Dynamic Differential Location Privacy with Personalized Error Bounds

Lei Yu, Ling Liu, Calton Pu
School of Computer Science, College of Computing, Georgia Institute of Technology
Email: leiyu@gatech.edu, ling.liu@cc.gatech.edu, calton.pu@cc.gatech.edu

***Abstract*—Location privacy continues to attract significant attentions in recent years, fueled by the rapid growth of location-based services (LBSs) and smart mobile devices. Location obfuscation has been the dominating location privacy preserving approach, which transforms the exact location of a mobile user to a perturbed location before its public release. The notion of location privacy has evolved from user-defined location k-anonymity to two statistical quantification based privacy notions: geo-indistinguishability and expected inference error. The former promotes differential location privacy but does not protect location against inference attacks of Bayesian adversary with using prior information, whereas the latter promotes the background inference resilient location privacy but does not guarantee differential location privacy with respect to geo-indistinguishability. In this paper we argue that geo-indistinguishability and expected inference error are two complementary notions for location privacy. We formally study the relationship between two privacy notions. By leveraging this relationship and a personalized error bound, we can effectively combine the two privacy notions. We develop PIVE, a two-phase dynamic differential location privacy framework. In Phase I, we take into account the user-defined inference error threshold and the prior knowledge about the user's location to determine a subset of locations as the protection location set for protecting the actual location by increasing adversary's expected location inference error. In Phase II, we generate pseudo-locations (i.e., perturbed locations) in the way that achieves differential privacy over the protection location set. This two-phase location obfuscation is constructed dynamically by leveraging the relationship between two privacy notions based on adversary's current prior information and user-specific privacy requirements on different locations and at different times. Experiments with real-world datasets demonstrate that our PIVE approach effectively guarantees the two privacy notions simultaneously and outperforms the existing mechanisms in terms of adaptive privacy protection in presence of skewed locations and computation efficiency.**

I. Introduction

We are entering a mobile Internet era where people and vehicles are constantly connected while on the move through mobile and wireless devices. Location becomes an important piece of information for enhancing such ubiquitous connectivity through a rich selection of location based services (LBSs) and applications, such as Uber [3], Yelp [4], Foursquare [1]. On one hand, the emergence of location aware computing and location-based services creates great opportunities for empowering business with new competitive edges and enriching citizen with life-enhancing experiences. On the other hand, such continuous publishing and sharing of mobile users' location information may open doors to potential misuse and abuse of private location information and serious location privacy risks, such as exposing places that a user has visited, the travel patterns of a user, and using the location information to infer users' activities and uncover many unauthorized personal information such as their political views, religious affiliation, or state of health.

Location privacy research has drawn significant interests in recent years. Considering the high utility of location information and personalized privacy risk variations, instead of cryptographic solutions, a large body of location privacy research have been centered on the location obfuscation mechanisms that allow mobile travelers to use LBSs with perturbed location instead of exact location, referred to as *pseudo-location*, such that the release of the pseudo-location can prevent the disclosure of user-specific and request-specific sensitive location information [5], [6], [10], [11], [20], [22], while maintaining desired utility of location information.

Recently, geo-indistinguishability [5] and expected inference error [21], [22] are proposed in the literature as the two statistical notions of location privacy. Geo-indistinguishability is derived from differential privacy [8] and ensures that for any two location points that are geographically close, the location obfuscation mechanism will produce a pseudo-location with similar probabilities. The expected inference error, as a statistical metric instead, takes into account the prior information of an adversary about user's location, and measures location privacy by the expected distance between the estimated location by the adversary and the true location. A number of location obfuscation mechanisms [22], [23] have been developed solely based on the privacy notion of expected inference error.

In this paper, we argue that geo-indistinguishability and expected inference error are two complementary notions for location privacy. Existing geo-indistinguishable mechanisms [5], [6] guarantees location privacy with respect to the information leakage through a differential privacy based location obfuscation mechanism, but they do not consider the inference attacks using prior knowledge [20]. We performed the bound analysis to formally study the relationship between geo-indistinguishability and expected inference error. We show that geo-indistinguishability may not adequately protect the absolute privacy of user's location against inference attacks

NDSS '17, 26 February - 1 March 2017, San Diego, CA, USA

http://dx.doi.org/10.14722/ndss.2017.23xxx

with prior information. On the other hand, the mechanisms with expected inference error as privacy metric are constructed based on the assumption of certain types of prior information that the adversary may have, but without consideration of constraint on the posterior information gain from the release of pseudo-locations. These mechanisms may be vulnerable to inference attacks with arbitrary prior knowledge. Thus, we argue that a strategic combination of the two privacy notions can double shield location privacy by simultaneously limiting information leakage of the location perturbation mechanism and ensuring the inference error to be constrained for inference attacks with prior information the adversary may have.

In addition to combining the two privacy notions for effective defense against inference attacks, we also argue that an effective location obfuscation mechanism should maintain desired location utility and service quality for respective mobile users and their LBSs. In practice, mobile users may have very different privacy requirements for different types of LBSs. Even for the same LBS, users may have different privacy demands for different locations or for the same location at different times. For example, a user may want the expected inference error of adversary to be larger than 1km when he is in a hospital or a religious event, but may reduce this requirement to 200 meters when he is in a restaurant with a lot of other restaurants nearby; or the user may not care about privacy at some places (e.g., her home or office) during certain periods of a day, but needs the privacy at other places, such as her travel routes and stops along some trajectories.

In this paper, we propose to design a dynamic differential location privacy mechanism with personalized error bounds. First, we formally study the relationship between geo-indistinguishability and expected inference error and examine their limitations through experimental study. The relationship between two privacy notions helps to determine the noise level of location obfuscation required for protecting a location against inference attacks. Second, we allow users to define personalized error bound for each of their locations and introduce the concept of *protection location set* for each location, which identifies the neighborhood locations based on both the personalized error bound constraint and the prior distribution that the adversary may have based on historical locations of a user, her mobility model or the population density. Based on the above development, we design a two-phase dynamic differential location privacy framework, called PIVE, which integrates geo-indistinguishability and expected inference error to effectively protect location privacy against two popular types of inference attacks: optimal inference attack and Bayesian inference attack. This framework constructs pseudo-locations dynamically and adaptively, based on multiple pieces of information that may change frequently in the spatial-temporal context of a mobile user, such as the user's current location at the time of her service request, her current location privacy requirements, her location utility and LBS quality preferences, and the prior information that the adversary may have at this time. In Phase I, we utilize the user-defined inference error threshold and the prior knowledge about the user's location to determine the protection location set for protecting the actual location of a user and ensuring the lower bound of adversary's expected location inference error over this protection location set. In Phase II, we generate pseudo-locations that achieve differential privacy on this protection location set. The former aims to bound the expected inference error in the worst case and the latter aims to scope the possible posterior information leakage. The PIVE approach provides dynamic differential location privacy with personalized error bound and can work adaptively in presence of skewed prior distribution of locations and efficiently for the scenarios in which users may have personalized and non-uniform privacy needs at different locations and for different LBSs.

Previous work [20] by Shokri is the first to identify the need for integrating the two privacy notions and to propose a joint optimization approach. This approach combines the two privacy notions together in parallel in a linear program and produce the distribution of perturbed locations statically once for all locations in an area, and we refer to it as the global optimization approach. Compared to the joint optimization [20], PIVE takes a sequential and local approach to combine two privacy notions. It separately applies the expected inference error metric first, which produces a neighborhood protection location set for the user's location by leveraging user defined error bound and the prior information, and then produces the perturbed location by ensuring geo-indistinguishability and at the same time increasing the resilience of perturbed location against inference attacks. Another feature of PIVE that is different from the joint optimization approach is to leverage the user defined personalized error bound (threshold) for different locations or for the same location at different times and for computing the protection location sets dynamically and adaptively. This allows PIVE to balance privacy and utility for different locations while meeting the personalized inference error bound constraint for perturbed locations.

PIVE algorithms are highly efficient in terms of computation complexity, compared to existing mechanisms that need to solve a linear program with $|\mathcal{X}|^2$ decision variables and up to $O(|\mathcal{X}|^3)$ constraints for previous joint optimization approach [20], where $\mathcal{X}$ is the number of all possible locations of a user. First, PIVE only requires the search of a protection location set locally within the neighborhood of a user's current location by leveraging user-defined error bound, and simple probability computation for the exponential mechanism. This locality based design enables PIVE to adapt to the dynamic changes of both prior information and privacy preferences per location more efficiently. Second, PIVE adaptively adjusts the noise level of location obfuscation to prior information through searching a protection location set under the minimum inference error bound constraint, which provides dynamic differentially private mechanism to generate perturbed locations. We implement the PIVE dynamic location obfuscation mechanism and evaluate PIVE with real-world datasets. Our experimental results show that the PIVE approach effectively guarantees the two privacy notions simultaneously and outperforms the existing mechanisms that secure geo-indistinguishability or that quantify location privacy by expected inference errors.

II. Related Work

Location privacy research started about ten years ago with the notion of location k-anonymity with two landmark results: (i) uniform location k-anonymity [13] and (ii) user-defined, personalized location k-anonymity [12]. The location k-anonymity based solutions hide a user's exact location point using a spatial region that meets the two constraints: (a) it

contains the exact location point of the user; and (b) there are at least $k-1$ other users who will use the same location region as their released location to meet the k anonymity requirement. Alternatively, some location obfuscation mechanisms achieve privacy by using landmark objects or random perturbation instead of k-anonymity. [14] proposes to use the location of a closest landmark object as the perturbed location such that the LBS severs process the location query based on the landmark. [25] proposes to search the region that has sufficient user footprints such that the user can feel safe for his location privacy. However, neither user-defined privacy notion nor any formal privacy notion is provided and guaranteed by the proposed region-based location cloaking mechanism.

Recently, two stronger privacy notions are proposed based on statistical quantification of attack resilience: expected inference error [21], [22] and geo-indistinguishability [5]. The former advocates the privacy notion based on its attack resilience to the prior information of adversary by measuring the expected inference error and the latter promotes the differential privacy notion to constrain the posterior information gain of an adversary based on the release of pseudo-locations of mobile user. A number of location obfuscation mechanisms [5]–[7], [20], [22] have been developed based on them. For example, based on the prior distribution of user's location, Shokri et al. [22] proposed an optimal construction mechanism for location perturbation against inference attacks through linear programming. The mechanism aims to maximize the expected inference error (resp. service quality) given the constraint on the service quality loss(resp. expected inference error). The service quality loss is characterized by the expected distance between real and reported locations. Based on Shokri et al.'s optimization framework, Theodorakopoulos et al [23] advocated to follow a user over his trajectory and maximizes privacy for each location with considering privacy leakage due to location correlation between past, current and future locations in a trajectory. Andrés et al. [5] proposed the notion of geo-indistinguishability. A Planar Laplace (LP) mechanism is developed to achieve the ϵ geo-indistinguishability by adding noise to actual location drawn from a polar Laplacian distribution. Several recent location privacy development projects [2], [10], [11] have adopted or extended ϵ geo-indistinguishability for location privacy protection. Bordenabe et al. [6] proposed an optimal geo-indistinguishable mechanism to minimize the service quality loss. Similar to [22], it uses linear programing to minimize global expected service quality loss, with a uniform privacy parameter for geo-indistinguishability. Chatzikokolakis et al [7] defines privacy mass over the point of interests on the plane and adaptively decide the privacy parameter of geo-indistinguishability for a location with considering local characteristics of each area.

