss and bandwidth constraints on privacy for the optimal dummy- and nearby precision-based LPPMs. As expected, when no extra bandwidth consumption is allowed ($B^{max}_{cost} = 0$) privacy increases with the amount of perturbation allowed by the quality constraint. For a given tolerable expected quality loss Q^{max}_{loss}, relaxing the bandwidth constraint increases the level of privacy achievable until it maxes out. Similarly, loosening the quality constraint increases the level of privacy for a given communication overhead.

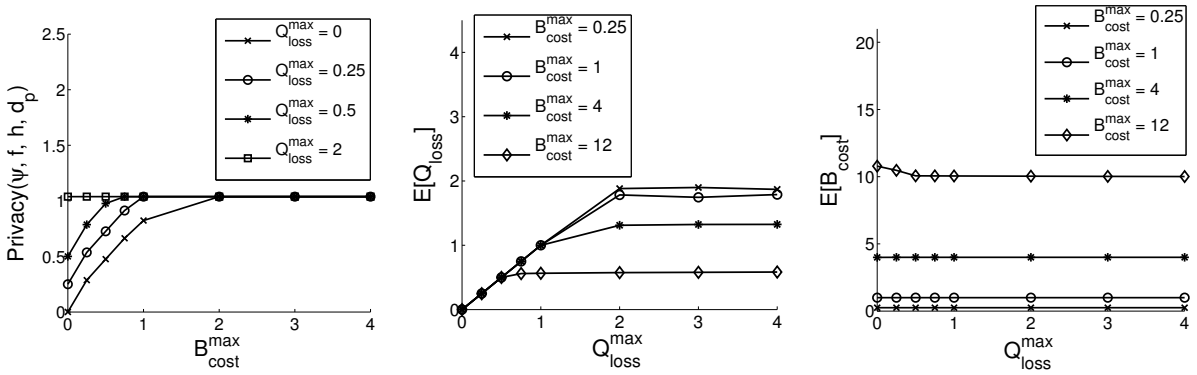
Next we examine the trade-off between the expected quality loss $E[Q_{loss}]$ and expected bandwidth overhead $E[B_{cost}]$ for given combinations of Q^{max}_{loss} and B^{max}_{cost}. Recall that when privacy maxes out, further loosening the quality constraint slows the growth of the average expected quality loss. Similarly, the more bandwidth is allowed the less expected quality loss needs to be traded-off for privacy (see Figures 7b and 8b); and the more quality loss is allowed, the less bandwidth needs to be used on average (see Figures 7c and 8c).

6.2.4 Dummy vs. Nearby Precision LPPMs

Finally, we compare dummy-based and nearby precision-based LPPMs in a 4x3 grid. Figure 9a shows the privacy level obtained by both algorithms for different quality and bandwidth constraints (the former showed in the legend, and the latter increased one unit at a time until privacy maxes out). Unsurprisingly, in Fig. 9a we see that for the same combination on constraints, the dummy LPPM performs better in terms of its achieved level of privacy. This is because the optimal nearby precision-based LPPM is restricted to choose $r' \in \mathcal{R}'$ that contain contiguous regions, while the optimal dummy-based LPPM has no such contiguity restriction and can make the most of the allowed bandwidth consumption.

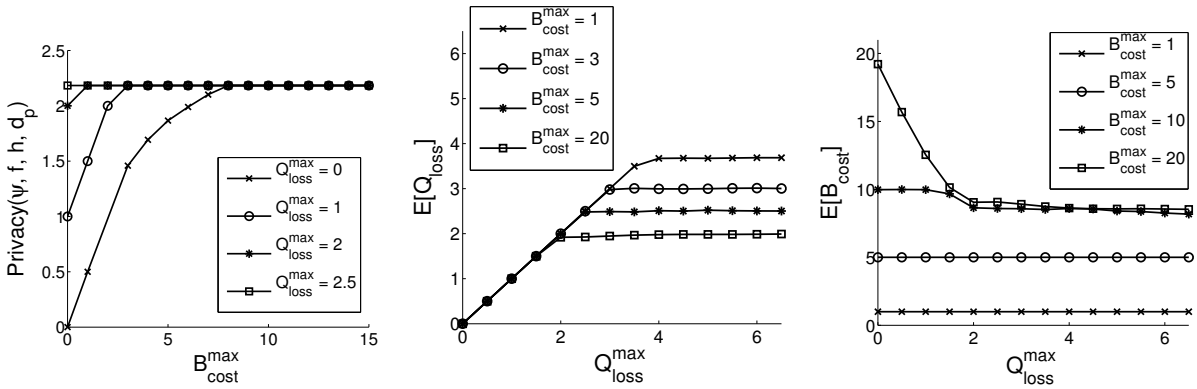
With respect to bandwidth overhead, we can see in Fig. 9b that the expected bandwidth consumption $E[B_{cost}]$ of both algorithms is the same until $E[Q_{loss}]$ stabilizes (i.e., when privacy maxes out). Once privacy has maxed out, the expected bandwidth consumption stabilizes for the nearby precision-based LPPM, but continues growing for the dummy-based LPPM. This is because we consider rectangular contiguous pseudo-locations in the precision-based case and therefore there are less eligible regions than in the dummy-based case, where there is no such restriction. For instance, in a 3x3 grid precision-based pseudo-locations can only be formed by 1, 2, 4, 6, and 9 contiguous regions in $\mathcal{R}$, while dummy-based LPPMs can output pseudo-locations containing any combination of 1 to 9 regions. Hence, even if the bandwidth constraint is loosened, the precision-based LPPM has fewer large pseudo-locations to choose from, and thus consumes less bandwidth than the dummy-based strategy, which can select more expensive alternatives.

