Releasing Correlated Trajectories: Towards High Utility and Optimal Differential Privacy

Lu Ou *Student Member, IEEE*, Zheng Qin *Member, IEEE*, Shaolin Liao *Senior Member, IEEE*, Yuan Hong *Senior Member, IEEE*, Xiaohua Jia *Fellow, IEEE*

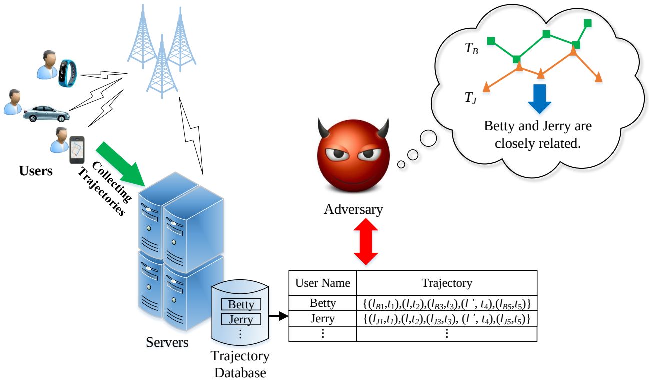
Abstract—Mutual correlation between trajectories of two users is very helpful to real-life applications such as product recommendation and social media. While providing tremendous benefits, the releasing of correlated trajectories may leak sensitive social relations, due to potential links between mutual correlations and social relations. To the best of our knowledge, we take the first step to propose a mathematically rigorous n-body Laplace framework, satisfying ε-differential privacy, which efficiently prevents social relations inference through the mutual correlation between n-node trajectories of two users. The problem is mathematically formulated by defining a trajectory correlation score to measure the social relation between two users. Then, under the n-body Laplace framework, we propose two Lagrange Multiplier-based Differentially Private (LMDP) approaches to optimize the privacy budgets, for the data utility measured by location distances and the data utility measured by location correlations, *i.e.*, UD-LMDP and UC-LMDP. Also, we present detailed analyses of privacy, data utility, adversary knowledge and the constrained optimizations. Finally, we perform experimental studies with real-life data. Our experimental results show that our proposed approaches achieve better privacy and data utility than the existing approaches.

Index Terms—Trajectory Correlation, Differential Privacy, Lagrange Multiplier Method, Constrained Optimization.

1 Introduction

With the growing ubiquity of smart phones, wearable devices and GPS enabled devices, movement trajectories can be collected by servers, such as location-based service (LBS) servers, anytime and anywhere. Two correlated trajectories tend to indicate some social relations between two users. So the correlation of users' trajectories is very helpful for link prediction, rating prediction, product recommendation, and community discovery.

While the correlation between trajectories provides great benefits, it also results in serious privacy concerns. Specifically, the correlation is a strong indication of a social relation [1]. Sensitive social relations may be inferred from the trajectory correlation. A typical example is shown in Fig. 1: assuming that adversaries have the prior background knowledge of Betty and Jerry, when adversaries know that they often go to the restaurants and movie theaters together based on the released trajectories, they can infer that Betty and Jerry are closely related. Adversaries can further confirm their social relation based on their prior knowledge of Betty and Jerry. So, such mutual correlation of trajectories should be hidden to protect social relations among users and thus their privacy.



User Name	Trajectory
Betty	$\{(l_{B1},t_1),(l,t_2),(l_{B3},t_3),(l\,', t_4),(l_{B5},t_5)\}$
Jerry	$\{(l_{J1},t_1),(l,t_2),(l_{J3},t_3), (l\,', t_4),(l_{J5},t_5)\}$
⋮	⋮

Fig. 1: A social relation inference example via a mutual correlation between trajectories

To the best of our knowledge, there is little research on differential privacy of the mutual correlation of a trajectory pair. Most research works deal with the privacy issues of a single trajectory, such as locations disclosure of a single trajectory and a mobility pattern divulgation of one user. According to the privacy problems whether caused by temporal-spatial correlations within a single trajectory or not, the existing privacy protection methods about trajectories can be summarized into two categories: 1) single-trajectory privacy without correlation; and 2) single-trajectory privacy with correlation within a single trajectory.

- *Lu Ou, Zheng Qin are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha, Hunan, China 410082. Zheng Qin is the Corresponding Author.*
E-mail: zqin@hnu.edu.cn, oulu9676@gmail.com
- *Shaolin Liao is with Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, IL, USA 60439; He is also affiliated with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, 10 W 31st Street, Chicago, IL 60616.*
E-mail: sliao@anl.gov and sliao5@iit.edu
- *Yuan Hong is with the Department of Computer Science, Illinois Institute of Technology, 10 W 31st Street, Chicago, IL 60616.*
E-mail: yuan.hong@iit.edu
- *Xiaohua Jia is with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong.*
E-mail: csjia@cityu.edu.hk

On one hand, some trajectory privacy protection methods based on k-anonymity [2] and suppression [3], have been proposed to protect individual privacy not caused by a temporal-spatial correlation within a single trajectory. Unfortunately, if protected locations in a sensitive area, after trajectory protection by using these solutions, adversaries still can obtain individual privacy via many types of privacy attacks, such as composition attack, deFinetti attack and background knowledge attack [4]. To overcome such drawback, differential privacy [5], which makes adversaries cannot distinguish whether an individual's record is included in the database or not, is adapted to protect trajectories' privacy which is not caused by trajectory correlations [4], [6]. However, although these solutions can protect trajectory privacy efficiently, they do not consider to protect two-user trajectories' correlation.

On the other hand, several works consider a correlation between locations within a single trajectory. In the earlier works [7], [8], [9], researchers utilized a prefix tree or the n-gram model to construct two-location correlation within a single trajectory, protecting individual privacy. Then Markov model [10] is adopted to achieve differentially private correlation protection within a single trajectory. Besides, the geographical and semantic correlations [11] are used to synthesize trajectories in order to protect individual privacy. However, none of them deals with the mutual-trajectory correlation between users and their social relation linked to them.

It is a challenging issue to deal with the privacy problem of the social relation inference through the mutual-trajectory correlation. First, it is difficult to quantify the social relation linked to the two-user trajectories in rigorous mathematical expressions. Second, it is hard to preserve data utility for a garuanteed privacy or vice versa. To fill in such gaps, in this paper, we propose an n-body privacy framework for protecting the social relation between any two users against adversarial inferences from their mutual-trajectory correlation.

Therefore, the main contributions of this paper are summarized as below:

- We propose a rigorous mathematical model to formulate a score measurement based on the trajectory distance to quantify the social relation between two users.

- We propose an n-body Laplace framework, for the score measurement of the mutual-trajectory correlation, satisfying ε-differential privacy. Specifically, we first define two important utilities: one is the location utility, *i.e.*, the data utility measured by location distances, and the other is the location correlation utility, *i.e.*, the data utility measured by location correlations; then we propose two Lagrange Multiplier-based Differentially Private (LMDP) approaches, called "UD-LMDP" for the given location utility and "UC-LMDP" for the given location correlation utility, respectively, by optimizing an n-element Laplace scale parameter set through the Lagrangian constrained optimization; and finally, we obtain the mathematical formulas of the optimal Laplace noise scale parameter sets for these two given utilities.

- We present detailed theoretical analyses of privacy, data utility, adversary knowledge and the constrained optimizations.

- We experimentally evaluate our approaches, *i.e.*, UD-LMDP and UC-LMDP, with real-life data. The experimental results demonstrate that our approaches achieve better privacy and data utility than the existing approaches.

The remainder of this paper is organized as follows. In Section 2, we review the related work in the literature. In Section 3, we formally define our problem, and then introduce some background knowledge, including differential privacy and the constrained optimization via the Lagrange Multiplier (LM) method. Then, in Section 4, we mathematically formulate a trajectory distance between users and obtain the Probability Density Function (PDF) of the trajectory distance; also, we define a correlation score to quantify the social relation between users. Later in Section 5, we introduce the n-body Laplace framework and the details of two privacy protection approaches. Furthermore, the privacy analysis, the data utility analysis, the analysis on the adversary knowledge and the constrained optimizations are depicted in Section 6. What's more, we present two privacy-preserving trajectory releasing algorithms in Section 7. The experimental performance evaluations are shown in Section 8. Finally, Section 9 concludes this paper.

2 Related Work

In this section, we review the literature on single-trajectory privacy, with or without consideration of correlations.

2.1 Single-trajectory Privacy without Correlations

Privacy protection is required for raw trajectories releasing. This is because, due to potential sensitive location information in trajectories, adversaries can utilize their background knowledge to infer users privacy. Literature reviews of different methods are available in [12], [13]. The most immediate and common solutions are dummy trajectories [14], suppression technique [15], techniques based on k-anonymity [2] and their variants. Terrovitis *et al.* proposed a suppression technique to publish trajectories. Huo *et al.* [16] proposed You Can Walk Alone (YCWA). This method anonymizes the stay points which are extracted efficiently on users' trajectories. Chen *et al.* [3] proposed a trajectory anonymity model based on local suppression. To deal with the trajectory similarity and direction, Gao *et al.* [17] proposed an anonymity region on the basis of the distance, considering the case that not all locations on the trajectory are sensitive. Liu *et al.* [18] proposed an anonymous privacy protection by introducing a game theorem. Cicek *et al.* [19] proposed p-confidentiality that can ensure the location diversity in trajectories. Unfortunately, social relations can be predicted by cloaked human trajectories [20].