The mechanisms in [6], [20], [22] follow a global optimization framework: given the privacy or service quality constraints, a linear programming model is formulated to maximize service quality or privacy respectively. Such formulation uses uniform differential privacy parameter and global privacy/quality metrics averaged over all locations, which offers uniform privacy/utility with respect to all locations and all LBSs. It could be a difficult task to pre-determine the constraint for every location where a user will ask for any LBS service request with his personalized and spatial-temporal dependent as well as LBS dependent privacy requirement. Besides, these techniques are computationally costly due to solving a linear program with $|\mathcal{X}|^2$ decision variables, and the perturbation solution is statically constructed once for all locations, which can be prohibitively expensive for frequently changing prior information and frequently changing privacy/utility preference by users at different locations and times.

Our work is primarily related to two recent research efforts in [20] and [7]. Concretely, Shokri [20] is the first to propose a joint mechanism to integrate the two privacy notions using a linear programming framework, demonstrating the potential for improvement on privacy protection. However, the joint optimization mechanism uses uniform differential privacy parameter and global privacy/utility metrics by averaging over all locations. We argue that an overall metric for all locations and a per-location based metric may result in different allocations of privacy and utility. Thus PIVE is more suitable to situations where mobile users may have different privacy/utility preferences for different locations, at different time and working with different LBSs. Next, unlike most existing geo-indistinguishable mechanisms that consider uniform differential privacy parameters for all users and all locations, Chatzikolakis and his co-authors [7] propose to adaptively decide the noise level of geo-indistinguishability according to the privacy characteristics of local area. They compute the density of a local area for each location and adds less noise for perturbed location if the density of the actual location area is high and more noise when the actual location falls into the low density areas. However, the density of a local area is defined in terms of the public locations such as restaurants, churches and hospitals. Thus this approach assumes that these different types of public locations are of the same privacy sensitivity for all mobile users at all time, and thus fails to model the personalized geo-indistinguishability with respect to different locations, different times and different LBSs. In comparison, PIVE adaptively adjusts the noise level of location obfuscation according to a personalized error bound and the prior distribution in local area.

III. Overview

In this section we first introduce the notation of differential privacy, describe the model of location obfuscation and the adversary model used in this paper. Then, we state the problem to be addressed in this paper.

A. Differential Privacy

Differential privacy is a rigorous mathematical framework that offers provable privacy guarantees for protecting individual data in statistical databases and has recently become a de-facto standard for privacy. It ensures that arbitrary changes to a single individual's row result in only statistically insignificant changes in the outcome of a data analysis. Formally,

Definition 1 (Differential Privacy [8]): A randomized mechanism $\mathcal{A}$ provides ϵ-differential privacy if for any two neighboring database D_1 and D_2 that differ in only a single entry, $\forall S \subseteq Range(\mathcal{A})$,

$$\frac{\Pr(\mathcal{A}(D_1) \in S)}{\Pr(\mathcal{A})(D_2) \in S)} \leq e^{\epsilon} \tag{1}$$

The standard approach to achieve differential privacy is the sensitivity method [8], [9] (e.g., Laplacian mechanism)

that adds to the query output the noise proportional to the sensitivity of the query function. The sensitivity measures the maximum change in the query answers due to the change of a single database entry.

Definition 2 (Sensitivity [9]): The sensitivity of a query function $q : \mathcal{D} \rightarrow \mathbb{R}^d$ is

$$\Delta q = \max_{D_1, D_2} \|q(D_1) - q(D_2)\|_1 \tag{2}$$

where D_1, $D_2 \in \mathcal{D}$ are any two neighboring datasets that differ at most one element, $\|\cdot\|_1$ denotes L_1 norm.

To achieve ϵ-differential privacy, the Laplacian mechanism perturbs the output by $q(D) + Lap(\Delta q/\epsilon)$, where $Lap(*) = (Z_1, \ldots, Z_d)$ in which Z_i are drawn i.i.d from Laplace distribution. Such differentially private mechanism ensures that two neighboring datasets are indistinguishable on the distribution of query answers.

The exponential mechanism [17] is another mechanism that preserves ϵ-differential privacy. Given the output range R, a utility function $u : D \times R \rightarrow \mathbb{R}$ is defined, which maps the dataset/output pairs to utility scores. The sensitivity of utility function u is

$$\Delta u = \max_{r \in R} \max_{D, D'} |u(D, r) - u(D', r)| \tag{3}$$

over any two neighboring datasets D and D'.

Definition 3 (The exponential mechanism [17]): The exponential mechanism selects and outputs an element $r \in R$ with probability proportional to $\exp(\frac{\epsilon u(D,r)}{2\Delta u})$.

B. Location Obfuscation Mechanism

In this paper we are interested in the location based services in which the users sporadically reveal their locations for issuing spatial queries, e.g., finding the nearby points-of-interests or friends. We do not consider the protection of the users' identities that prevents the adversary to discover which user issues the query. In this case, the typical way to preserve the users' location privacy is to randomly obfuscate the user's actual location to a pseudo-location and report this pseudo-location to the location based service providers. In this paper we assume discretized locations as in [6], [22] and use $\mathcal{X}$ to denote the set of the user's possible locations. An obfuscation mechanism determines the random mapping between the user's actual locations A and pseudo-locations O, with following the probability distribution

$$f(x'|x) = \Pr(O = x'|A = x) \qquad x, x' \in \mathcal{X} \tag{4}$$

That is, it takes the actual location x as input and chooses a pseudo-location x' by sampling from the distribution $f(x'|x)$. An obfuscation mechanism is indeed a specification of probability distributions $f(\cdot|\cdot)$ over $\mathcal{X}$. Different obfuscation mechanisms determine such probability distributions in different ways.

C. Adversary Model

This paper assumes the adversary that has prior knowledge about user's location. We argue that the prior information about users' locations inherently exists because of the publicly available transportation information, geographical information of points of interest, road networks, residential area, population distribution, and human movement pattern, etc. Following previous works [21], [22], the prior knowledge is captured by a prior (probability) distribution π over the set of possible locations of the user, $\mathcal{X}$. The adversary can build π for the target user in multiple ways:

- Using the population density or popularity [7], [25] of every place as π that can be obtained from public traces, check-in datasets or demographic information;
- Using the user's historical access information to a location based service that records his locations from which he sent location based queries [22].
- Using the mobility pattern modeled by Markov chain to infer the possible locations of a user at current time and their probabilities given his previous disclosed locations [24].

In this paper we assume the adversary with prior knowledge of π regardless of in which way it is derived. We also assume that the adversary also knows the location obfuscation mechanism, i.e, how it works and the distribution f. Such adversary is called an informed adversary [9].

The adversary's goal is to infer the user's actual location x. Once the adversary observes the pseudo-location x' reported by the user, he computes the posterior probability distribution, $\Pr(x|x')$ for $x \in \mathcal{X}$, i.e., the probability that x is the actual location that generated x':

$$\Pr(x|x') = \frac{\pi(x) f(x'|x)}{\sum_{x \in \mathcal{X}} \pi(x) f(x'|x)} \tag{5}$$

Based on the posterior distribution, a *Bayesian adversary* can perform **optimal inference attack** [22] which aims to minimize his expected inference error, i.e., the expected distortion between the estimated location $\hat{x}$ and user's actual location x, given an observed pseudo-location x'. That is,

$$\hat{x} = \arg\min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \Pr(x|x') d_p(\hat{x}, x) \tag{6}$$

where d_p can be Hamming distance or Euclidean distance between locations, or their semantic dissimilarity, which captures the privacy loss from inference attack. We assume d_p to be Euclidean distance d for optimal inference attack.

If d_p is Hamming distance, for which $d_p(\hat{x}, x) = 0$ if $\hat{x} = x$, and $d_p(\hat{x}, x) = 1$ otherwise, it is easy to see that the optimal inference attack actually guesses the actual location as the one having the maximum posterior probability. We call this attack as **Bayesian inference attack**, represented by

$$\hat{x} = \arg\max_{x \in \mathcal{X}} \Pr(x|x') \tag{7}$$

D. Problem Statement

We can categorize the existing research on quantifying location privacy into two broad categories based on two notions of location privacy: geo-indistinguishability and expected inference error. The location privacy solutions that promote geo-indistinguishability are primarily based on the theory of differential privacy [8]. The solutions that quantify location privacy by the amount of expected inference error are

typically based on Bayesian theory and thus are referred to as Bayesian optimal mechanisms. The class of solutions based on geo-indistinguishability protect location privacy without any assumption of adversary's prior information but consequently do not consider absolute location privacy against inference attacks in terms of expected inference error when the adversary has some prior knowledge about the user's exact location or past released locations. In contrast, the Bayesian optimal mechanisms advocate the background inference resilient location privacy but are not as robust as geo-indistinguishability against adversary with arbitrary prior information.

The problem statement can be summarized from three dimensions. First, geo-indistinguishability and expected inference error are two complementary privacy notions for protecting location privacy against inference attacks. It is critical to understand the relationship between the two privacy notions, and the limitations of existing location obfuscation mechanisms that support only one of the two privacy notions. Second, it is not only beneficial but also feasible to develop a location obfuscation mechanism that can effectively integrate the two privacy notions. Third, incorporating user-defined constraint, such as minimum inference error bound, not only improves the usability perspective, which is critical for the wide deployment of privacy protection models, but also enables adaptive noise adjustment for geo-indistinguishability and supports customizable privacy/utility requirement of mobile users that allows personalized error bounds at different locations, different times, and for different LBSs. This motivates the design and implementation of PIVE, a two-phase dynamic differential location privacy framework for ensuring both notions of location privacy with personalized error bounds.

IV. Location Privacy Notions

In this section we provide a detailed analysis and illustration of the two location privacy notions: expected inference error and geo-indistinguishability. We first briefly describe each notion, its respective location perturbation model, compare the mechanisms based on these two privacy notions and identify and illustrate their inherent problems through both formal and experimental analysis.