In terms of quality loss, the dummy-based LPPM suffers more quality degradation than the precision-based LPPM (see Fig. 9c). This is due to the freedom of the dummy-based strategy to select any combination of locations. This



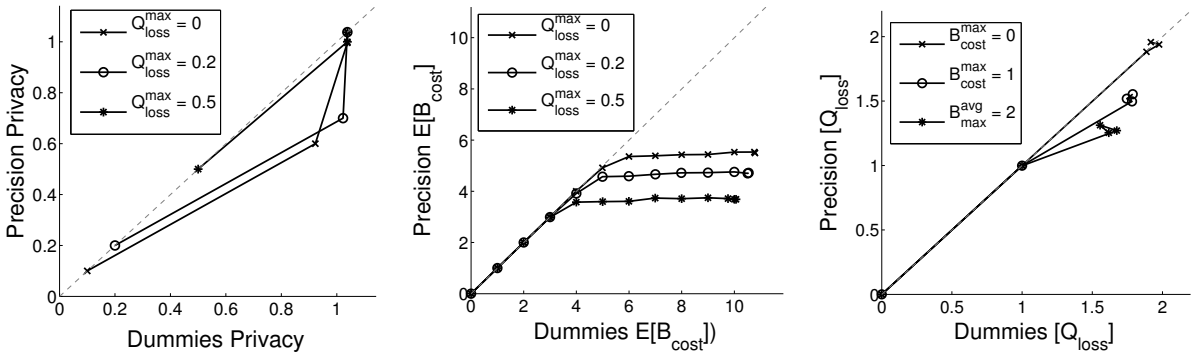
(a) Privacy (b) Expected quality loss (c) Expected bandwidth

Figure 7: Dummy-based LPPM. (4x3 grid)



(a) Privacy. (b) Expected quality loss. (c) Expected bandwidth.

Figure 8: Precision-based LPPM. (8x6 grid)



(a) Privacy. (b) Bandwidth. (c) Quality of service loss.

Figure 9: Comparison of optimal dummy-pased LPPM vs. nearby precision-based LPPM.

allows dummy-based LPPMs to squeeze the quality constraint more efficiently than the precision-based strategy, which is limited to choosing contiguous locations. The clusters at the end of the lines in the figure reflect that the values $E[Q_{loss}]$ and $E[B_{cost}]$ fluctuate slightly once they have stabilized (Fig. 9b).

7. CONCLUSIONS

Location Privacy-Preserving Mechanisms (LPPMs) mitigate privacy risks derived from the disclosure of location data when using Location Based Services (LBSs). Shokri *et al.* proposed in prior work a framework to design optimal LPPMs towards strategic adversaries, aware of the LPPM algorithm and the users' mobility patterns [23], for applications in which users only reveal their location sporadically. The proposed framework allows users to set a limit on the maximum tolerated quality loss incurred by the LPPM, but it fails to capture constraints on the resource consumption (e.g., bandwidth) introduced by some LPPM strategies, such as sending dummies, or decreasing the precision of exposed locations.

In this work we have extended Shokri *et al.*'s framework to allow the user to specify a bandwidth constraint. Furthermore, we have modeled two popular strategies to trade-off bandwidth for privacy: a scheme based on sending dummy locations to the LBS, and a scheme based on reducing the precision of the location sent to the LBS.