What's more, Li *et al.* [21] proposed a differentially private trajectory publication without consideration of correlations of trajectories. This approach consists of a differentially private location generation algorithm and a bounded Laplace mechanism. Wang *et al.* [22] proposed a differentially private event histogram for trajectories releasing.

Riboni *et al.* [23] proposed trajectory privacy protection in a context-aware recommender system through the (L, j)-density mechanism to achieve ε-differential privacy. Quan *et al.* [24] proposed a trajectory obfuscation mechanism based on the Laplace mechanism. In this mechanism, a polar Laplacian is used to add noise on trajectories. Cao *et al.* [25] proposed an ℓ-trajectory privacy model to protect a trajectory whose length is ℓ. This model satisfies ε-differential privacy for releasing privacy-preserving real-time statistics over trajectory stream. However, these approaches also do not consider mutual-trajectory correlation that may lead to the social relation leakage.

2.2 Single-trajectory Privacy with Correlations

Generally, there exists spatio-temporal correlations in trajectories that could be harmful for users' privacy. For example, for a single user, adversaries may utilize Markov model to build the user's mobility pattern and thus infer the single-user trajectory [26], [27]. There are several works that consider such spatio-temporal correlation within a single trajectory [7], [8], [11], [13], [28]. To release differentially private trajectories with spatio-temporal correlations [29], Chen *et al.* [7], [30] proposed a data-dependent solution by recursively constructing a noisy prefix tree based on the existed trajectory data. However, with the growth of the prefix tree, this solution has a poor data utility. To solve this problem, Chen *et al.* [8] proposed a differentially private trajectory publishing method by using the variable n-gram model as well as a set of novel techniques under the Markov assumption. He *et al.* [31] proposed a trajectory synthesis, called "DPT", according to correlations between locations within a single trajectory. Their proposed a system considers individual movements' speeds and constructs prefix tree counts that satisfies ε-differential privacy. Then, they presented a tool, called "VisDPT" [32], helping data curators understand privacy problems for data releasing and their proposed mechanism. Wang *et al.* [9] proposed a private reference system by cluster-based anchor points under X-order Markov assumption for raw discrete trajectories. The noisy calibrated trajectories are released by using differentially private prefix trees. For the privacy-preserving trajectory releasing, Song *et al.* [10] proposed Markov Quilt Mechanism (MQM) that satisfies Pufferfish privacy framework to protect the spatio-temporal correlations within a single trajectory. At last, Wang *et al.* [33], [34] proposed "RescueDP" to release crowd-sourced data with spatio-temporal correlations by using differential privacy. However, these works only consider the spatio-temporal correlations within a single trajectory, but not the correlation of the mutual trajectories.

Above all, according to our knowledge, there is little literature dealing with the mutual-trajectory correlation. However, it is an important privacy problem because social relations can be inferred through the mutual-trajectory correlation [20]. Thus in this paper, we propose an n-body differential privacy framework for the mutual-trajectory correlation to protect the social relation against the inference attack.

3 Preliminaries

In this section, in order to explain our problem, we illustrate the statement of our problem and the basic idea of our approaches. Then, we introduce some basic concepts, including differential privacy and the constrained optimization via the LM method.

3.1 Problem Statement and Basic Idea

To represent the correlation of mutual trajectories, we first define a user's daily trajectory.

Definition 3.1 (Daily Trajectory)**.** A trajectory of a user u, denoted by $\mathbb{T}_u$, is a sequence of locations that a user u moves on a single day in which there are n sampled time slots. It is defined as

$$\mathbb{T}_u = \{(x_u^i, y_u^i) | i = 1, 2, \cdots, n\},$$

where x_u^i and y_u^i are the u's longitude and latitude at the i-th time slot, respectively. In this paper, we assume that the coordinates x_u^i and y_u^i are independent from each other and thus can be treated individually.

A trajectory $\mathbb{T}_u$, represents a user u's trajectory on a single day. A user's trajectory data set usually contains a user's trajectories of many days. When we study the mutual-trajectory correlation of two users, we always compare these trajectories on the same day. We assume all users' trajectories are aligned in the same time slots sequence.

When adversaries obtain users' trajectory data sets, they may compute the distance of each two-user trajectories of the same days to form a distance database DB. According to the distance, they could obtain a mutual-trajectory correlation between two users. Then they could relate this correlation information to their background knowledge to see whether there is a social relation between them. In such case, the social relation linked to the mutual-trajectory correlation should be protected in order to avoid such inference attack.

In this paper, we first quantify the mutual-trajectory correlation to measure the social relation between two users. We then propose an n-body Laplace framework to protect the social relation from mutual-trajectory correlation inference attack. Finally, based on this framework, for two given utilities, we propose two Lagrange multiplier-based differetially private approaches by optimizing Laplace scale parameters at all n locations of a daily trajectory through the constrained optimization via the LM method, *i.e.*, UD-LMDP and UC-LMDP.

3.2 Basic Concept

Before we go into the details of our proposed approaches, we introduce some mathematical background here. To start, **Table 1** defines the variables we use across this paper.

3.2.1 Differential Privacy

Differential privacy [5] is a powerful tool to provide privacy-preserving query answers over databases. It guarantees that the PDF of noisy query answers changes very little with the addition, deletion or modification of any tuple.

Definition 3.2 (Differential Privacy)**.** Let $\mathcal{A}$ be a randomized algorithm, $\mathbb{A}$ be any output of $\mathcal{A}$, $\mathbb{D}_1$ and $\mathbb{D}_2$ be two

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2018.2853105, IEEE Transactions on Dependable and Secure Computing

TABLE 1: Notation Definition

Symbol	Description
$\mathbb{T}_u$	A daily trajectory of a user u.
$\mathbb{D}(a,b)$	A daily trajectory distance between the user a and b.
(d_x^i, d_y^i)	The element of $\mathbb{D}(a,b)$ at the i-th time slot.
$\mathbf{D}^i$	The location distance data subset at the i-th time slot.
$\mathbf{D}$	A dataset of trajectory distances over all days: $\{\mathbf{D}^i \mid i=1,2,\cdots,n\}$.
$\boldsymbol{\delta}$	The noise set of the trajectory distance.
R	The range of the correlation.
$(\widetilde{d_x^i}, \widetilde{d_y^i})$	The normalized location distance at the i-th time slot: $(d_x^i/R, d_y^i/R)$.
$\boldsymbol{\rho}$	A location correlation set: $\{(\rho_x^i, \rho_y^i) \mid i=1,2,\cdots,n\}$.
$\boldsymbol{\sigma}_{\mathbb{D}}$	A standard deviation set of $\mathbb{D}(a,b)$: $\{\sigma_{d^i} \mid i=1,2,\cdots,n\}$.
S	The score between trajectories.
$\mathbf{b}$	An n-element Laplace scale parameter set: $\{b_i \mid i=1,2,\cdots,n\}$.
$\tilde{\boldsymbol{b}}$	A normalized Laplace scale parameter set: $\mathbf{b}/R$.
$\mathcal{U}_d$	A location utility.
$\mathcal{U}_c$	A location correlation utility.
$\mathcal{L}$	A Lagrange operator.

neighboring datasets that differ in only one tuple. $\mathcal{A}$ will be ε-differential private, if

$$\frac{Pr(\mathcal{A}(\mathbb{D}_1) \in \mathbb{A})}{Pr(\mathcal{A}(\mathbb{D}_2) \in \mathbb{A})} \leq exp(\varepsilon).$$

Differential privacy is to protect data privacy by adding noise for a query function. And a sensitivity is a key parameter for noise addition. It means that the maximum difference of the query response for two database which differ only one tuple. Its definition is given below.

Definition 3.3 (Sensitivity)**.** If there is a function $F : \mathbb{D} \to R^n$, its input is a dataset $\mathbb{D}_1$ and its output is an n-dimensional vector, then for the given dataset $\mathbb{D}_1$ and its neighboring dataset $\mathbb{D}_2$, a sensitivity ΔF of the function F is defined as follow:

$$\Delta F = \max_{\mathbb{D}_1, \mathbb{D}_2} ||F(\mathbb{D}_1) - F(\mathbb{D}_2)||_1.$$

Differential privacy can be generalized to Group Differential Privacy [10], denoted as "Group-DP". And its definition is as follow.

Definition 3.4 (Group Differential Privacy)**.** Let $\mathbb{D}$ be a database with n tuples, and let $\mathbb{G} = \{G_1, G_2, \cdots, G_n\}$, G_i be a subset of $\mathbb{G}$. A privacy mechanism S is said to be ε-group differentially private with respect to $\mathbb{G}$ if for all G_i in $\mathbb{G}$, for all pairs of databases $\mathbb{D}_1$ and $\mathbb{D}_2$ which differ in the private values of the individuals in G_i, and for all $\mathbb{S} \subseteq Range(S)$, we have:

$$\frac{Pr(S(\mathbb{D}_1) \subseteq \mathbb{S})}{Pr(S(\mathbb{D}_2) \subseteq \mathbb{S})} \leq exp(\varepsilon).$$

3.2.2 The Constrained Optimization Problem

The constrained optimization problem is a strategy for finding the local maxima and minima of a function $f(x,y)$ subject to equality constraint $g(x,y) = 0$. The Lagrange Multiplier method can achieve this and it is defined as follow,

$$\begin{cases} \mathcal{L}(x,y|\lambda) = f(x,y) - \lambda g(x,y), \\ g(x,y) = 0. \end{cases}$$

4 Quantization

Before we present our approaches, we assume that the PDF of trajectories is Gaussian and give some mathematical quantizations. In general, the smaller the distance between two users, the stronger the relation between them. We first give a measurement of a trajectory distance which consists of multiple location distances. Then we quantify a social relation between two users based on the distance.