A. Expected Inference Error

Under the inference attack of Bayesian adversary, the location privacy offered by a mechanism is measured by the expected inference error of the adversary averaged over all possible locations in $\mathcal{X}$, referred to as unconditional expected inference error [21], [22], computed as

$$\sum_{x' \in \mathcal{X}} \Pr(x') \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \Pr(x|x') d_p(\hat{x}, x) \tag{8}$$

$$= \sum_{x' \in \mathcal{X}} \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi(x) f(x'|x) d_p(\hat{x}, x) \tag{9}$$

Similarly, the service quality loss is measured by the unconditional expected distance between actual location and reported pseudo-location over the quality metric $d_q(\cdot)$, i.e.,

$$\sum_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} \pi(x) f(x'|x) d_q(x', x) \tag{10}$$

where d_q determines the quality loss by reporting x' instead of actual location x. Since the accuracy of location based queries like nearest neighbor and range queries usually depends on the Euclidean distance between the actual location and reported location, we use the Euclidean distance d as d_q, as in previous works [6], [20].

An optimal mechanism [22] has been proposed to maximize the expected inference error (resp. service quality) given the constraint on the service quality loss (resp. expected inference error). In such approach, privacy and quality are controlled in terms of these global performance metrics that are averaged over all locations, which does not provide users a straightforward way to explicitly specify different privacy/quality requirements at different locations and times. Also, for the prior information that is dynamically built by the adversary with mobility model [24], a linear program has to be recomputed under every change. More importantly, the construction relies on the assumption about adversary's prior information, different prior information with higher accuracy level may cause privacy degradation of the mechanism, as shown in [20].

We note that the upper limit of expected inference error is achieved when the maximum tolerable service quality loss becomes sufficiently large or not bounded. In this case, the pseudo-locations are generated independently of user's locations, and the adversary's best strategy is to make guess based on prior distribution. Therefore, the upper limit of expected inference error is

$$ExpErr_{max} = \min_{\hat{x}} \sum_{x \in \mathcal{X}} \pi(x) d_p(\hat{x}, x) \tag{11}$$

B. Geo-indistinguishability

A mechanism satisfies ϵ_g-geo-indistinguishability [5] iff for all x, y,

$$\frac{f(x'|x)}{f(x'|y)} \leq e^{\epsilon_g d(x,y)} \tag{12}$$

where $d(x, y)$ is the Euclidean distance between x and y. It ensures that for two locations that are geographically close, the probability distributions of pseudo-locations generated at them are similar. Note, as shown in [5], ϵ_g is decided by a privacy parameter ϵ (≥ 0) and the range of circular region centered at the user's location x. Essentially it means that geo-indistinguishability aims to protect this circular region with guaranteeing ϵ-differential privacy over it. Because the actual location is protected by being hidden among all the locations in the region due to their similar probability distributions for generating pseudo-locations, we call such region as the *protection region* and the set of locations within the region as the *protection location set*. Let D be the diameter of protection region and $\epsilon_g = \epsilon/D$, the mechanism is ϵ-differentially private for any two locations x and y in the protection region, i.e.,

$$e^{-\epsilon} \leq \frac{f(x'|x)}{f(x'|y)} \leq e^{\epsilon}. \tag{13}$$

Upper bound of posterior probability: Let Φ be the protection region. An upper bound of the posterior distribution of location $x \in \Phi$, given any observed pseudo-location x', can be obtained as follows:

$$\Pr(x|x') = \frac{\pi(x) f(x'|x)}{\sum_{y \in \mathcal{X}} \pi(y) f(x'|y)} \tag{14}$$

$$= \frac{\pi(x)f(x'|x)}{\sum_{y\in\Phi}\pi(y)f(x'|y) + \sum_{y\in\mathcal{X}\setminus\Phi}\pi(y)f(x'|y)} \quad (15)$$

$$\leq \frac{\pi(x)f(x'|x)}{\sum_{y\in\Phi}\pi(y)f(x'|y)} \quad (16)$$

$$= \frac{\pi(x)}{\sum_{y\in\Phi}\pi(y)f(x'|y)/f(x'|x)} \quad (17)$$

Applying (13),we have

$$\leq \frac{\pi(x)}{\pi(x) + e^{-\epsilon}\sum_{y\in\Phi, y\neq x}\pi(y)} \quad (18)$$

Since $0 < e^{-\epsilon} < 1$,we have

$$\leq e^{\epsilon}\frac{\pi(x)}{\sum_{y\in\Phi}\pi(y)} \quad (19)$$

The upper bound of posterior probability (19) implies that no matter what prior information the adversary has, geo-indistinguishability constrains the multiplicative distance between posterior distribution $\Pr(x|x')$ and prior distribution $\frac{\pi(x)}{\sum_{y\in\Phi}\pi(y)}$ within e^{ϵ}, and thus limits the posterior information gain of the adversary. This makes location obfuscation more robust against Bayesian adversary compared with the Bayesian mechanism [22] that could be constructed with incomplete knowledge about the adversary's prior information.

Lower bound of inference error: We further consider location privacy in terms of expected inference error. Let z be the estimated location by the adversary, i.e., $z = arg\min_{\hat{x}}\sum_{x\in\mathcal{X}}\Pr(x|x')d(\hat{x}, x)$. The conditional expected inference error is

$$\sum_{x\in\mathcal{X}}\Pr(x|x')d_p(z, x) \quad (20)$$

Here we consider the lower bound for it, which is indeed achieved in the worst case that the adversary narrows possible guesses to the location set within the protection region that contains the user's actual location. Therefore, the lower bound is

$$\min_{\hat{x}\in\mathcal{X}}\sum_{x\in\Phi}\frac{\Pr(x|x')}{\sum_{y\in\Phi}\Pr(y|x')}d_p(\hat{x}, x) \quad (21)$$

Let $z' = arg\min_{\hat{x}\in\mathcal{X}}\sum_{x\in\Phi}\frac{\Pr(x|x')}{\sum_{y\in\Phi}\Pr(y|x')}d_p(\hat{x}, x)$, the above becomes

$$= \sum_{x\in\Phi}\frac{\Pr(x|x')}{\sum_{y\in\Phi}\Pr(y|x')}d_p(z', x) \quad (22)$$

$$= \sum_{x\in\Phi}\frac{\pi(x)f(x'|x)}{\sum_{y\in\Phi}\pi(y)f(x'|y)}d_p(z', x) \quad (23)$$

Using (13), we have

$$\geq e^{-\epsilon}\sum_{x\in\Phi}\frac{\pi(x)}{\sum_{y\in\Phi}\pi(y)}d_p(z', x) \quad (24)$$

$$\geq e^{-\epsilon}\min_{\hat{x}\in\Phi}\sum_{x\in\Phi}\frac{\pi(x)}{\sum_{y\in\Phi}\pi(y)}d_p(\hat{x}, x) \quad (25)$$

where we have the derivation from (24) to (25) given that Φ is convex and thus the minimum is obtained when $\hat{x}$ is the weighted geometric median of Φ that lies in the region.

The bounds of posterior probability (19) and inference error (25) indicate the capability of geo-indistinguishability for defending against Bayesian inference attack (7) and optimal inference attack (6) respectively. Both of them depend on the prior distribution over protection region Φ, which suggests that geo-indistinguishability may not provide enough location protection against Bayesian adversary with sufficient prior information. The protection of geo-indistinguishability only measures the impact of user's location on the output, but not the inference capability of Bayesian adversary with his prior information. We have argued that certain prior knowledge to identify the user's location inherently exists, but geo-indistinguishable mechanisms produce pseudo-locations as if the adversary does not have any prior knowledge.

Also, we can see that it has limitations for geo-indistinguishable mechanisms in existing works [5], [6], [10], [11] to use uniform differential privacy parameter and protection region radius, independently of the user's locations. Because the prior distribution over protection regions around different locations are mostly different, geo-indistinguishablity may not achieve the same level privacy against Bayesian adversary, indicated by bounds (19) and (25) that change with priors. For example, in an urban area with many possible locations densely distributed, the user can use a small radius r for his protection region in which ϵ-differential privacy is achieved; but in a rural area, when the user's location is only possible location within it, using a small radius to generate a pseudo-location does not provide sufficient protection. This is indicated by that the upper bound (19) achieves maximum e^{ϵ} (≥ 1)and the lower bound (25) becomes zero, which actually means no bound for the posterior probability and inference error. Indeed, the adversary can easily associate the pseudo-location with the actual location given the prior knowledge that there is only one possible location in this area.

C. Experimental Illustration

In this section we evaluate the privacy of geo-indistinguishability against optimal and Bayesian inference attack to validate our analysis result in previous section. In order to see the difference between mechanisms built with two different privacy notions, we compare a geo-indistinguishable mechanism with the mechanism that is optimal against two attacks in terms of expected inference error. Specifically, we consider two following mechanisms:

- The optimal ϵ_g-geo-indistinguishable mechanism [6], denoted by M_{ϵ_g}, that minimizes the service quality loss (10) subject to geo-indistinguishability (12);
- The optimal Bayesian mechanism [22], denoted by M_B, that maximizes the expected inference error (8) under the constraint of the maximum tolerable service quality loss $Q_{loss}^{\max}$ for (10).

We choose them because it has been shown in [6] that M_{ϵ_g} and M_B can achieve the same level of location privacy in terms of expected inference error (8) defined with Euclidean distance, which enables us to make a fair comparison of them under optimal inference attack. To achieve that, given M_{ϵ_g} and the minimum quality loss q it obtains, M_B is derived with letting $Q_{loss}^{\max} = q$. Besides, we are particularly interested in the local performance of the mechanisms for protecting each location, rather than only considering the global average metrics as previous works [6], [22].

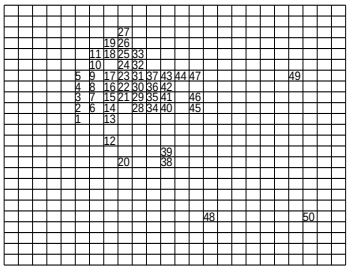


Fig. 1: The 50 regions in the location dataset.

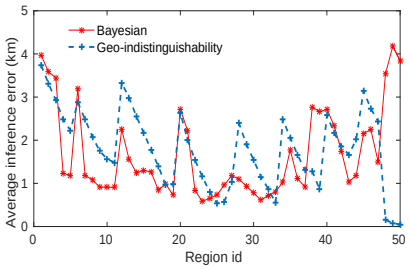


Fig. 2: The average inference error of optimal inference attack.

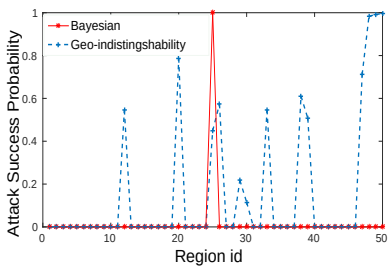


Fig. 3: The success probability of Bayesian inference attack.