We have evaluated the performance of LPPMs that consume bandwidth using the CRAWDAD taxi dataset. Our results show that the optimal dummy- and precision- based LPPMs provide more privacy than their respective naive counterparts. Furthermore, both LPPMs perform better than perturbation-based strategies if communication overhead is allowed by the user, with dummy-based LPPMs being the the best choice for a given combination of quality and bandwidth constraints. Furthermore, the results of our simulations show that users can achieve the maximum privacy allowed by their mobility profiles by either permitting a sufficiently large quality of service loss, or bandwidth consumption, or an adequate combination of both.

Acknowledgments.

We thank Reza Shokri for sharing his optimization code. This research was supported in part by the European Union under project LIFTGATE (Grant Agreement Number 285901) and the European Regional Development Fund (ERDF); and by the projects: IWT SBO SPION, FWO G.0360.11N, FWO G.0686.11N, and GOA TENSE (GOA/11/007).

8. REFERENCES

[1] Foursquare. https://foursquare.com/.

[2] Google maps. https://maps.google.com/.

[3] Zeit online: Betrayed by our own data. http://www.zeit.de/digital/datenschutz/2011-03/data-protection-malte-spitz.

[4] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: Differential privacy for location-based systems. *CoRR*, abs/1212.1984, 2012.

[5] A. Beresford and F. Stajano. Mix zones: user privacy in location-aware services. In *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, pages 127–131, 2004.

[6] A. R. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, Jan. 2003.

[7] J. Freudiger, R. Shokri, and J.-P. Hubaux. On the optimal placement of mix zones. In I. Goldberg and M. J. Atallah, editors, *9th Privacy Enhancing Technologies Symposium*, volume 5672 of *LNCS*, pages 216–234. Springer, 2009.

[8] J. Freudiger, R. Shokri, and J.-P. Hubaux. Evaluating the privacy risk of location-based services. In G. Danezis, editor, *Financial Cryptography and Data Security*, volume 7035 of *LNCS*, pages 31–46. Springer Berlin Heidelberg, 2012.

[9] B. Gedik and L. Liu. Location privacy in mobile systems: A personalized anonymization model. In *Distributed Computing Systems, 2005. ICDCS 2005. Proceedings. 25th IEEE International Conference on*, pages 620–629, 2005.

[10] P. Golle and K. Partridge. On the anonymity of home/work location pairs. In H. Tokuda, M. Beigl, A. Friday, A. J. B. Brush, and Y. Tobe, editors, *7th International Conference on Pervasive Computing*, volume 5538 of *LNCS*, pages 390–397. Springer, 2009.

[11] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the 1st international conference on Mobile systems, applications and services*, MobiSys '03, pages 31–42, New York, NY, USA, 2003. ACM.

[12] B. Hoh and M. Gruteser. Protecting location privacy through path confusion. In *Security and Privacy for Emerging Areas in Communications Networks, 2005. SecureComm 2005. First International Conference on*, pages 194–205, 2005.

[13] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Enhancing security and privacy in traffic-monitoring systems. *Pervasive Computing, IEEE*, 5(4):38–46, 2006.

[14] H. Kido, Y. Yanagisawa, and T. Satoh. An anonymous communication technique using dummies for location-based services. In *Pervasive Services, 2005. ICPS '05. Proceedings. International Conference on*, pages 88–97, 2005.

[15] J. Krumm. Inference attacks on location tracks. In A. LaMarca, M. Langheinrich, and K. Truong, editors, *Pervasive Computing*, volume 4480 of *LNCS*, pages 127–143. Springer Berlin Heidelberg, 2007.

[16] H. Lu, C. S. Jensen, and M. L. Yiu. Pad: privacy-area aware, dummy-based location privacy in mobile services. In *Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*, MobiDE '08, pages 16–23, New York, NY, USA, 2008. ACM.

[17] J. T. Meyerowitz and R. R. Choudhury. Hiding stars with fireworks: location privacy through camouflage. In K. G. Shin, Y. Zhang, R. Bagrodia, and R. Govindan, editors, *15th Annual International Conference on Mobile Computing and Networking (MOBICOM)*, pages 345–356. ACM, 2009.

[18] F. G. Olumofin, P. K. Tysowski, I. Goldberg, and U. Hengartner. Achieving efficient query privacy for location based services. In M. J. Atallah and N. J. Hopper, editors, *10th International Privacy Enhancing Technologies Symposium, PETS 2010,*, volume 6205 of *LNCS*, pages 93–110. Springer, 2010.

[19] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser. CRAWDAD data set epfl/mobility (v. 2009-02-24). Downloaded from http://crawdad.cs.dartmouth.edu/epfl/mobility, Feb. 2009.