4.1 The Trajectory Distance

For each two daily trajectories moved by two users, we will compute a trajectory distance between them. And the definition of the trajectory distance is as follow.

Definition 4.1 (Trajectory Distance)**.** A trajectory distance between the user a and b, denoted by $\mathbb{D}(a,b)$, is a sequence of distances of location coordinates between a and b, and its definition is as below:

$$\mathbb{D}(a,b) = \{(d_x^i, d_y^i) | i = 1, 2, \cdots, n\},$$

where $d_x^i = x_a^i - x_b^i$ and $d_y^i = y_a^i - y_b^i$ are the location distances of the longitudes and the latitudes at the same i-th time slot, respectively.

4.2 The PDF of the Trajectory Distance

We assume that two users' social relation highly depends on their trajectory distance $\mathbb{D}$: the closer the users, *i.e.*, the smaller (d_x^i, d_y^i) in $\mathbb{D}$, the stronger the correlation and vice versa. If the two users' trajectories are not at the same time slots, we can perform interpolation to obtain their approximate locations on the same time slots. Now let's look at the PDF of the trajectory distance $\mathbb{D}$ first. Assuming that the trajectory given in **Definition 3.1** follows the joint correlated Gaussian distribution (see **Appendix A**), the probability of the trajectory distance $\mathbb{D}$ between two users is then given by

$$Pr(\mathbb{D}(a,b)) = \prod_{i=1}^{n} \frac{\exp\left\{ -\left(\frac{(d_x^i)^2}{2(\sigma_{d_x^i})^2} + \frac{(d_y^i)^2}{2(\sigma_{d_y^i})^2} \right) \right\}}{2\pi \sigma_{d_x^i} \sigma_{d_y^i}}, \quad (1)$$

where $\sigma_{d_x^i}$ and $\sigma_{d_y^i}$ are standard deviations of d_x^i and d_y^i respectively, and $\sigma_{d_x^i} = \sqrt{(\sigma_{x_a}^i)^2 + (\sigma_{x_b}^i)^2 - \rho_x^i \sigma_{x_a}^i \sigma_{x_b}^i}$ and $\sigma_{d_y^i} = \sqrt{(\sigma_{y_a}^i)^2 + (\sigma_{y_b}^i)^2 - \rho_y^i \sigma_{y_a}^i \sigma_{y_b}^i}$, where $\sigma_{x_a}^i(\sigma_{x_b}^i)$ is the standard deviation of the longitudes x about the user $a(b)$ at the i-th time slot, $\sigma_{y_a}^i(\sigma_{y_b}^i)$ is the standard deviation of the latitudes y about the user $a(b)$ at the i-th time slot, and $\rho_x^i(\rho_y^i)$ is the location correlation coefficient of longitudes (latitudes) at the i-th time slot.

4.3 Quantization of the Social Relation

The smaller distance between two users means the stronger correlation between them and vice versa. Therefore, we measure the mutual-trajectory correlation based on their trajectory distance as defined in **Definition 4.1**. Before we quantify a social relation between two users, we define a trajectory correlation score according to the trajectory distance, as shown in **Definition 4.2**.

Definition 4.2 (Trajectory Correlation Score)**.** The trajectory correlation score S is a measurement of the correlation between a pair of daily trajectories, defined as

$$S = S(x)S(y),$$

where $S(x) = 1 - \prod_{i=1}^{n} \left\{1 - \exp\left(-\frac{|d_x^i|}{R}\right)\right\}$ and $S(y) = 1 - \prod_{i=1}^{n} \left\{1 - \exp\left(-\frac{|d_y^i|}{R}\right)\right\}$, where R is a parameter that defines the range of the correlation.

Then, we quantify the two-user social relation through S. From **Definition 4.2**, there is a parameter R that defines the correlation range of the location distance to indicate the social relation. As shown in **Table 2**, there are score values for some typical location distance values of a 1-node trajectory pair: when the location distances d_x^i and d_y^i are close to the range R, the score $S \approx 0$ is small, indicating a weak social relation; and when the location distances d_x^i and d_y^i are close to zero, the score $S \approx 1$ is large, indicating a strong social relation.

TABLE 2: Typical Correlation Score Examples

Location Distance d_x^1	**Location Distance** d_y^1	**Score** S
0	0	1
0	$R/2$	0.607
0	$R/\sqrt{2}$	0.493
0	R	0.368
$R/2$	$R/2$	0.368
$R/\sqrt{2}$	$R/\sqrt{2}$	0.243
R	R	0.135

5 Our Approaches

We propose an n-body Laplace framework to deal with the mutual-trajectory correlation by optimizing the n-element Laplace scale parameter set $\mathbf{b}$. Intuitively, a high data utility means that raw trajectories cannot be distorted too much. Specifically, for relation-based services, such as product recommendations, the raw location correlations cannot be changed too much in the released trajectory data. Then, while ensuring a high level output data utility, we propose two approaches to optimize the privacy budgets through the constrained optimization via the LM method, for two types of utilities: the location utility and the location correlation utility, respectively.

5.1 n-Body Laplace Framework

Due to n locations in a daily trajectory, we should deal with the n location distances at n time slots for trajectory correlation privacy. We propose an n-body Laplace framework to solve the mutual correlation between trajectories which contain n locations on each day. Before we present the n-body Laplace noise mechanism, let's look at the maximally different neighboring trajectory distances and the sensitivity of the problem.

Before we present the maximally different neighboring trajectory distances, let's first define the maximum one-time-slot location distance and the maximum two-user location distance as shown in **Definition 5.1** and **5.2**.

Definition 5.1 (Maximum One-time-slot Location Distance)**.** Denote a data subset of location distances of longitudes or latitudes at the i-th time slot for all days as $\mathbf{D}^i \in \left\{\mathbf{D}_x^i, \mathbf{D}_y^i\right\}$, and the maximum location distances at different time slots for a large longitude or latitude data subset are obtained from Eq. (1),

$$Pr\left(\max\left(\mathbf{D}^i\right) = 3\sigma_{d^i}\right) \geq 99.7\%,$$

which includes 99.7% trajectories for the Gaussian distribution, where $\sigma_{d^i} \in \left\{\sigma_{d_x^i}, \sigma_{d_y^i}\right\}$ is the standard deviation of location distances of longitudes or latitudes at the i-th time slot for all days.

In this paper, we assume that we have the data subsets are large so that σ_{d^i} can be calculated accurately.

Then, according to **Definition5.1**, we will define the maximum two-user location distance as shown in **Definition5.2**. Thus, we can achieve the neighboring trajectory distances as shown in **Definition5.3**.

Definition 5.2 (Maximum Two-user Location Distance)**.** For all days, the maximum location distance between two users, denoted as $\max(\mathbf{D})$, is defined as below,

$$\max(\mathbf{D}) = \max\left(\{3\sigma_{d^i} | i = 1, 2, \cdots, n\}\right),$$

where $\mathbf{D} = \{\mathbf{D}^1 \cup \mathbf{D}^2 \cup \cdots \cup \mathbf{D}^n\}$.

Definition 5.3 (Neighboring Trajectory Distances)**.** There are two trajectory distances between users a and b, denoted by $\mathbb{D}_1(a, b)$ and $\mathbb{D}_2(a, b)$, if they differ only at the i-th time slot, they are neighboring trajectory distances.

Lemma 5.1. When $\mathbb{D}_{max} = \{0, \cdots, 0, \max(\mathbf{D}), 0, \cdots, 0\}$ and $\mathbb{D}_0 = \{0, \cdots, 0\}$, they are the maximally different neighboring trajectory distances.

Proof. For two neighboring trajectory distances given in **Definition 5.3**, to obtain the maximum privacy given in **Definition 3.2**, the ratio of score probabilities $Pr(S)$ has to be maximized. This occurs when $\mathbb{D}_{max} = \{0, \cdots, 0, \max(\mathbf{D}), 0, \cdots, 0\}$ and $\mathbb{D}_0 = \{0, \cdots, 0\}$, according to the score given in **Definition 4.2**.

With the maximally different neighboring trajectory distances, let's define the trajectory correlation sensitivity.

Definition 5.4 (Trajectory Correlation Sensitivity)**.** Trajectory correlation sensitivity is the maximum difference of the trajectory correlation score between neighboring trajectory distances, its definition is as below:

$$\Delta S = \max_{\mathbb{D}_1(a,b), \mathbb{D}_2(a,b)} ||S(\mathbb{D}_1(a, b)) - S(\mathbb{D}_2(a, b))||_1 = 1. \quad (2)$$

In achieving Eq. (2), we have used **Lemma 5.1** and the score in **Definition 4.2** and the following,

$$S(0) = 1, \;\; S(\max(\mathbf{D})) = 1 - \left(1 - \exp\left\{-\frac{\max(\mathbf{D})}{R}\right\}\right) = 0.$$

Now let's introduce our proposed n-body Laplace framework as shown in **Definition 5.5**.