In our experiment we use location data extracted from Geo-life dataset [26]–[28]. The details of data processing is described in Section VI. Specifically, here we use all-day location data of a single user with id 0, and consider 50 regions shown in Figure 1 as $\mathcal{X}$. The prior for the user is computed by counting and normalizing the number of his/her location points falling into each of 50 regions.

1) Optimal inference attack: We use $\epsilon_g = 0.9$ for M_{ϵ_g} that incurs minimum quality loss 0.89km. Letting $Q_{loss}^{\max} = 0.89$km, we follow the approach in [22] to obtain M_B with maximum expected inference error 0.89km. We simulate a user using two mechanisms at every region in $\mathcal{X}$, repeat the simulation 1000 times, and measure the inference error, i.e., the distance between the actual location and the location inferred by optimal inference attack (6), averaged over 1000 times for every region. The result is shown in Figure 2.

Though both M_{ϵ_g} and M_B can guarantee the expected inference error 0.89km at most locations, we can see that M_{ϵ_g} has significantly lower inference error at regions with id 48, 49 and 50 (almost zero for latter two), and M_B performs significantly better than M_{ϵ_g}. It indicates that geo-indistinguishability does not provide sufficient privacy protection for these locations against optimal inference attack. The essential reason is that geo-indistinguishability does not consider any prior distribution the adversary may have. As we can see from Figure 1, region 48, 49 and 50 are isolated locations on the prior distribution over $\mathcal{X}$, which means zero probabilities for any other locations in their neighborhood. Such skewed probability distribution lets the lower bound of expected inference error in (25) for these three regions to be zero, meaning no guarantee for location privacy in terms of expected inference error. Consequently, with minimization on the quality loss, M_{ϵ_g} has probability larger than 0.9 to report the actual locations at these regions.

In Figure 4, we vary ϵ_g from 0.7 to 0.1 and $Q_{loss}^{\max}$ from 1 to 2 and measure the expected inference error (8). We can see that both M_{ϵ_g} and M_B have the expected inference error to increase to 1.178 and remain the same after that. This value is exactly the value calculated by (11), which validates the upper limit (11). We note M_{ϵ_g} and M_B achieve this limit in different ways: M_{ϵ_g} always chooses the same region 25 as pseudo-location for any user's location, but M_B turns out to uniformly sample a location from $\mathcal{X}$ as the pseudo-location. In essence, both break the dependency between pseudo-locations and actual locations.

2) Bayesian inference attack: For Bayesian inference attack, we are interested in the probability that Bayesian inference attack makes correct guesses about the actual location for

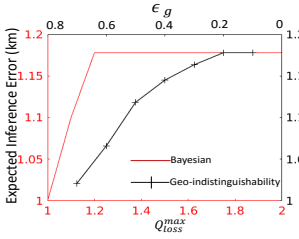


Fig. 4: Max Expected Error.

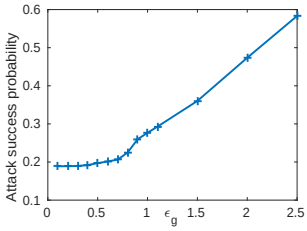


Fig. 5: ϵ_g v.s. P_s.

a user. We construct optimal Bayesian mechanism against this attack with using Hamming distance for expected inference error and the same $Q_{loss}^{\max} = 0.89$km. We replace optimal inference attack in the above simulation with Bayesian inference attack, repeat the simulation 1000 times, and calculate the percentage of successful guesses at every location. The result is shown in Figure 3. We can see that at most locations M_B has zero attack success probability, only at region 25 with probability one. It is because that the posterior distribution of M_B $\Pr(x|x')$ have maximum value at $x = 25$ for any x', which means that the adversary always guess 25 for any observed x'. Thus, when the user is at region 25, the attack success probability can be one. For M_{ϵ_g}, there are multiple locations with high attack success probabilities. In particular, we examine M_{ϵ_g} and find the mechanism has probability larger than 0.9 to report the true locations at region 48, 49 and 50. Also, given the observed pseudo-locations 48, 49 and 50, the posterior probabilities are close to one on the actual locations, which explains the high attack success probabilities (close to one) for M_{ϵ_g} at these regions in the figure. This is consistent with our analysis. The upper bound of posterior probability (19) on these isolated locations becomes e^{ϵ} (larger than 1), which means no bound for the posterior probability and thus it can get close to one on the actual locations.

Because the user has different probabilities to visit every region, we evaluate the expected success probability of Bayesian inference attack as follows:

$$P_s = \sum_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} \pi(x) f(x'|x) c(x, x') \tag{26}$$

where $c(x, x') = 1$ if the actual location x is correctly inferred given pseudo-location x', i.e., $x = \arg\max_{y \in \mathcal{X}} \Pr(y|x')$; otherwise $c(x, x') = 0$. We compute P_s for two mechanisms shown in Figure 3, and obtain $P_s = 0.19$ for M_B and $P_s = 0.27$ for M_{ϵ_g}. M_B that is constructed with prior information against Bayesian inference attack achieves better privacy than M_{ϵ_g}. We also measure the expected attack success probability for

M_{ϵ_g} with different privacy parameter ϵ_g, shown in Figure 5. Smaller ϵ_g indicates higher privacy. When ϵ_g gets close to zero, the multiplicative distance between posterior and prior distribution approaches to one, which means that the adversary cannot do better than just guessing the actual location by prior knowledge. That's why the curve becomes flat when ϵ_g approaches to zero in Figure 5. In contrast, when ϵ_g (as also ϵ) increases, the upper bound of posterior probability (19) increases, letting the mechanism truthfully report the locations with higher probabilities under quality loss minimization.

D. Our Design Objective

From our formal and experimental analysis, we can see that geo-indistinguishability limits the privacy leakage by bounding the relative information gain of the adversary given observed pseudo-locations, regardless of what kind of prior information the adversary may have. But it does not ensure absolute location privacy guarantee in terms of expected inference error against inference attacks (6) and (7). Bayesian optimal mechanisms protect location privacy by maximizing expected inference error against inference attacks but require assuming a prior location distribution that the adversary has, which is not robust against adversaries with arbitrary knowledge. Thus, it is desirable to have both privacy notions in a location obfuscation mechanism. On the other hand, existing geo-indistinguishable mechanisms suppose uniform differential privacy parameters over every location, which may either cause unnecessarily large noise level at some locations or insufficient noise level at others leading to privacy disclosure, and their formulations do not provide the user a straightforward way to customize his privacy preference for his current location. Considering these issues, we aim to design a location obfuscation mechanism that can effectively combine geo-indistinguishability and expected inference error, while operating adaptively with supporting customizable privacy preferences for the users.

V. Our Solution Approach

In this section we describe PIVE, a two-phase dynamic approach to protect location privacy in terms of both geo-indistinguishability and expected inference error. We first present the PIVE two phase location obfuscation framework and then describe each phase in detail. In the first phase, we determine a set of locations (i.e., protection location set) to protect user's actual location, with guaranteeing the expected location inference errors with the user-defined threshold and the adversary's prior knowledge with respect to the user's location. we develop a Hilbert curve based method and its optimization for efficiently and accurately determining the protection location set. In the second phase, we devise a differentially private mechanism to generate pseudo-locations with strong utility guarantee with respect to the service quality.

A. PIVE Two-Phase Framework

Our goal is to design a mechanism that achieves geo-indistinguishability while providing lower bound on expected inference error against optimal inference attacks. A challenging problem is how to integrate both privacy notions to a mechanism designed to obfuscate locations instantly and adaptively. Basically, our solution is to dynamically choose a protection location set to guarantee expected inference error and produce pseudo-locations in a differentially private way for every location in the set.

To introduce our approach, we first define ϵ-differential location privacy over an arbitrary region containing the actual location, as opposed to geo-indistinguishability that is defined over the circular neighborhood centered at the actual location.

Differential privacy requires that a query function has unsubstantial difference for the outputs over any two neighboring datasets that differ only in a single element. The location obfuscation mechanism for a user only involves a single data record, i.e., his current location. Differentially private location obfuscation requires the definition of "neighboring" location points to the user's location, such that they have the similar probabilities to produce a pseudo-location. The neighborhood consisting of all "neighboring" locations indeed functions as "a minimum crowd" for the actual location to "be hidden in a crowd". As mentioned in Section IV-B, previous geo-indistinguishabile mechanisms [5], [6] actually regard the circular region centered at the user's location with a uniform radius as such neighborhood for protection. In this paper, we define "neighboring" relationship over a set of locations in an arbitrary region that contains the user's actual location, referred to as *protection location set*, and accordingly differentially private location obfuscation is defined as follows:

Definition 4: A randomized location obfuscation mechanism $f(\cdot|\cdot)$ satisfies ϵ-differential privacy on protection location set Φ, if for any locations $x, y \in \Phi$, and any output x',

$$\frac{f(x'|x)}{f(x'|y)} \leq e^{\epsilon} \tag{27}$$

Based on the upper bound of posterior probability (19), here ϵ is chosen to achieve a desired bound of the multiplicative distance between posterior distribution and prior distribution, to limit the adversary's posterior information gain.

Then, we consider how to guarantee the expected inference error via protection location set. Let

$$ExpEr(x') = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \Pr(x|x') d(\hat{x}, x) \tag{28}$$

$$E(\Phi) = \min_{\hat{x} \in \Phi} \sum_{x \in \Phi} \frac{\pi(x)}{\sum_{y \in \Phi} \pi(y)} d(\hat{x}, x) \tag{29}$$

where we choose Euclidean distance as privacy metric as previous works [6], $ExpEr(x')$ is the conditional expected inference error given any observed pseudo-location x'. For optimal inference attack with using x', according to the lower bound result for expected inference error in (25), we have

$$ExpEr(x') \geq e^{-\epsilon} E(\Phi) \tag{30}$$

To ensure a lower bound for conditional expected inference error $ExpEr(x')$, we introduce privacy parameter E_m that is specified by the user according to his current location's sensitivity, such that $\forall x'$, $ExpEr(x') \geq E_m$. To ensure that, it is sufficient to satisfy that

$$E(\Phi) \geq e^{\epsilon} E_m \tag{31}$$

Then, we have the following theorem (with the above as a proof):

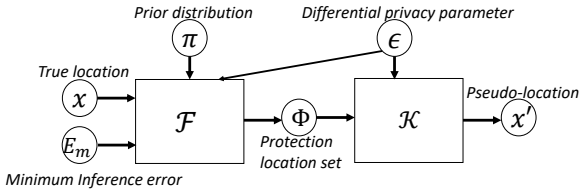


Fig. 6: The framework of PIVE. $\mathcal{F}$ is the algorithm to determine the protection location set Φ and $\mathcal{K}$ is the differential mechanism to produce a pseudo-location.