[20] R. Shokri, J. Freudiger, and J.-P. Hubaux. A unified framework for location privacy. *3rd Hot Topics in Privacy Enhancing Technologies (HotPETs)*, 2010.

[21] R. Shokri, G. Theodorakopoulos, J.-Y. L. Boudec, and J.-P. Hubaux. Quantifying location privacy. In *32nd IEEE Symposium on Security and Privacy, S&P 2011*, pages 247–262. IEEE Computer Society, 2011.

[22] R. Shokri, G. Theodorakopoulos, G. Danezis, J.-P. Hubaux, and J.-Y. L. Boudec. Quantifying location privacy: The case of sporadic location exposure. In S. Fischer-Hübner and N. Hopper, editors, *11th International Symposium Privacy Enhancing Technologies, PETS 2011*, LNCS, pages 57–76. Springer, 2011.

[23] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. L. Boudec. Protecting location privacy: optimal strategy against localization attacks. In T. Yu, G. Danezis, and V. D. Gligor, editors, *ACM Conference on Computer and Communications Security CCS'12*, pages 617–627. ACM, 2012.

[24] R. Shokri, C. Troncoso, C. Díaz, J. Freudiger, and J.-P. Hubaux. Unraveling an old cloak: k-anonymity for location privacy. In E. Al-Shaer and K. B. Frikken, editors, *ACM Workshop on Privacy in the Electronic Society, WPES 2010*, pages 115–118. ACM, 2010.

[25] Y. Wang, D. Xu, X. He, C. Zhang, F. Li, and B. Xu. L2p2: Location-aware location privacy protection for location-based services. In *INFOCOM, 2012 Proceedings IEEE*, pages 1996–2004, 2012.

[26] T.-H. You, W.-C. Peng, and W.-C. Lee. Protecting moving trajectories with dummies. In *Mobile Data Management, 2007 International Conference on*, pages 278–282, 2007.

of observing r' as follows:

$$\Pr(r|r') = \frac{\Pr(r,r)}{\Pr(r')} = \frac{f(r'|r)\psi(r)}{\sum_r f(r'|r)\psi(r)}\,, \tag{28}$$

$$\Pr(r') = \sum_r \psi(r) f(r'|r)\,. \tag{29}$$

The goal of the adversary is to choose the estimated location $\hat{r}$ that minimizes the expected privacy of the user conditioned to the exposed location being r':

$$\min_{\hat{r}} \sum_r \Pr(r|r') d_p(\hat{r}, r) \tag{30}$$

Combining Eqs (28), (29), and (30), we can express the unconditional expected privacy that the user aims at maximizing as:

$$\sum_{r'} x_{r'}\,, \tag{31}$$

where we have defined

$$x_{r'} \doteq \min_{\hat{r}} \sum_r \psi(r) f(r'|r) d_p(\hat{r}, r')\,. \tag{32}$$

Shokri *et al.* note that $x_{r'}$ can be transformed as a series of linear constraints $x_{r'} \leq \min_{\hat{r}} \sum_r \psi(r) f(r'|r) d_p(\hat{r}, r'), \forall r$ and hence $x_{r'}$ can be use as decision variable representing the privacy offered by an LPPM.

Similarly, if we consider the attack $h(\hat{r}, r')$ given that a true location is r and corresponding exposed pseudo-location r', the conditional expected user privacy is:

$$\sum_{\hat{r}} h(\hat{r}|r') d_p(\hat{r}, r)\,. \tag{33}$$

Taking into account the prior knowledge of the adversary on the user's profile $\psi(r)$ the unconditional expected user privacy can be written as:

$$\sum_r \psi(r)\, y_r\,, \tag{34}$$

where

$$y_r \doteq \max_{r'} \sum_{\hat{r}} h(\hat{r}|r') d_p(\hat{r}, r)\,. \tag{35}$$

Shokri *et al.* note that y_r can be transformed as a series of linear constraints $y_r \geq \sum_{\hat{r}} h(\hat{r}|r') d_p(\hat{r}, r), \forall r'$ and hence y_r can be use as decision variable representing the privacy obtained against an attack $h(\hat{r}, r)$.

APPENDIX

A. PRIVACY DECISION VARIABLES

In this section we sketch the derivation of the privacy decision variables used in the linear programs in Sections 4.1 and 5. We refer the reader to [23] for more details on the linear programs derivation.

Recall that in the Stackelberg approach the adversary knows the user's choice of LPPM $f(\cdot)$, as well as the user's profile $\psi(r)$. Hence, the adversary can compute the posterior probability $\Pr(r|r')$ that the user being at r when the exposed pseudo-location is r', as well as the probability $\Pr(r)$