Definition 5.5 (n-body Laplace Framework)**.** The sanitized trajectory distance $\mathbb{D}'(a, b)$ after a noise set $\boldsymbol{\delta}$ added is given by

$$\mathbb{D}'(a, b) = \mathbb{D}(a, b) + \boldsymbol{\delta},$$

where $\boldsymbol{\delta}$ follows the joint Laplace distribution that is typical for differential privacy,

$$Pr\left(\boldsymbol{\delta}\right) = \prod_{i=1}^{n} \frac{\exp\left\{\frac{-|\delta_x^i|}{\left(3\sigma_{d_x^i}\right)b_x^i}\right\} \exp\left\{\frac{-|\delta_y^i|}{\left(3\sigma_{d_y^i}\right)b_y^i}\right\}}{4b_x^i b_y^i \left(3\sigma_{d_x^i}\right)\left(3\sigma_{d_y^i}\right)}, \quad (3)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2018.2853105, IEEE Transactions on Dependable and Secure Computing

where $\boldsymbol{\delta} = \left\{(\delta_x^i, \delta_y^i) \middle| i = 1, 2, \cdots, n\right\}$ is a noise set for longitudes and latitudes at each time slot; $\mathbb{D}(a,b)$ is the raw trajectory distance between two users a and b; $\mathbf{b} = \{(b_x^i, b_y^i)|i = 1, 2, \cdots, n\}$ is the Laplace scale parameter set, where b_x^i and b_y^i are the Laplace scale parameters of the longitude and latitude at the i-th time slot respectively, and they are determined by the sensitivity and the privacy budgets ε_x^i and ε_y^i for the longitude and latitude at the i-th time slot, *i.e.*, $b_x^i = \frac{1}{\varepsilon_x^i}$, $b_y^i = \frac{1}{\varepsilon_y^i}$.

Our goal is to obtain the optimal Laplace noise scale parameter set $\mathbf{b}$ for the maximum privacy (*i.e.*, the minimum privacy budget ε) given in **Definition 3.2**, which can be shown to be equivalent to obtain the optimal Laplace noise scale parameter set $\mathbf{b}$ for the maximum data utility $\mathcal{U}$, as shown in **Lemma 5.2** below.

Lemma 5.2. The maximization of the data utility for a given privacy budget is equivalent to the minimization of the privacy budget for a given data utility, *i.e.*, $\max\{\mathcal{U}|\varepsilon\} \Leftrightarrow \min\{\varepsilon|\mathcal{U}\}$.

Proof. Let's look at the two optimization cases: 1) on one hand, the maximization of the data utility for a given privacy budget requires that $\partial\mathcal{U}/\partial\mathbf{b} = 0$; and 2) on the other hand, the minimization of the privacy budget for a given data utility requires that $\partial\varepsilon/\partial\mathbf{b} = 0$. However, it is well known that the privacy budget is a monotonically decreasing function of the data utility, *i.e.*, $\partial\varepsilon/\partial\mathcal{U} < 0$. So we have, $\partial\varepsilon/\partial\mathbf{b} = \partial\varepsilon/\partial\mathcal{U} \times \partial\mathcal{U}/\partial\mathbf{b} = 0$, which reduces to $\partial\mathcal{U}/\partial\mathbf{b} = 0$. This is exactly the maximization of the data utility for a given privacy budget shown in case 1). However, due to $\partial\varepsilon/\partial\mathcal{U} < 0$, we are now dealing with the minimization instead of the maximization and **Lemma 5.2** is proved.

5.2 UD-LMDP: LMDP for the Data Utility Measured by Location Distances

For services based on accurate locations, the location distortion cannot be too large after injecting the noise during trajectories sanitization. In other words, as the noise induced the distortion increases, the data utility becomes lower. Since such distortion can be measured by a location distance deviation between locations before and after the noise is added, we define a location utility measure based on such location distance deviation as below.

Definition 5.6 (Location Utility)**.** The location utility is a measure of the data utility from the aspect of the absolute distance deviation and given as

$$\mathcal{U}_d \equiv \frac{1}{n}\sum_{i=1}^{n}\left[\mathrm{E}\left(\left|\delta_x^i\right|\right) + \mathrm{E}\left(\left|\delta_y^i\right|\right)\right],$$

where δ_x^i and δ_y^i are the additive noise of the longitude and the latitude at the i-th time slot, respectively.

Then we propose a mutual-trajectory correlation protection under the n-body Laplace framework according to **Definition 5.5** for a given location utility $\mathcal{U}_d$ given in **Definition 5.6**, *i.e.*, UD-LMDP. According to **Lemma 5.2**, we can obtain the optimal Laplace noise scale parameter set $\mathbf{b}_{u_d}$ to achieve the maximum privacy (*i.e.*, the minimum privacy budget ε) for the given utility u_d. This is a standard constrained optimization problem and we can realize it via the LM method, *i.e.*, $\min\{\varepsilon|\mathcal{U}_d = u_d\}$.

5.3 UC-LMDP: LMDP for the Data Utility Measured by Location Correlations

For services based on two-user location correlations, we do not want to distort the location correlation too much after injecting the noise during trajectories sanitization. Thus, the data utility is also related to the location correlation. Before the definition of the location correlation utility, we should define the location correlation coefficients as below.

Definition 5.7 (Location Correlation Coefficients)**.** There are two users a and b, and their location correlation coefficients of the longitude and the latitude are given below:

$$\rho_x^i \equiv \frac{Cov\{x_a^i, x_b^i\}}{\sqrt{Var\{x_a^i\}Var\{x_b^i\}}}, \rho_y^i \equiv \frac{Cov\{y_a^i, y_b^i\}}{\sqrt{Var\{y_a^i\}Var\{y_b^i\}}}, \quad (4)$$

where $Cov\{x_a^i, x_b^i\}$ and $Cov\{y_a^i, y_b^i\}$ are the covariances of the two users' longitudes and latitudes at the i-th time slot for all days, respectively, and $Var\{x_a^i\}$, $Var\{x_b^i\}$, $Var\{y_a^i\}$, $Var\{y_b^i\}$ are their variances.

According to the location correlation coefficients, we define the location correlation utility.

Definition 5.8 (Location Correlation Utility)**.** The location correlation utility is a measure of the data utility from the aspect of the location correlation, and it is given by

$$\mathcal{U}_c \equiv \frac{1}{n}\sum_{i=1}^{n}\left\{\left|\rho_{x0}^i - \rho_x^i\right| + \left|\rho_{y0}^i - \rho_y^i\right|\right\},$$

where ρ_{x0}^i (ρ_{y0}^i) and ρ_x^i (ρ_y^i) are location correlation coefficients of the longitude (latitude) at the i-th time slot before and after the noise is added, respectively.

Similar to UD-LMDP, we propose the other mutual-trajectory correlation protection under the n-body Laplace framework according to **Definition 5.5** for a given location correlation utility given in **Definition 5.8**, *i.e.*, UC-LMDP. Also, we can obtain the optimal Laplace noise scale parameter set $\mathbf{b}_{u_c}$ to achieve the maximum privacy (*i.e.*, the minimum privacy budget ε) for the given location correlation utility u_c, *i.e.*, $\min\{\varepsilon|\mathcal{U}_c = u_c\}$.

Note that UC-LMDP does not affect the PDF of locations of a user's trajectory, as shown in **Theorem 5.1**.

Theorem 5.1. The distortion of the location correlation set $\boldsymbol{\rho}$ does not affect the probability of each individual user.

Proof. Let's look at the marginal probability of each user's location at the i-th time slot, which can be obtained by integrating out the other users'. For example, the probability of the user a's location at the i-th time slot, is given by

$$\begin{aligned} Pr\left((x_a^i, y_a^i)\,|\boldsymbol{\rho}\right) &= \iint_{-\infty}^{+\infty} Pr\left((x_a^i, y_a^i, x_b^i, y_b^i)\,|\boldsymbol{\rho}\right) dx_b^i dy_b^i \\ &= \frac{1}{2\pi\sigma_{x_a}^i\sigma_{y_a}^i}\exp\left\{-\frac{1}{2}\left[\frac{(x_a^i)^2}{(\sigma_{x_a}^i)^2} + \frac{(y_a^i)^2}{(\sigma_{y_a}^i)^2}\right]\right\}, \end{aligned}$$

from which we can see that the change of the location correlation set $\boldsymbol{\rho}$ between two users does not affect the PDF of locations of a single user.

6 Analysis

In this section, we give the details of the privacy analysis, the data utility analysis, the adversary knowledge and the constrained optimizations.

6.1 Privacy Analysis

According to **Definition 3.2**, to prove that the n-body Laplace framework satisfies differential privacy, we need to find the maximum ratio of the trajectory correlation score probabilities for any neighboring trajectory distances.

Lemma 6.1. The maximum ratio of score probabilities occurs at the score $S\left(|(d^i)'|\approx\infty\right)\approx 0$ for the maximally different neighboring trajectory distances $\mathbb{D}_{max}$ and $\mathbb{D}_0$, where $(d^i)'$ is the noisy location distance for the both longitude and latitude at the i-th time slot.

Proof. According to **Definition 4.2**, the score probability decreases monotonically with $(d^i)'$. For the maximally different neighboring trajectory distances $\mathbb{D}_{max}$ and $\mathbb{D}_0$, the maximum ratio of score probability occurs at $|(d^i)'| \approx \infty$, and $S((d^i)'\approx\infty)\approx 0$.