Theorem 1: For a location obfuscation mechanism that achieves ϵ-differential privacy on protection location set Φ, if $E(\Phi) \geq e^{\epsilon}E_m$, the optimal inference attack using any observed pseudo-location x', $ExpEr(x') \geq E_m$.

Based on (31), we regard Φ as a variable and propose to dynamically search a region of Φ where the user is located to satisfy $E(\Phi) \geq e^{\epsilon}E_m$. Then, with the protection location set Φ, we propose an exponential mechanism that generates a pseudo-location in the way that achieves ϵ-differential privacy on Φ, defined in Definition 4. Because the maximum change of the user's location is within the range of Φ, the sensitivity method to achieve differential privacy introduces the noise perturbation proportional to Φ's diameter $D(\Phi)$, as shown in Section V-C. To maximize the utility, the noise magnitude should be minimized, and thus it is desired to find Φ that satisfies (31) with a minimum diameter.

Figure 6 illustrates the workflow of PIVE. It shows two components with their inputs: the algorithm $\mathcal{F}$ for generating the protection location set and the differentially private mechanism $\mathcal{K}$ for producing a pseudo-location. In essence, via protection location set that is determined with prior distribution π and $e^{\epsilon}E_m$, PIVE achieves differential privacy while guaranteeing a lower bound for the adversary's expected inference error. This framework offers adaptive location protection for users according to their current locations and requirements on two privacy notions (expressed by E_m and ϵ), and the latest prior distribution the adversary could have known (e.g., by inference with the mobility model of users). Previous geo-indistinguishable mechanisms [5], [6] can be regarded as the special cases of our framework with $\mathcal{F}$ using the circular neighborhood with a fixed radius as the protection location set without considering any prior distribution and inference error bound.

PIVE provides two privacy control knobs: 1) the minimum inference error E_m and 2) the differential privacy parameter ϵ. Through these parameters, we allow users to define their desired privacy preferences at different locations. The min error parameter E_m aims to bound the expected inference error in the worst case. The differential privacy parameter ϵ allows users to constrain the posterior information leakage via the provisioning of differential privacy. Given that ϵ-differential privacy is the property of the random mechanism $\mathcal{K}$ producing pseudo-locations, one possible way for a user to set these two parameters is to use a fixed ϵ for $\mathcal{K}$ and set E_m according to this user's tolerance estimation on the lowest bound of expected inference error of the adversary against the protection region, for example, E_m=0.1km.

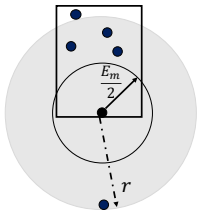


Fig. 7: The non-optimal case for the simple approach

B. Determining Protection Location Set

Given the user's location x, the problem is how to efficiently determine its protection location set Φ ($x \in \Phi$) that satisfies $E(\Phi) \geq e^{\epsilon}E_m$ with a diameter as small as possible. Meanwhile, we note the diameter of the protection location set cannot be less than $e^{\epsilon}E_m$, given by the following theorem.

Theorem 2: Let $D(\Phi)$ be the diameter of protection location set Φ that is the largest distance between any two locations in Φ. If $E(\Phi) \geq e^{\epsilon}E_m$, we have $D(\Phi) \geq e^{\epsilon}E_m$.

Proof: $D(\Phi) \geq d(\hat{x}, x)$ for $\forall \hat{x}$, x in Φ, so

$$e^{\epsilon}E_m \leq E(\Phi) \leq \min_{\hat{x}\in\Phi} \sum_{x\in\Phi} \frac{\pi(x)}{\sum_{y\in\Phi}\pi(y)} D(\Phi) = D(\Phi)$$

■

A simple way to determine the protection location set is to gradually increase the radius of circular region centered at the user's location from $e^{\epsilon}E_m/2$ (given by Theorem 2) until it satisfies $E(\Phi) \geq e^{\epsilon}E_m$. However, this approach may produce unnecessarily large diameter, leading to significant service quality loss. Figure 7 shows an example in which the desired protection location set is obtained by increasing radius to r to include four location points, resulting in $2r$ diameter. However, in this way we cannot find another qualified set that is the rectangle area in the figure with a much smaller diameter.

Hilbert Curve based Search: To efficiently search over the plane for the protection location set, we propose a Hilbert curve based search algorithm. Hilbert curve [16] is a popular member in the family of space-filling curves. It provides a mapping from a data point in a 2-D space to a point in one dimensional space that preserves the proximity of data. That is, points which are close to one another in the 2-D space will also remain close to each other in the transformed 1-D space. It has been shown that Hilbert curves have the superior distance preserving properties [18]. Figure 8 shows the Hilbert curves for 4×4 and 16×16 grids in 2-D space. The Hilbert curve maps a location point x to a 1-D value denoted by $H(x)$. We call $H(x)$ as the Hilbert value of x. The locations in $\mathcal{X}$ is sorted by their Hilbert values, and the rank of a location x in the sorted $\mathcal{X}$ is denoted by $R(x)$.

Given user's location x, our algorithm searches the neighborhood of x along the Hilbert curve to find a protection location set Φ that satisfies $x \in \Phi$ and $E(\Phi) \geq e^{\epsilon}E_m$. The basic search strategy in the algorithm can be generally described as follows. Let $x_{-l}, x_{-l+1}, \ldots, x_0(= x), x_1, x_2, \ldots, x_r$ be the sequence of locations in the searching neighborhood of x along the Hilbert curve, sorted by their Hilbert values. For each x_i ($-l \leq i \leq 0$), the algorithm checks every interval from x_i to x_j for $0 \leq j \leq r$ in the sequence, denoted by $[x_i, x_j]$, and evaluate

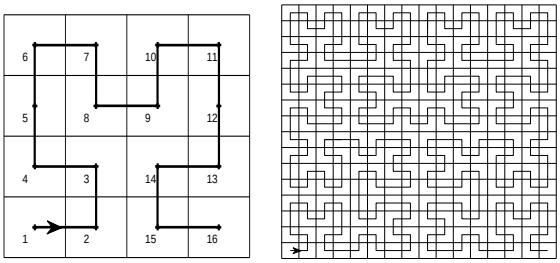


Fig. 8: Hilbert Curve for 4×4 and 16×16 grid

Algorithm 1: Protection Location Set Search Algorithm

Input: x: user's location E_m: error bound, ϵ: privacy parameter
if $\pi(x) = 0$ **then**
  $S \leftarrow \{l \mid H(l) \in [H(x) - range, H(x) + range]$ on $H\}$;
else
  $S \leftarrow \{l \mid H(l) \in [R(x) - range, R(x) + range]$ on sorted $\mathcal{X}\}$;
Let S be $x_{-l}, x_{-l+1}, \ldots, x_0 = x, x_1, x_2, \ldots, x_r$;
$L \leftarrow \emptyset$;
for *i from −l to 0* **do**
  for *j from 0 to r* **do**
    $\Phi = \{x_k \mid i \le k \le j\}$;
    Calculate $E(\Phi)$ by(29) ;
    if $E(\Phi) \ge e^{\epsilon} E_m$ **then**
      Add Φ to L;
      break ;
return a set having the smallest diameter in L;

$E([x_i, x_j])$ by (29). Once an interval that has $E([x_i, x_j]) \ge e^{\epsilon} E_m$ is found for x_i, the algorithm stops interval check for x_i, adds location set in $[x_i, x_j]$ to a candidate list, and repeats with the next x_i. Finally, the set having the smallest diameter in the list is returned, with breaking ties by a random choice.

In the case that all locations in Φ have zero prior probabilities, i.e., $\sum_{y \in \Phi} \pi(y) = 0$ in (29), we define $\frac{\pi(x)}{\sum_{y \in \Phi} \pi(y)} = \frac{1}{|\Phi|}$, because a uniform distribution is assumed in an area when the adversary does not have any prior information about it. Accordingly, the algorithm searches Φ for locations with zero and non-zero prior probability in $\mathcal{X}$ in different ways. Because over the plane any locations outside $\mathcal{X}$ (e.g., the un-numbered regions in Figure 1) indeed have zero prior probabilities, the protection location set Φ for the user's location x with $\pi(x) = 0$ can involve them with $E(\Phi)$ being computed in the defined way. Thus, the searching range is determined as all the locations on the plane with Hilbert values in a range $[H(x) - range, H(x) + range]$. For the protection location set of x with $\pi(x) > 0$, the locations with zero prior probabilities contribute zero to $E(\Phi)$ in (29) and thus the locations outside $\mathcal{X}$ are not considered. The searching range is defined as the locations with ranks in $[R(x) - range, R(x) + range]$ over the sorted sequence of $\mathcal{X}$. The algorithm applies the search strategy mentioned above to the searching range to obtain the protection location set. The pseudo-code is given in Algorithm 1.

The *range* must be large enough to have better chance to find a qualified protection location set. Given $T = e^{\epsilon} E_m$ and Theorem 2, *range* can be decided heuristically. We can traverse along the Hilbert curve in both directions from user's location x. Once reaching the locations a and b in each direction with their distances to x being some multiple of T, we set *range* = $\max(|H(a) - H(x)|, |H(b) - H(x)|)$. In our implementation we simply choose sufficiently large range that incurs low failure rate for finding protection location set. We can also specify an upper bound for *range* to limit the searching cost and avoid large region that causes unacceptable quality loss. Note the algorithm may not find any qualified protection location set, for example, in the case that the user's current location is only possible location for him and all other locations has zero prior probabilities on the plane. Thus, if an empty set is returned, which indicates the location privacy cannot be protected, the user can choose to suppress location report.

Improvement with Multiple Rotated Hilbert Curves: Although using Hilbert curve enables efficient search over 2-D plane, a drawback is that the search is conducted along a single direction and the searched regions can only be ones that consist of neighboring locations on the curve. A cell actually can have four neighbors on the plane while two neighboring cells may be far apart on the curve (e.g., location 2 and 15 in Figure 8). Since there are regions where locations are not adjacent on the curve, we propose to use multiple different Hilbert curves to connect locations in different ways such that more possible regions can be involved, which can improve the chance to find the protection location set with a smaller diameter. Previous works have utilized multiple Hilbert curves to improve the quality of k-Nearest Neighbor queries [15] and reduce cloaking area [19]. In PIVE, given a Hilbert curve H over $2^n \times 2^n$ grid, other three Hilbert curves are generated by rotating it 90, 180, 270 degrees clockwise about the center point. We use Algorithm 1 to find the protection set of the user's location x for each Hilbert curve, and choose the one with smallest diameter among four results.