Now before we prove that the n-body Laplace framework given in **Definition 5.5** satisfies differential privacy given in **Definition 3.2**, we need to calculate the PDF of the score $S(|(d^i)'| \approx \infty) \approx 0$ after the Laplace noise is added, for the maximally different neighboring trajectory distance $\mathbb{D}_{max}$ and $\mathbb{D}_0$. The PDF of the score for an arbitrary trajectory distance $\mathbb{D}$ is given by,

$$Pr(S(|(d^i)'|)\approx 0) = \int_{S((d^i)')\approx 0}\cdots\int \prod_{i=1}^{n} \frac{\exp\left\{-\frac{|\delta^i|}{3\sigma_{d^i} b^i}\right\}}{2b^i(3\sigma_{d^i})} d\delta^i, \quad (5)$$

where δ^i denotes δ^i_x and δ^i_y, b^i denotes b^i_x and b^i_y from now.

From **Definition 4.2**, we have $S = 1 - \prod_{i=1}^{n}(S^i)'$ with $(S^i)' = 1 - \exp\left\{-\frac{|d^i+\delta^i|}{R}\right\}$. Changing the variable in Eq. (5) from δ^i to $(S^i)'$, we have $d\delta^i = \mp\frac{R}{1-(S^i)'}d(S^i)'$, where the minus sign is for $d^i + \delta^i > 0$ and the positive sign is for $d^i + \delta^i < 0$. Now Eq. (5) reduces to the following,

$$\begin{aligned} Pr(S') =& \int_{S'=1-S}\cdots\int \prod_{i=1}^{n} \frac{1}{\widetilde{b^i}} \cosh\left(\frac{\widetilde{d^i}}{\widetilde{b^i}}\right) \left(1-(S^i)'\right)^{\frac{1}{\widetilde{b^i}}-1} dS' \\ =& \int_{1-S}^{1} d(S^n)' \int_{\frac{1-S}{(S^n)'}}^{1} d(S^{n-1})' \cdots \\ & \int_{\frac{1-S}{\prod_{i=3}^{n}(S^i)'}}^{1} \frac{\prod_{i=1}^{n}\frac{1}{\widetilde{b^i}}\cosh\left(\frac{\widetilde{d^i}}{\widetilde{b^i}}\right)\left(1-(S^i)'\right)^{\frac{1}{\widetilde{b^i}}-1}}{(S^2)'} d(S^2)', \quad (6) \end{aligned}$$

where $(S^i)' \approx 1$ and we have defined the normalized quantities as follows,

$$\widetilde{b^i} \equiv \frac{3\sigma_{d^i} b^i}{R}, \text{ and, } \widetilde{d^i} \equiv \frac{(d^i)'}{R}. \quad (7)$$

Lemma 6.2. The 2-node probability is given by,

$$Pr(S \approx 0|n=2) = C_{12}(\tilde{\mathbf{b}}) \left(1 - \frac{1-S}{\prod_{i=3}^{n}(S^i)'}\right)^{\frac{1}{\widetilde{b^1}}+\frac{1}{\widetilde{b^2}}-1} \quad (8)$$

with $C_{12}(\tilde{\mathbf{b}}) = \prod_{i=1}^{2} C'_i \cosh\left(\frac{\widetilde{d^i}}{\widetilde{b^i}}\right)$, and $C'_1 = \frac{1}{\widetilde{b^1}}; C'_2 = \frac{1}{\widetilde{b^2}}\left[\widetilde{b^1}(\frac{1}{2})^{\frac{1}{\widetilde{b^1}}} + \widetilde{b^2}(\frac{1}{2})^{\frac{1}{\widetilde{b^2}}}\right]$.

Proof. See **Appendix B**.

Lemma 6.3. The n-node probability $Pr(S \approx 0)$ is in the Laplace form and given by,

$$Pr(S \approx 0) \approx C(\tilde{\mathbf{b}}) \exp\left\{\frac{\ln S}{b_R}\right\},$$

with

$$\begin{aligned} b_R &= \frac{1}{\sum_{i=1}^{n}\frac{1}{\widetilde{b^i}} - 1}, \quad C(\tilde{\mathbf{b}}) = \prod_{i=1}^{n} C'_i \cosh\left(\frac{\widetilde{d^i}}{\widetilde{b^i}}\right), \\ C'_i &= \frac{1}{\widetilde{b^i}}\left[\widetilde{b^i}(\frac{1}{2})^{\frac{1}{\widetilde{b^i}}} + \frac{(\frac{1}{2})^{\sum_{j=1}^{i-1}\frac{1}{\widetilde{b^j}}}}{\sum_{j=1}^{i-1}\frac{1}{\widetilde{b^j}}}\right]. \end{aligned}$$

Proof. From **Appendix C**, we have

$$Pr(S) \approx C(\tilde{\mathbf{b}})\,(S)^{\sum_{i=1}^{n}\frac{1}{\widetilde{b^i}}-1} = C(\tilde{\mathbf{b}}) \exp\left\{\left(\sum_{i=1}^{n}\frac{1}{\widetilde{b^i}}-1\right)\ln S\right\},$$

from which **Lemma 6.3** is proved.

Theorem 6.1. The n-body Laplace framework satisfies ε-differential privacy.

Proof. According to the definition of differential privacy in **Definition 3.2**, the maximum ratio of score probability is for the maximally different neighboring trajectory distances given in **Lemma 5.1**:

$$\frac{Pr(S(\mathbb{D}'_1) \in \mathbb{S}|\theta_{\boldsymbol{\rho}})}{Pr(S(\mathbb{D}'_2) \in \mathbb{S}|\theta_{\boldsymbol{\rho}})} \leq \frac{Pr(S(\mathbb{D}'_{max}) \in \mathbb{S}|\theta_{\boldsymbol{\rho}})}{Pr(S(\mathbb{D}'_0) \in \mathbb{S}|\theta_{\boldsymbol{\rho}})}, \quad (9)$$

where $\theta_{\boldsymbol{\rho}}$ is the distribution of all possible location correlation set $\boldsymbol{\rho} = \{(\rho^i_x, \rho^i_y)|i = 1, 2, \cdots, n\}$.

From **Lemma 6.3**, we know that the probability $Pr(S \approx 0)$ has the Laplace form. Substituting Eq. (9) into Eq. (9), we have

$$\begin{aligned} \frac{Pr(S(\mathbb{D}'_1) \in \mathbb{S}|\theta_{\boldsymbol{\rho}})}{Pr(S(\mathbb{D}'_2) \in \mathbb{S}|\theta_{\boldsymbol{\rho}})} &\leq c \max_i \left\{\cosh\left(\varepsilon^i_x\right)\cosh\left(\varepsilon^i_y\right)\right\} \quad (10) \\ = c\exp\left(\varepsilon\right) \frac{[1+\exp(-2\varepsilon^m_x)]\left[1+\exp(-2\varepsilon^m_y)\right]}{4} &\leq c\exp\left(\varepsilon\right), \end{aligned}$$

with

$$\begin{cases} c = \dfrac{\prod_{i=1}^{n} C_i\left(\widetilde{\mathbf{b}}_{max}\right)}{\prod_{i=1}^{n} C_i\left(\widetilde{\mathbf{b}}_0\right)}; \quad \varepsilon = \varepsilon^m_x + \varepsilon^m_y, \\ m = \arg\max_i \left\{\cosh\left(\varepsilon^i_x\right)\cosh\left(\varepsilon^i_y\right)\right\}, \end{cases}$$

where $\widetilde{\mathbf{b}}_{max}$ and $\widetilde{\mathbf{b}}_0$ are Laplace scale parameter sets for $\mathbb{D}'_{max}$ and $\mathbb{D}'_0$, respectively. Thus we have proved **Theorem 6.1**, *i.e.*, ε-differential privacy is satisfied as follows,

$$\frac{Pr(S(\mathbb{D}'_1) \in \mathbb{S}|\theta_{\boldsymbol{\rho}})}{Pr(S(\mathbb{D}'_2) \in \mathbb{S}|\theta_{\boldsymbol{\rho}})} \leq c\exp\left(\varepsilon\right).$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2018.2853105, IEEE Transactions on Dependable and Secure Computing

6.2 Utility Analysis

Privacy and data utility form a contradicted pair that has to be balanced when noise addition mechanisms are chosen. Usually, better privacy means poorer data utility and vice versa. Here we are investigating our proposed approaches for two types of utilities to achieve the optimized privacy, *i.e.*, UD-LMDP for the location utility $\mathcal{U}_d$, and UC-LMDP for the location correlation utility $\mathcal{U}_c$.

6.2.1 Location Utility

With the help of Eq. (3), the location utility $\mathcal{U}_d$ defined in **Definition 5.6** can be obtained as follows,

$$\begin{aligned}\mathrm{E}(|\delta_x^i|) &= \int_{-\infty}^{\infty} |\delta_x^i| \, \mathrm{Pr}\left((d_x^i)' = d_x^i + \delta_x^i\right) d\delta_x^i \\ &= \frac{1}{\widetilde{b_x^i}} \int_{-\infty}^{\infty} |\delta_x^i| \frac{\exp\left\{\frac{-|\delta_x^i|}{\left(3\sigma_{d_x^i}\right) b_x^i}\right\}}{4b_x^i\left(3\sigma_{d_x^i}\right)} d\delta_x^i = 3\sigma_{d_x^i} b_x^i = R\widetilde{b_x^i}.\end{aligned}$$

Similarly, we have $\mathrm{E}(|\delta_y^i|) = R\widetilde{b_y^i}$, and with the help of Eq. (7), the location utility $\mathcal{U}_d$ is then given by

$$\mathcal{U}_d = \frac{R}{n} \sum_{i=1}^{n} \left\{ \widetilde{b_x^i} + \widetilde{b_y^i} \right\}. \tag{11}$$

6.2.2 Location Correlation Utility

With the help of Eq. (1), Eq. (3), and Eq. (4), the correlation utility given in **Definition 5.8** after the noise is added can be obtained as follows,

$$\mathcal{U}_c \;=\; \frac{1}{n} \sum_{i=1}^{n} \left\{ \rho_{x0}^i \left|1 - v_x^i\right| + \rho_{y0}^i \left|1 - v_y^i\right| \right\}, \tag{12}$$

where $v_x^i = \left\{\left[1 + 2\left(\frac{R\widetilde{b_x^i}}{\sigma_{x_a}^i}\right)^2\right]\left[1 + 2\left(\frac{R\widetilde{b_x^i}}{\sigma_{x_b}^i}\right)^2\right]\right\}^{-1/2}$ and $v_y^i = \left\{\left[1 + 2\left(\frac{R\widetilde{b_y^i}}{\sigma_{y_a}^i}\right)^2\right]\left[1 + 2\left(\frac{R\widetilde{b_y^i}}{\sigma_{y_b}^i}\right)^2\right]\right\}^{-1/2}$.

6.3 Adversary Knowledge

Our goal is to prevent adversaries from mining users' privacy. To achieve better protection, the prior knowledge should be close to the posterior knowledge. Now we investigate the adversary's prior knowledge $Pr(S \approx 0)$ and posterior knowledge $Pr(S \approx 0|\mathbb{D})$ of two users given that the adversary has known the trajectory distance $\mathbb{D}$.

6.3.1 Prior Probability

Without any knowledge of the location correlation set $\boldsymbol{\rho} = \{(\rho_x^i, \rho_y^i)|i = 1, 2, \cdots, n\}$ of the two users, the adversary might uniformly pick $\rho^i = \rho_x^i, \rho_y^i \in [0, 1]$, *i.e.*, $\pi_0\left(\rho^i\right) = 1$. From **Lemma 9**, the prior probability is given by,

$$Pr(S) \approx \int \cdots \int_{\rho^i \in [0,1]} C(\tilde{\mathbf{b}}) \exp\left\{\frac{\ln S}{b_R}\right\} d\rho^i.$$

6.3.2 Posterior Probability

When the trajectory distance $\mathbb{D}$ is released, the adversary can infer the location correlation distribution $\theta_{\boldsymbol{\rho}}$ through Bayes formula,

$$\pi\left(\rho^i|\mathbb{D}\right) = \frac{Pr\left(\mathbb{D}|\,\rho^i\right)\pi_0(\rho^i)}{Pr(\mathbb{D})} = \frac{Pr\left(\mathbb{D}|\,\rho^i\right)\pi_0\left(\rho^i\right)}{\int_{\mathbb{R}^n} Pr\left(\mathbb{D}|\,\rho^i\right)\pi_0\left(\rho^i\right)d\rho^i}.$$

Assuming that ρ^i has a uniform prior distribution $\pi_0\left(\rho^i\right) = 1$, we have the posterior distribution of ρ^i as follows,

$$\pi\left(\rho^i|\,\mathbb{D}\right) = \frac{\exp\left\{-\frac{|d^i|^2}{2(\sigma_{d^i})^2}\right\}}{\int_{\mathbb{R}^n} \exp\left\{-\frac{|d^i|^2}{2(\sigma_{d^i})^2}\right\} d\rho^i} = \frac{1}{\mathcal{N}_{\rho^i}} \exp\left\{-\frac{|d^i|^2}{2(\sigma_{d^i})^2}\right\},$$

where $\mathcal{N}_{\rho^i}$ is the normalization constant and the posterior probability of score $Pr(S)$ is given by,

$$Pr(S|\mathbb{D}) \approx \int \cdots \int_{\rho^i} Pr(S)\pi(\rho^i|\mathbb{D})d\rho^i. \tag{13}$$

6.3.3 Posterior-over-prior Knowledge Gain

Now, let's look at the posterior-over-prior knowledge gain of an adversary.

Theorem 6.2. The posterior-over-prior knowledge gain of an adversary satisfies

$$\frac{Pr(S|\mathbb{D})}{Pr(S)} \le c\exp(\varepsilon).$$

Proof. From **Lemma 9**, we can see that the prior knowledge $Pr(S)$ is not smaller than the probability $Pr(S|\mathbb{D}_0)$,

$$Pr(S) \ge Pr(S|\mathbb{D}_0),$$

from which we have,

$$\frac{Pr(S|\mathbb{D})}{Pr(S)} \le \frac{Pr(S|\mathbb{D})}{Pr(S|\mathbb{D}_0)} \le c\exp(\varepsilon),$$

where **Theorem 6.1** and Eq. (13) have been used. Thus **Theorem 6.2** is proved.

6.4 The Constrained Optimizations

In this Section, we present the constrained optimizations for both UD-LMDP and UC-LMDP.

6.4.1 Analysis for UD-LMDP

For a given $\mathcal{U}_d = u_d$ in Eq. (11), the Lagrangian for constrained optimization $\min\{\varepsilon|\mathcal{U}_d = u_d\}$ in **Section 5.2** is given by,

$$\mathcal{L}(\widetilde{\boldsymbol{b}_{u_d}}) = Pr(S|\widetilde{\boldsymbol{b}_{u_d}}) + \lambda\left[\frac{1}{n}\sum_{i=1}^{n}\widetilde{b_{u_d}^i} - u_d\right].$$

where we have used the relation of $\varepsilon \propto Pr(S|\widetilde{\boldsymbol{b}_{u_d}})$ as shown in **Theorem 6.1**.

The probability extrema can be obtained when we have the following,

$$\frac{dPr(S|\widetilde{b_{u_d}^i})}{d\widetilde{b_{u_d}^i}} = -\frac{\lambda_d}{n}; \quad \frac{1}{n}\sum_{i=1}^{n}\widetilde{b_{u_d}^i} = u_d, \tag{14}$$

with the help of the differential privacy given in Eq. (9).

What's more, we further prove that the probability $Pr(S \approx 0)$ is maximized through evaluating the determinants of the k-order principal minors H_B^{k+1} of the bordered Hessian matrix, *i.e.*, we have numerically proved that,

$$sign\left\{\left|H_B^{k+1}\right|\right\} = (-1)^k, \quad k > 2. \tag{15}$$

Lemma 6.4. The LM method given in Eq. (14) has the following solution of λ_d,

$$\frac{1}{\widetilde{b^i_{u_d}}}\left[1 + \frac{\widetilde{d^i}}{\widetilde{b^i_{u_d}}}\tanh\left(\frac{\widetilde{d^i}}{\widetilde{b^i_{u_d}}}\right) + \frac{\ln S}{\widetilde{b^i_{u_d}}}\right] = -\lambda_d, \tag{16}$$

from which we can see that $\widetilde{\boldsymbol{b}_{u_d}}$ is a function of $\widetilde{d^i}$ and λ_d.

Proof. See **Appendix D**.

Now we can obtain the relationship between $\widetilde{\boldsymbol{b}_{u_d}}$ and the given location utility u_d given in Eq. (11),

$$\frac{R}{n}\sum_{i=1}^{n}\widetilde{b^i_{u_d}}(\widetilde{d^i}, \lambda_d) \quad = \quad u_d, \tag{17}$$

from which we can first solve for λ_d and then for $\widetilde{\boldsymbol{b}_{u_d}}$.

Rigorously, Eq.(16) can only be solved numerically. However, approximate solutions exist for two cases.

Case I: When $\widetilde{d^i} << \widetilde{b^i_{u_d}}$, Eq. (16) reduces to

$$\frac{1}{\widetilde{b^i_{u_d}}}\left[1 + \left(\frac{\widetilde{d^i}}{\widetilde{b^i_{u_d}}}\right)^2 + \frac{\ln S}{\widetilde{b^i_{u_d}}}\right] = -\lambda_d, \tag{18}$$

which is a cubic function and the solution is

$$\widetilde{b^i_{u_d}} \approx A - B(\widetilde{d^i})^2. \tag{19}$$

Case II: When $\widetilde{d^i} >> \widetilde{b^i_{u_d}}$, Eq. (16) reduces to

$$\frac{1}{\widetilde{b^i_{u_d}}}\left\{1 + \frac{1}{\widetilde{b^i_{u_d}}}\left[\ln S + \widetilde{d^i}\right]\right\} = -\lambda_d', \tag{20}$$

with the solution of

$$\widetilde{b^i_{u_d}} = \frac{\sqrt{1 - 4c\left[\ln S + \widetilde{d^i}\right]} - 1}{2\lambda_d}. \tag{21}$$

We can see that $b^i \approx A - B\widetilde{d^i}$ is at large value of $\widetilde{d^i}$. As an example, Fig. 2(a) shows the plot of the theoretical values of $\mathbf{b}$ and its numerical solution for $d^i \in [-1000, 1000]$ and $\ln S = -7$. Also shown are approximations at two limit regimes: $\widetilde{d^i} << \widetilde{b^i_{u_d}}$ given in Eq. (19) and $\widetilde{d^i} >> \widetilde{b^i_{u_d}}$ given in Eq. (21).