C. Differentially Private Mechanism

Given the protection location set Φ, PIVE achieves differential privacy on it through the exponential mechanism [17]. Considering the set $\mathcal{X}$ as the output range of location obfuscation, the utility of output x' is measured by the distance between x' and user's location x in Φ. Smaller distance has higher utility. As the protection location set decides "neighboring" locations to the user's location, the sensitivity of the utility function u is

$$\Delta u = \max_{x' \in \mathcal{X}} \max_{x,y \in \Phi} |d(x, x') - d(y, x')| \tag{32}$$

It is easy to see, according to triangle inequality, for any $x, y \in \Phi$, $|d(x, x') - d(y, x')| \le d(x, y) \le D(\Phi)$, so $\Delta u = D(\Phi)$ where $D(\Phi)$ is the diameter of Φ.

Exponential mechanism $\mathcal{K}$: Given the user's location x and location protection set Φ, the exponential mechanism $\mathcal{K}$ selects and outputs a location $x' \in \mathcal{X}$ with probability proportional to $\exp(\frac{-\epsilon d(x,x')}{2D(\Phi)})$.

The mechanism $\mathcal{K}$ samples each location x' from $\mathcal{X}$ with the probability $w_x \exp(\frac{-\epsilon d(x,x')}{2D(\Phi)})$ where w_x is the normalization factor for the probability distribution over $\mathcal{X}$,

$$w_x = 1 \Big/ \Big(\sum_{x' \in \mathcal{X}} exp(\frac{-\epsilon d(x, x')}{2D(\Phi)}) \Big) \tag{33}$$

Following the proof of Theorem of McSherry and Talwar [17], we can easily obtain the theorem below,

Theorem 3: The exponential mechanism $\mathcal{K}$ preserves ϵ-differential privacy on the protection location set Φ.

Proof:

$$\begin{aligned}\frac{f(x'|x)}{f(x'|y)} &= \frac{w_x \exp\Big(-\epsilon d(x,x')/(2D(\Phi))\Big)}{w_y \exp\Big(-\epsilon d(y,x')/(2D(\Phi))\Big)} \\ &\le \frac{w_x}{w_y} e^{\epsilon|d(x,x')-d(y,x')|/(2D(\Phi))} \le \frac{w_x}{w_y} e^{\epsilon d(x,y)/2D(\Phi)} \\ &\le \frac{w_x}{w_y} e^{\epsilon/2} \le \frac{\Big(\sum_{x'\in\mathcal{X}} exp(\frac{-\epsilon d(y,x')}{2D(\Phi)})\Big)}{\Big(\sum_{x'\in\mathcal{X}} exp(\frac{-\epsilon d(x,x')}{2D(\Phi)})\Big)} e^{\epsilon/2} \\ &\le \frac{\Big(\sum_{x'\in\mathcal{X}} exp(\frac{-\epsilon(d(x,x')-D(\Phi))}{2D(\Phi)})\Big)}{\Big(\sum_{x'\in\mathcal{X}} exp(\frac{-\epsilon d(x,x')}{2D(\Phi)})\Big)} e^{\epsilon/2} \\ &\le \frac{\Big(\sum_{x'\in\mathcal{X}} exp(\frac{-\epsilon d(x,x')}{2D(\Phi)})\Big)}{\Big(\sum_{x'\in\mathcal{X}} exp(\frac{-\epsilon d(x,x')}{2D(\Phi)})\Big)} e^{\epsilon/2} e^{\epsilon/2} \le e^{\epsilon}\end{aligned} \tag{34}$$

■

The exponential mechanism provides strong utility guarantees since it discounts the pseudo-locations exponentially quickly as their distances to the actual location increase. To see that, we have the following theorem

Theorem 4: Given the actual location x, let x' be the pseudo-location randomly sampled from $\mathcal{X}$ by the exponential mechanism $\mathcal{K}$, with probability at least $1-\delta$ we will have

$$d(x,x') \le \frac{2D(\Phi)}{\epsilon}\Big(\ln|\mathcal{X}| + \frac{\epsilon}{2} - \ln|\Phi| - \ln\delta\Big) \tag{35}$$

Proof: For any x' that has $d(x,x') \ge c$, the probability it is sampled with is at most $w_x \exp(\frac{-\epsilon c}{2D(\Phi)})$. Thus, the total probability of $d(x,x') \ge c$ for all x' is at most $w_x|\mathcal{X}|\exp(\frac{-\epsilon c}{2D(\Phi)})$. On the other hand,

$$\begin{aligned} w_x &= 1\Big/\Big(\sum_{x'\in\Phi} exp(\frac{-\epsilon d(x,x')}{2D(\Phi)}) + \sum_{x'\in\mathcal{X}\setminus\Phi} exp(\frac{-\epsilon d(x,x')}{2D(\Phi)})\Big) \\ &\le 1\Big/\Big(\sum_{x'\in\Phi} exp(\frac{-\epsilon d(x,x')}{2D(\Phi)})\Big) \le 1\Big/\Big(\sum_{x'\in\Phi} e^{-\epsilon/2}\Big) \\ &= \frac{e^{\epsilon/2}}{|\Phi|}\end{aligned}$$

Thus, we have $\Pr(d(x,x') \ge c) \le \frac{|\mathcal{X}|e^{\epsilon/2}}{|\Phi|}\exp(\frac{-\epsilon c}{2D(\Phi)})$. Let δ be the right-hand side and we can derive (35). ■

Because the searching range is limited in Algorithm 1, $\ln|\Phi|$ is bounded by $\ln(2range)$ that is a small constant (e.g, at most 4 in our experiment), while $D(\Phi)$ can vary a lot given possible sparse location distribution. Therefore, the value of right hand side of (35) mainly depends on $\frac{D(\Phi)}{\epsilon}$. Given fixed E_m, increasing ϵ can incur a protection location set with a larger diameter since $D(\Phi) \ge e^{\epsilon}E_m$ given by Theorem 2. Because $D(\Phi)$ increases exponentially with ϵ, with increasing ϵ from the value close to zero, $\frac{D(\Phi)}{\epsilon}$ will decrease first and then increase. Therefore, we expect that the service quality and also location privacy will exhibit the similar changing pattern, which is demonstrated in our evaluation.

VI. Evaluation

In this section we first evaluate the performance of our PIVE mechanism, and compare PIVE approach with other mechanisms on location privacy and service quality. Our evaluation shows that PIVE effectively combines two privacy notions, and efficiently addresses the issues of existing location obfuscation mechanisms.

We use the dataset provided by authors of [6]. The dataset was extracted from the GeoLife GPS Trajectories dataset [26]–[28], which contains 17621 traces collected from 182 users in Beijing, China, during a period of over five years. The traces record users outdoor movements with locations being logged every 1-5 seconds or every 5-10 meters. The details of data processing can be found in [6] and here we provide a brief description. The map of Beijing is divided into a grid of regions 0.658km wide and 0.712km high, the 50 "most popular" regions of the grid is used as the set of all locations $\mathcal{X}$, as shown in Figure 1, and the users who have few recorded points for each time period at these regions are filtered out. The final dataset contains 84 users. The prior for each user is computed by counting and normalizing the number of points falling in each of 50 regions with in different time periods (all day, morning, afternoon and night). In this paper we use all-day prior to construct mechanisms. In order to demonstrate the performance in a single user setting, at default we always choose the user with id 0, as in Section IV-C.

A. Performance of Protection Location Set Search

Given ϵ and E_m, a threshold $T = e^{\epsilon}E_m$ is determined and Algorithm 1 searches a location protection set Φ for user's location that has $E(\Phi)$ in (29) no less than T while with the smallest diameter. In this section we study the performance of our search algorithm in terms of the diameter $D(\Phi)$ and value $E(\Phi)$. In the algorithm, we choose sufficient large *range*=50 at default. The searching range *range* decides the chance to find a qualified location protection set for a location. To see its impact, we test *range* with values 20, 40, 50 and 60 for $T = 2$ that is corresponding to the largest average diameter in Figure 9. The number of regions for which the algorithm fails to find qualified protection location set is 19, 8, 6, 6 for each *range* value respectively. We can see that from *range*=50 that is the size of $\mathcal{X}$, the number of such regions remains to be 6. Smaller *range* 40 has approximate number of failures as 50. Since the size of 50 regions is small, in our experiment we choose 50 that incurs smallest number of failures. Within a large size of $\mathcal{X}$, *range* can be a relatively smaller value.

We vary T from 0.1 and 2.0 and measure the diameters of protection (location) sets obtained by our algorithm for user's location at each of 50 regions. The results are shown in Figure 9 where the whiskers represents minimum and maximum diameters in each group. It is clear that the average diameter of all regions increases with the threshold T. The diameters for isolated region 48, 49, and 50 remain between 4km and 5km under different T. They are maximum ones in the results from T=0.1 to 0.8. For some regions like 24, 25, 32, 33 and 34, the diameters become higher than 12km from T=1.2. From T=1.4 to 2.0, the algorithm cannot find qualified protection sets for regions like 24, 25, 26, 32 and 33. Figure 10 shows the corresponding $E(\Phi)$ values of obtained protection sets for every region under different T. As we can see, the average

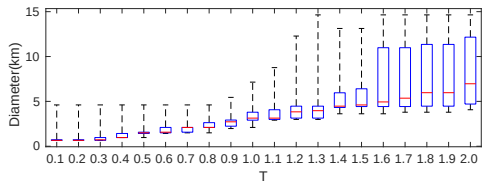


Fig. 9: Boxplot of Diameter with different T

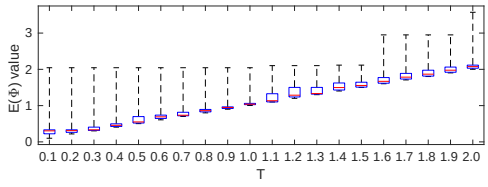
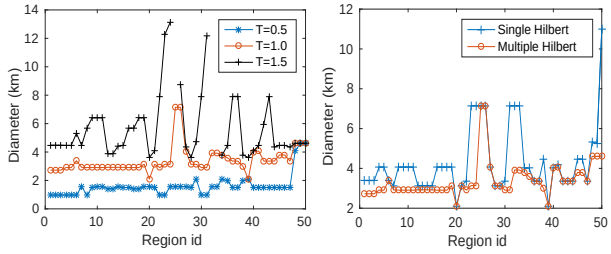


Fig. 10: Boxplot of $E(\Phi)$ with different T



(a) Diameter of Φ for regions (b) Single V.S. multiple Hilbert

Fig. 11: Diameter of protection location set for every region.