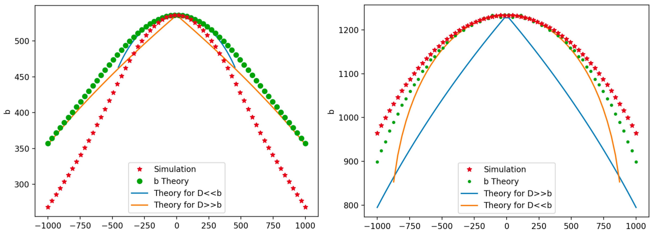
6.4.2 Analysis for UC-LMDP

Following the similar analysis given in Section 6.4.1, let's start from a location correlation utility, which is defined as **Definition 5.7**.

Following similar procedure as Eq. (14), we can obtain the optimized $\mathbf{b}_{u_c}$ for a given location correlation utility $\mathcal{U}_c = u_c$ as follows,

$$\frac{dPr}{d\rho^i_{u_c}} = \frac{-1}{\widetilde{b^i_{u_c}}}\left[1 + \frac{\widetilde{d^i}}{\widetilde{b^i_{u_c}}}\tanh\left(\frac{\widetilde{d^i}}{\widetilde{b^i_{u_c}}}\right) + \frac{\ln S}{\widetilde{b^i_{u_c}}}\right]\frac{db^i}{d\rho^i_{u_c}}, \tag{22}$$

where $\widetilde{b^i_{u_c}} = m\sigma_{d^i}$ and $\frac{db^i_{u_c}}{d\rho^i} = -\frac{m^2\sigma^i_a\sigma^i_b}{b^i_{u_c}}$.



(a) $\mathbf{b}_{u_d}$ of UD-LMDP (b) $\mathbf{b}_{u_c}$ of UC-LMDP

Fig. 2: Simulated $\mathbf{b}$ and their theoretical approximations

Applying the LM method $\min\{\varepsilon|\mathcal{U}_c = u_c\}$ in **Section 5.3** and with the help of Eq. (22), we have the following,

$$\frac{1}{(\widetilde{b^i_{u_c}})^2}\left[1 + \frac{\widetilde{d^i}}{\widetilde{b^i_{u_c}}}\tanh\left(\frac{\widetilde{d^i}}{\widetilde{b^i_{u_c}}}\right) + \frac{\ln S}{\widetilde{b^i_{u_c}}}\right] = -\lambda_c. \tag{23}$$

Therefore, we can obtain the relationship between $\widetilde{\boldsymbol{b}_{u_c}}$ and the given location correlation utility u_c in Eq. (12),

$$\frac{1}{n}\sum_{i=1}^{n}\rho_0^i\left|1 - v^i\left(\widetilde{\boldsymbol{b}_{u_c}}(\widetilde{d^i}, \lambda_c)\right)\right| = u_c,$$

form which we can solve for λ_c and then $\widetilde{\boldsymbol{b}_{u_c}}$.

Similarly, we have approximate solutions for two cases.

Case I: when $\widetilde{d^i} << \widetilde{b^i_{u_c}}$, Eq. (23) reduces to

$$\frac{1}{(\widetilde{b^i_{u_c}})^2}\left[1 + \left(\frac{\widetilde{d^i}}{\widetilde{b^i_{u_c}}}\right)^2 + \frac{\ln S}{\widetilde{b^i_{u_c}}}\right] = -\lambda_c, \tag{24}$$

which is a quartic equation with the following root,

$$\widetilde{b^i_{u_c}} \approx A - B(\widetilde{d^i})^2. \tag{25}$$

Case II: When $\widetilde{d^i} >> \widetilde{b^i_{u_c}}$, Eq. (23) reduces to

$$\frac{1}{(\widetilde{b^i_{u_c}})^2}\left\{1 + \frac{1}{\widetilde{b^i_{u_c}}}\left[\ln S + \widetilde{d^i}\right]\right\} = -\lambda_c, \tag{26}$$

whose solution is approximately given by

$$\widetilde{b^i_{u_c}} = \sqrt[3]{\frac{A - \widetilde{d^i}}{B}}. \tag{27}$$

Fig. 2(b) shows the numerical solution of $\mathbf{b}_{u_c}$ and its theoretical value, together with approximations at two limit regimes: $\widetilde{d^i} << \widetilde{b^i_{u_c}}$ in Eq. (25) and $\widetilde{d^i} >> \widetilde{b^i_{u_c}}$ in Eq. (27).

7 UD-LMDP and UC-LMCP Algorithms

In this section, we present numerical recipes of UD-LMDP and UC-LMDP, for privacy-preserving trajectories releasing of two users a and b, *i.e.*, $\{\mathbb{T}_a, \mathbb{T}_b\}$.

For location-based services, when we publish the trajectory set $\{\mathbb{T}_a, \mathbb{T}_b\}$, we cannot change raw locations too much in order to obtain high data utility. The numerical recipe is shown in **Algorithm** 1, which is summarized here: we first compute the score S and the trajectory distance for the two users. Then, in order to optimize the privacy for a given

Algorithm 1 UD-LMDP Trajectory Releasing

```
Input: Original two-user trajectory set $\{\mathbb{T}_a, \mathbb{T}_b\}$ and a given location utility $\mathcal{U}_d = u_d$.
Output: Privacy-preserving trajectory set $\{\mathbb{T}'_a, \mathbb{T}'_b\}$.
The trajectory noise set $\mathbb{N} = \emptyset$.
for Two users' daily trajectories do
    Calculate the trajectory distance $\mathbb{D}(a, b)$.
    Calculate the score $S$.
    Obtain $\lambda_d$ through Newton-Raphson method: $\frac{1}{n}\sum_1^n \widetilde{b^i_{u_d}}(\widetilde{d^i}, \lambda_d) = u_d$.
    Obtain the Laplace scale $\widetilde{\mathbf{b}_{u_d}}$ via the LM method: $\frac{1}{\widetilde{b^i_{u_d}}}\left[1 + \frac{\widetilde{d^i}}{\widetilde{b^i_{u_d}}}\tanh\left(\frac{\widetilde{d^i}}{\widetilde{b^i_{u_d}}}\right) + \frac{\ln S}{\widetilde{b^i_{u_d}}}\right] = -\lambda_d$.
    $\mathbb{D}'(a, b) = \mathbb{D}(a, b) + \boldsymbol{\delta}$ with $Pr\left(\boldsymbol{\delta}\right) = Lap\left(\boldsymbol{\delta} \middle| \widetilde{\mathbf{b}_{u_d}R}\right)$.
    $\mathbb{N} = \mathbb{N} \cup \mathbb{D}'(a, b)$
end for
$\mathbb{T}'_a = \mathbb{T}_b - \mathbb{N}$; and $\mathbb{T}'_b = \mathbb{T}_a + \mathbb{N}$.
return $\{\mathbb{T}'_a, \mathbb{T}'_b\}$
```

location utility, we use the Newton-Raphson method to get the Lagrangian parameter λ_d, from which we can obtain the optimal Laplace scale parameter set $\mathbf{b}_{u_d}$ for the longitudes or the latitudes at n time slot in each daily trajectory. Finally, we release the sanitized trajectory set $\{\mathbb{T}'_a, \mathbb{T}'_b\}$ according to the optimal Laplace noise parameter set $\mathbf{b}_{u_d}$.

Similarly, for correlation-based services, when we publish the trajectory set $\{\mathbb{T}_a, \mathbb{T}_b\}$, we do not want to change the location correlations too much. The numerical recipe is shown in **Algorithm** 2, which is summarized here: we first compute the score S and the trajectory distance for the two users. Then, in order to optimize privacy for the given location correlation utility, we use the Newton-Raphson method to get the Lagrange multiplier λ_c, from which we obtain the optimal Laplace scale parameter set $\mathbf{b}_{u_c}$ for the longitudes or the latitudes at n time slots in each daily trajectory. Finally, we release the sanitized trajectory set $\{\mathbb{T}'_a, \mathbb{T}'_b\}$ according to the optimal Laplace noise parameter set $\mathbf{b}_{u_c}$.

8 EVALUATION

We implement our simulations with Python 2.7 on a laptop with Intel Core i7-6500U, 2.59GHz, and Windows 10 system equipped with 8GB main memory. All simulation experiments are executed with real-life data [35]. The data is collected by Yonsei University in Seoul of Korea. It contains mobility data in 62 days. During our simulations, we focus on two metrics of performance: privacy and data utility. And we compare our UD-LMDP and UC-LMDP to Group-DP and MQM [10].

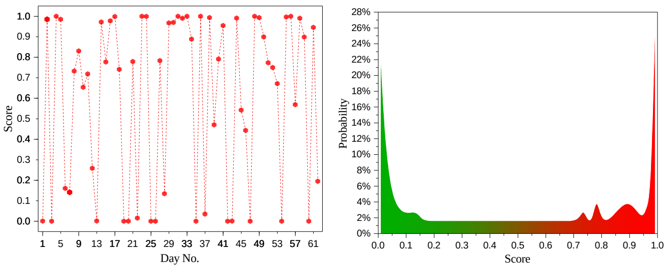
8.1 Privacy

Our goal is to protect the social relation between two users, *i.e.*, to prevent adversaries from gaining too much knowledge of the real trajectory correlation score S_0 between the two users after Laplace noise is added.