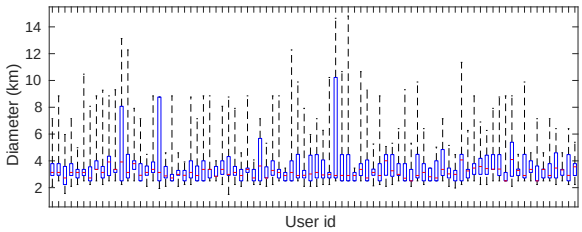
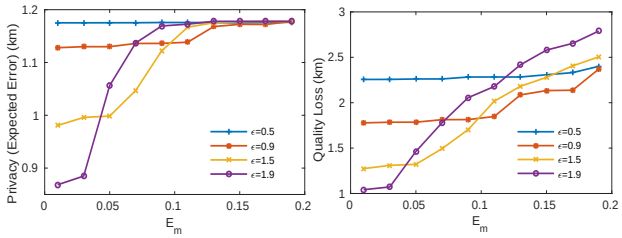


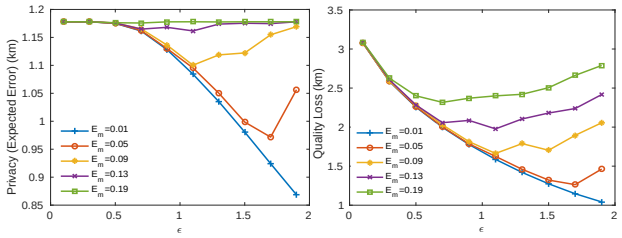
Fig. 12: The diameters under $T = 1$ for every user.

$E(\Phi)$ value increases linearly with T and is approximate to T. This indicates that our algorithm effectively finds the qualified protection location set with $E(\Phi) \geq T$ as desired. We also observe that the maximum $E(\Phi)$ for each T is about 2km from T=0.1 to 1.5, which is because that the protection set for region 49 always has maximum $E(\Phi)$ 2km. By further looking into the results, we find that for both region 49 and 50, the protection location set remains the same from T=0.1 to 1.5, resulting the same diameter and $E(\Phi)$. Region 49 always has protection set $\{47, 49\}$, and 50 has $\{48, 50\}$. From Figure 1, we can see the reason is that they are isolated regions and their nearest neighbors are 47 and 48 respectively that provide qualified protection sets. For $T > 1.5$, region 50 has to involve another region 45 to satisfy $E(\Phi) \geq T$.

To see how the diameter of protection location set varies among different regions, we show the results of T=0.5, 1.0 and 1.5 in Figure 11a. It is clear that the diameter of protection location set for each region increases with T. The curve is discontinuous at some points for T=1.5 because the algorithm cannot find qualified set at those locations. The diameters for regions 49 and 50 remains the same with three different T due to the reasons mentioned above.



(a) Privacy V.S. E_m (b) Quality Loss V.S. E_m



(c) Privacy V.S. ϵ (d) Quality Loss V.S. ϵ

Fig. 13: Impact of privacy parameters ϵ and E_m

Improvement with Multiple Hilbert curves: Our algorithm utilizes multiple Hilbert curves that are generated by the rotation of the original Hilbert curve to find protection location set with the smallest diameter. To see the effectiveness of such improvement, we compare the diameter of every region with the search algorithm using one single Hilbert curve and multiple ones respectively, under a given T. Figure 11b shows the result with T=1.0. We can see that using multiple Hilbert curves effectively reduces the diameter of protection location set. At some regions the improvement is significant. For example, the diameter is reduced by more than half at region 31 and 50. Such improvement holds for different T values.

We further investigate the diameters for all 84 users in the dataset and show the result with T=1 in Figure 12. All users have approximate average diameters between 2km and 4km, but the maximum diameter for some users can be as large as 14km. Large diameter will incur significant noise and extremely low utility. To avoid that, a maximum tolerable diameter D_m can be specified in the mechanism, such that the mechanism can use the location set with maximum $E(\Phi)$ among those with diameters no larger than D_m if the diameter of the produced protection location set exceeds D_m.

B. Location Privacy and Service Quality

In this section, we evaluate the impact of differential privacy parameter ϵ and inference error threshold E_m on location privacy and service quality under a single user setting. Although PIVE allows different privacy parameters at different locations, we use uniform parameters over all locations and unconditional expected inference error (8) and quality loss (10) as privacy and quality metric, in order to examine the effects of different ϵ and E_m on the performance.

Figure 13a and 13b show that under different ϵ, both location privacy and quality loss monotonically increase with E_m. This is because that higher E_m leads to larger diameter of the protection location set, and the pseudo-location is more likely to be further from the actually location and thus incurs

lower utility, which is indicated by Theorem 4. The monotonic relationship between E_m and the corresponding location privacy (i.e., expected inference error) indicates that E_m is an effective control knob to guarantee the expected inference error. The difference in order of magnitude between E_m and corresponding expected inference error is because E_m is the lower bound of the expected inference error given any pseudo-locations in the worst case that the adversary have identified the protection set. Therefore, E_m should be determined with consideration of such worst case to protect location privacy in terms of unconditional expected inference error.

For smallest ϵ=0.5 indicating the strongest privacy guarantee, $e^{\epsilon}E_m$ increases linearly with E_m with a small factor e^{ϵ}, that is to say, the impact of diameter changes on the privacy and quality is much smaller compared with that of ϵ. In contrast, under larger ϵ like 1.9 that incur weak requirement for differential privacy, $E(\Phi)$ increases with E_m with a much larger factor e^{ϵ}, which incurs larger diameter variance. Therefore, E_m has more significant impact on location privacy and quality loss for ϵ=1.9, indicating by its highest curve steepness in Figure 13a and 13b. Also, in Figure 13a, location privacy for different ϵ increases to the same upper limit 1.178 as in Section IV-C1. ϵ=0.5 achieves this limit regardless of E_m. Other cases have location privacy approximate to the upper limit starting from E_m=0.1. Therefore, we can choose E_m no larger than 0.1 for improving utility. Accordingly, in our comparison experiment we focus on E_m=0.05 and 0.09.

We further examine the impact of ϵ on location privacy and quality loss with given E_m. The results are shown in Figure 13c and 13d. We can see that the relationship between ϵ and location privacy as well as quality loss is not monotonic. Location privacy and quality loss first decrease with ϵ and then increase. This result confirms our discussion following Theorem 4. The reason is that, at first ϵ takes control of location privacy and quality loss, and thus increasing ϵ incurs lower location privacy and quality loss. As the diameter increases exponentially with ϵ, the diameter takes effects, thus increasing ϵ causes higher privacy and quality loss. Comparing Figure 13c and 13d, we can see that the turning points of both metrics under the same E_m occurs at the same ϵ values.

C. Comparison with other mechanisms

In this section we compare PIVE with typical geo-indistinguishable mechanisms to verify the advantage of introducing inference error bound. Because PIVE focuses on local performance of privacy protection for every region rather than the global average performance examined in previous works [6], [20], [22], we compare PIVE with other mechanisms mostly in a single user setting, in order to check the privacy protection performance at each individual region. We examine the performance of PIVE for every user, and only show results with regard to user with id 0 due to the similar behaviors of these mechanisms for other users. It is worth to note that PIVE provides the users a way to specify different privacy requirements for different locations through two privacy parameters E_m and ϵ. Given that the existing mechanisms like optimal geo-indistinguishable mechanism do not support different privacy specifications for different locations, we set the same privacy parameters everywhere for PIVE in order to make meaningful comparisons.

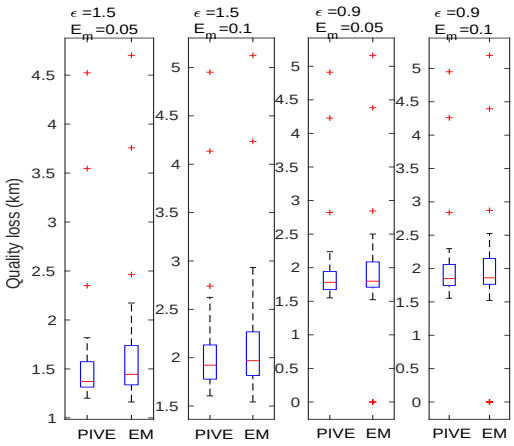
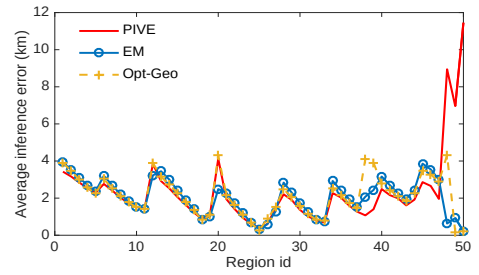


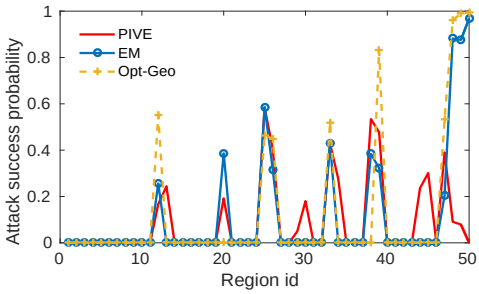
Fig. 14: Boxplot of quality loss of 84 users

We first consider an exponential mechanism, named as EM, that is like the one proposed in PIVE except using uniform constant diameter for every location's protection location set. It represents geo-indistinguishable mechanisms like discrete Planar Laplace Mechanism that ensure ϵ-differential privacy in the circular neighborhood centered at the user's location. To make a fair comparison, for each user, we run PIVE with different ϵ and E_m, and obtain its location privacy (i.e., expected inference error (8)). Then, given the same ϵ, we derive EM by choosing the diameter to achieve the same location privacy. To deal with floating point comparisons, two values with less than 0.005 difference are regarded to be equal for location privacy. Figure 14 shows boxplots of the quality losses of all users for PIVE and EM respectively under different pairs of ϵ and E_m. In each subfigure, we can see that overall PIVE achieves smaller quality loss than EM, though they have the same location privacy. This is because that PIVE adaptively determines protection location sets to implement geo-indistinguishability but EM uses protection regions of uniform radius everywhere. At some locations with sufficient number of possible locations in their neighborhood, PIVE can use smaller diameters than at locations in sparse areas for providing the same level of location privacy. Comparing these subfigures, we can see that lower ϵ or higher E_m, both indicating higher privacy requirements, incur larger quality loss.

We further look into the level of privacy protection for a single user at every region, with using the same simulation approach described in Section IV-C1. Suppose E_m = 0.05 and ϵ = 1.5 for PIVE. We derive EM with ϵ = 1.5 and also the optimal geo-indistinguishable mechanism Opt-geo (M_{ϵ_g} in Section IV-C) with ϵ_g = 0.7 such that they achieve the same expected inference error as PIVE. Figure 15a and 15b show the average error of optimal inference attack and success probability of Bayesian inference attack against each region for three mechanisms. We can see that both E_M and Opt-geo show insufficient protection for regions like 48, 49 and 50. These weak regions have zero inference error for the optimal inference attack and high success probability for Bayesian inference attack. PIVE is resilient to such vulnerable cases due to the skewed probability distribution on these isolated regions by finding a sufficient protection location set and ensuring the lower bound of inference error in the worst cases. That is the reason of why PIVE has much larger inference error and



(a) Optimal inference attack



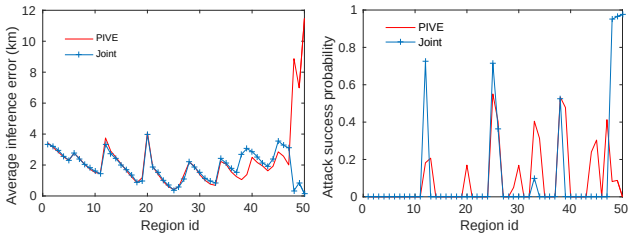
(b) Bayesian inference attack

Fig. 15: Comparison of local privacy protection at every region

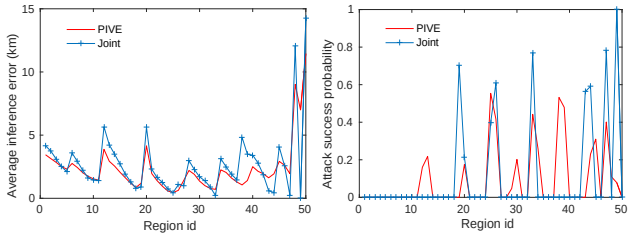
approximate zero attack success probabilities at these most vulnerable locations. Furthermore, with PIVE, the Bayesian inference attack success probability is capped to be no more than 60% as shown in Figure 15b. Comparing the service quality losses of three mechanisms, we have EM=1.49 > PIVE=1.32 > Opt-geo=1.02. PIVE has smaller quality loss than EM, which has been explained above, and Opt-geo achieves smallest quality losses due to its global optimization on service quality.

Next, we compare PIVE with the joint optimization mechanism [20] in terms of effectivess for privacy protection by combining geo-indistinguishability and expected inference error. We choose the same parameters for PIVE as in the previous expeirments, and then use its expected inference error as the minimum desired distortion privacy level d_m for constructing the optimal joint mechanism. ϵ_g in the joint mechanism is chosen to achieve the same location privacy as PIVE in terms of unconditional expected inference error. Figure 16a shows the average inference error and success probability for two inference attacks respectively at each region, with d_m=0.9986km and ϵ_g=0.8. It can be seen that (1) the joint mechanism and PIVE exhibit similar performance at most locations with small variation; and (2) the joint mechanism incurs the weak regions, e.g., region id 48, 49 and 50, against inference attacks, despite having bound on global expected distortion metric. These weak regions represent some skewedness as they are far away from the rest of the regions. Concretely, PIVE has the attackers average inference error bounded to be no lower than 0.22 and at the same time the Bayesian inference attack success probability capped to be no higher than 60%. In comparison, the joint optimization achieves good privacy in most of the locations but fail to avoid the worst case scenarios when the location dataset contains some skewed locations.

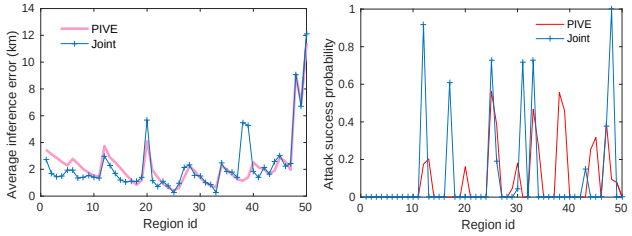
Given d_m=0.9986km, we vary ϵ_g from 0.8 to 2.2. The resulting joint mechanism achieves the same location privacy 0.9986km as PIVE from 0.8 to 2.1, and 0.9997km at ϵ_g=2.2.



(a) ϵ_g = 0.8 for joint mechanism



(b) ϵ_g = 2.1 for joint mechanism



(c) ϵ_g = 2.2 for joint mechanism

Fig. 16: Comparison of PIVE and joint mechanism under optimal inference attack (left column) and Bayesian inference attack (right column)

When ϵ_g increases, the behaviors of the joint mechanism get close to the optimal Bayesian mechanism, that is being more resilient to the skewed locations. The reason is that the linear program of the joint mechanism is equal to the optimal geo-indistinguishable mechanism with minimum distortion constraint, or the optimal Bayesian mechanism with geo-indistinguishability constraint. When ϵ_g is small, geo-indistinguishability overpowers the expected inference error in privacy protection, and therefore the limitation of geo-indistinguishability against inference attacks shows effects, as shown in Figure 16a; When ϵ_g increases, the constraint of geo-indistinguishability is relaxed, and the expected inference error takes control, helping to overcome weak protection on skewed locations against optimal inference attack (only 49 left being weak in Figure 16b, none in Figure 16c). By comparing Figure 16a(left) with 16c(left), PIVE shows the benefits of both privacy notions against optimal inference attack simultaneously: similar privacy protection as the joint mechanism with ϵ_g=0.8 except regions 48-50, and similar privacy protection as the joint mechanism with ϵ_g=2.2 for regions 48-50.

Even though with PIVE, all 50 regions are ensured to be above the user-defined inference error bound with geo-indistinguishability and good expected inference errors comparing to existing approaches, there are some location points in which PIVE offers slightly lower level of protection compared to the joint mechanism. We would like to make two remarks: (1) The level of privacy protection offered by both PIVE and joint optimization are exceeding the user-defined lower error

bound at those locations, thus are acceptable for users as good privacy protection, even though the privacy metric of PIVE is slightly lower. (2) For the weak locations, PIVE shows high resilience and adaptivity to the skewed distribution against inference attacks, compared to all three existing approaches (see Figure 15 and Figure 16).

VII. Conclusions

We have presented PIVE, a two-phase dynamic differential location privacy framework for providing stronger notion of location privacy in terms of background knowledge based inference attacks. This paper makes three novel contributions. First, we formally study the relationship between geo-indistinguishability and expected inference error, and demonstrate inherent problems of using geo-indistinguishability alone as the ultimate goal of location privacy protection through formal analysis and experimental illustration. Second, we propose a dynamic differential location privacy protection framework, where we first determine a set of protection locations by guaranteeing the expected inference error bound defined by a mobile user with respect to her service request by taking into account the adversary's prior distribution of the user's locations. Then, we generate the pseudo-locations in a differentially private way. Third, this two-phase framework constructs location obfuscation dynamically by capturing the relationship between two privacy notions based on adversary's current prior information and user-specific privacy requirements for different spatial-temporal contexts. Our experimental evaluation shows that the proposed PIVE approach effectively guarantees the two privacy notions simultaneously and outperforms the existing mechanisms that either offer geo-indistinguishability or quantify location privacy by expected inference errors in terms of adaptive privacy protection and computation efficiency.

Acknowledgment

The authors would like to thank Dr. Shokri, our shepherd, and all the reviewers for their helpful comments and suggestions, which help improving both the technical quality and the presentation of our paper.

References

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

[2] Location guard. https://addons.mozilla.org/en-US/firefox/addon/location-guard/.

[3] Uber. https://www.uber.com/.

[4] Yelp. http://www.yelp.com.

[5] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of ACM CCS*, pages 901–914, New York, NY, USA, 2013. ACM.

[6] N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Optimal geo-indistinguishable mechanisms for location privacy. In *Proceedings of ACM CCS*, pages 251–262, New York, USA, 2014. ACM.

[7] K. Chatzikokolakis, C. Palamidessi, and M. Stronati. Constructing elastic distinguishability metrics for location privacy. *PoPETs*, 2015(2):156–170, 2015.

[8] C. Dwork. Differential privacy. In *Automata, Languages and Programming*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12. Springer Berlin Heidelberg, 2006.

[9] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Conference on Theory of Cryptography*, TCC'06, pages 265–284, Berlin, Heidelberg, 2006. Springer-Verlag.

[10] K. Fawaz, H. Feng, and K. G. Shin. Anatomization and protection of mobile apps' location privacy threats. In *Proceedings of the 24th USENIX Conference on Security Symposium*, SEC'15, pages 753–768, Berkeley, CA, USA, 2015. USENIX Association.

[11] K. Fawaz and K. G. Shin. Location privacy protection for smartphone users. In *Proceedings of ACM CCS*, pages 239–250, New York, NY, USA, 2014. ACM.

[12] 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, June 2005.

[13] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of MobiSys*, pages 31–42, New York, USA, 2003. ACM.

[14] J. I. Hong and J. A. Landay. An architecture for privacy-sensitive ubiquitous computing. In *Proceedings of MobiSys*, pages 177–189, New York, NY, USA, 2004. ACM.

[15] A. Khoshgozaran and C. Shahabi. Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy. In *Proceedings of the 10th International Conference on Advances in Spatial and Temporal Databases*, SSTD'07, pages 239–257, Berlin, Heidelberg, 2007. Springer-Verlag.

[16] J. K. Lawder and P. J. H. King. Using space-filling curves for multi-dimensional indexing. In *Proceedings of the 17th British National Conferenc on Databases: Advances in Databases*, BNCOD 17, pages 20–35, London, UK, UK, 2000. Springer-Verlag.

[17] F. McSherry and K. Talwar. Mechanism design via differential privacy. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '07, pages 94–103, 2007.

[18] B. Moon, H. v. Jagadish, C. Faloutsos, and J. H. Saltz. Analysis of the clustering properties of the hilbert space-filling curve. *IEEE Trans. on Knowl. and Data Eng.*, 13(1):124–141, Jan. 2001.

[19] H. Ngo and J. Kim. Location privacy via differential private perturbation of cloaking area. In *2015 IEEE 28th Computer Security Foundations Symposium*, pages 63–74, July 2015.

[20] R. Shokri. Privacy games: Optimal user-centric data obfuscation. *PoPETs*, 2015(2):299–315, 2015.

[21] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux. Quantifying location privacy. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pages 247–262, May 2011.

[22] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec. Protecting location privacy: Optimal strategy against localization attacks. In *Proceedings of ACM CCS*, pages 617–627, New York, USA, 2012. ACM.

[23] G. Theodorakopoulos, R. Shokri, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec. Prolonging the hide-and-seek game: Optimal trajectory privacy for location-based services. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, WPES '14, pages 73–82, New York, NY, USA, 2014. ACM.

[24] Y. Xiao and L. Xiong. Protecting locations with differential privacy under temporal correlations. In *Proceedings of ACM CCS*, pages 1298–1309, New York, NY, USA, 2015. ACM.

[25] T. Xu and Y. Cai. Feeling-based location privacy protection for location-based services. In *Proceedings of ACM CCS*, pages 348–357, New York, NY, USA, 2009. ACM.

[26] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on gps data. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, UbiComp '08, pages 312–321, New York, NY, USA, 2008. ACM.

[27] Y. Zheng, X. Xie, and W.-Y. Ma. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data(base) Engineering Bulletin*, June 2010.

[28] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 791–800, New York, NY, USA, 2009. ACM.