We first measure the raw trajectory correlation score between the two users. Fig. 3(a) shows the real score S_0 of each day and Fig. 3(b) shows its PDF $Pr(S_0)$. We can see from Fig. 3(b) that the PDF of the real score $S_0 > 0.5$ is higher than the score $S_0 \leq 0.5$, *i.e.*, $Pr(S_0 > 0.5) > Pr(S_0 \leq 0.5)$. So we can consider that there is a strong social relation between the two users on average for all 62 days.

Algorithm 2 UC-LMDP Trajectory Releasing

```
Input: Original two-user trajectory set $\{\mathbb{T}_a, \mathbb{T}_b\}$ and a given location utility $\mathcal{U}_c = u_c$.
Output: Privacy-preserving trajectory set $\{\mathbb{T}'_a, \mathbb{T}'_b\}$.
The trajectory noise set $\mathbb{N} = \emptyset$.
for Two users' daily trajectories do
    Calculate the trajectory distance $\mathbb{D}(a, b)$.
    Calculate the score $S$.
    Calculate the standard deviations of trajectory distance at all time slot $\{\sigma_{d^i} | i = 1, 2, \cdots, n\}$ and its correlation coefficients $\{\rho_0^i | i = 1, 2, \cdots, n\}$.
    Obtain $\lambda_c$ through Newton-Raphson method: $\frac{1}{n}\sum_{i=1}^n \rho_0^i \left|1 - v^i\left(\widetilde{\boldsymbol{b}_{u_c}}(\widetilde{d^i}, \lambda_c)\right)\right| = u_c$.
    Obtain Laplace scale $\widetilde{\mathbf{b}_{u_c}}$ via the LM method: $\frac{1}{(\widetilde{b^i_{u_c}})^2}\left[1 + \frac{\widetilde{d^i}}{\widetilde{b^i_{u_c}}}\tanh\left(\frac{\widetilde{d^i}}{\widetilde{b^i_{u_c}}}\right) + \frac{\ln S}{\widetilde{b^i_{u_c}}}\right] = -\lambda_c$.
    $\mathbb{D}'(a, b) = \mathbb{D}(a, b) + \boldsymbol{\delta}$ with $Pr\left(\boldsymbol{\delta}\right) = Lap\left(\boldsymbol{\delta} \middle| \widetilde{\mathbf{b}_{u_c}R}\right)$.
    $\mathbb{N} = \mathbb{N} \cup \mathbb{D}'(a, b)$
end for
$\mathbb{T}'_a = \mathbb{T}_b - \mathbb{N}$; and $\mathbb{T}'_b = \mathbb{T}_a + \mathbb{N}$.
return $\{\mathbb{T}'_a, \mathbb{T}'_b\}$
```

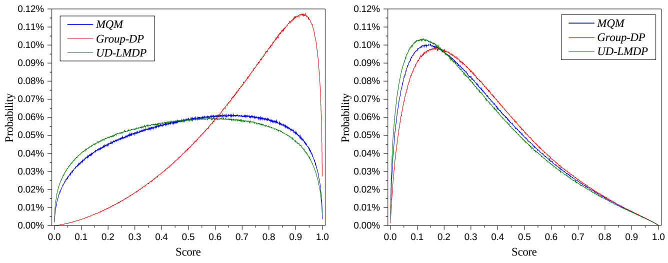


(a) The real score (S_0) (b) The real score's PDF ($Pr(S_0)$)

Fig. 3: The real score of two users and its PDF ($R = 18$)

8.1.1 The Noisy Scores' PDFs

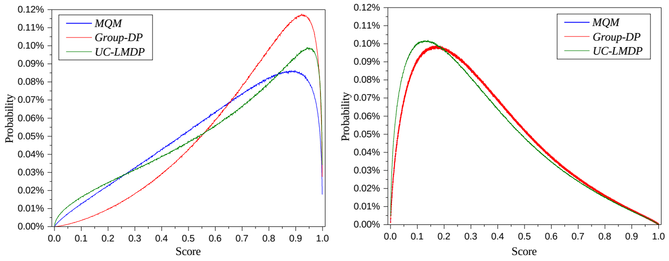
To evaluate our approaches' privacy and performance against the inference attack, PDFs of the noisy scores of two typical pairs of trajectories are studied: a strongly correlated trajectories pair on day #32 with $S_0 = 0.99$; and a weakly correlated trajectories pair on day #46 with $S_0 = 0.44$.



(a) Day #32 ($S \approx 0.99$) (b) Day #46 ($S \approx 0.44$)

Fig. 4: The noisy score's PDF $Pr(S)$ for UD-LMDP

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2018.2853105, IEEE Transactions on Dependable and Secure Computing



(a) Day #32 ($S \approx 0.99$) (b) Day #46 ($S \approx 0.44$)

Fig. 5: The noisy score's PDF $Pr(S)$ for UC-LMDP

For UD-LMDP, we use a location utility $u_d = 0.1\sigma_{\mathbb{D}}$ with $\sigma_{\mathbb{D}} = \sum_{i=1}^{n} \sigma_{d^i}/n$ being the standard deviation of the trajectory distance $\mathbb{D}$ of the two users. Two typical scenarios are studied: 1) the strong correlation case (Day #32) is shown in Fig. 4(a), from which we can see that $Pr(S \approx 0)$ of UD-LMDP is much better than that of Group-DP. To further show the performance of UD-LMDP, we compare it with MQM whose Laplace scale parameters of locations at all time slots are the same. We can also see that UD-LMDP is also better than MQM. What's more important, we can see that our UD-LMDP can provide effective protection against the inference attack: after the noise is added, the maximum score probability $Pr_{max}(S)$ is at $S \approx 0.5$, which differs from its real score value $S_0 \approx 0.99$; and 2) for the weak correlation case, *i.e.*, Day #46 in Fig. 4(b), our UD-LMDP is also better than both Group-DP and MQM. However, since the correlation is already weak, only small noise is added and the differences between different approaches are not so obvious.

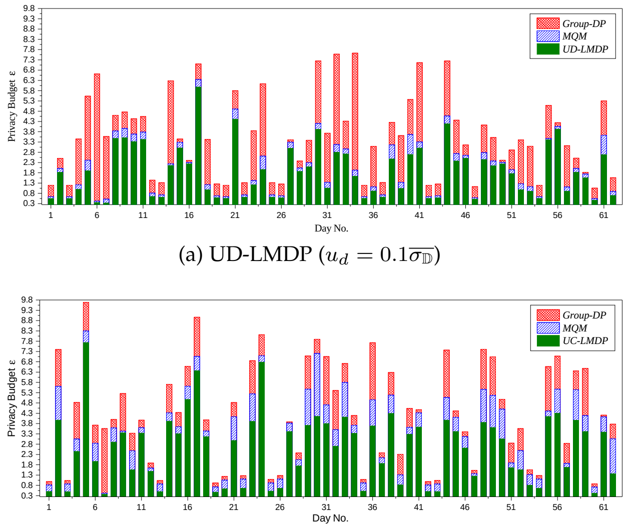
Similarly, for UC-LMDP, we use a location correlation utility $u_c = 0.2$ and study the same scenarios as UD-LMDP: 1) the strong correlation case (Day #32) is shown in Fig. 5(a), from which we can see that $Pr(S \approx 0)$ of UC-LMDP is the best among all approaches; and 2) for the weak correlation case, *i.e.*, Day #46 on Fig. 5(b), our UC-LMDP is also better than both Group-DP and MQM.

8.1.2 *Privacy Budget ε*

We further study the privacy budget ε for our proposed approaches. Note that smaller ε means better privacy. First, we evaluate ε of 62 days for the location utility $u_d = 0.1\overline{\sigma_{\mathbb{D}}}$ and the location correlation utility $u_c = 0.85$. As shown in Fig. 6(a), the ε of our proposed UD-LMDP is smaller than that of MQM and much better than the that of Group-DP. Similar results are shown in Fig. 6(b), for our proposed UC-LMDP, in comparison with both Group-DP and MQM.

Then, Fig. 7(a) and Fig. 7(b) show ε vs. $\mathcal{U}_d$ in unit of $\sigma_{\mathbb{D}}$ among our UD-LMDP, Group-DP and MQM for the two days, *i.e.*, Day #32 with the strong correlation and Day #46 with the weak correlation. We can see that our UD-LMDP is better than the other two approaches.

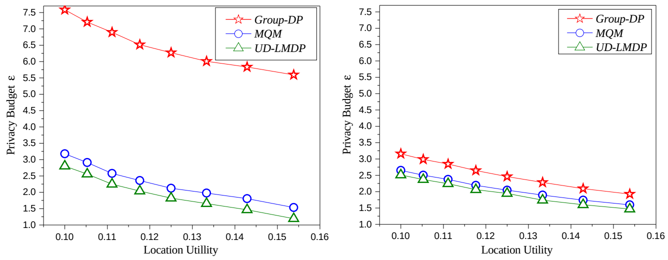
Lastly, Fig. 8(a) and Fig. 8(b) show ε vs. $\mathcal{U}_c$ among our UC-LMDP, Group-DP and MQM. We can see that our UC-LMDP is also better than both Group-DP and MQM.



(a) UD-LMDP ($u_d = 0.1\overline{\sigma_{\mathbb{D}}}$)

(b) UC-LMDP ($u_c = 0.85$)

Fig. 6: Privacy budget evaluation of 62 days



(a) Day #32 ($S \approx 0.99$) (b) Day #46 ($S \approx 0.44$)

Fig. 7: Privacy budget vs. location utility

8.2 Data Utility Evaluation

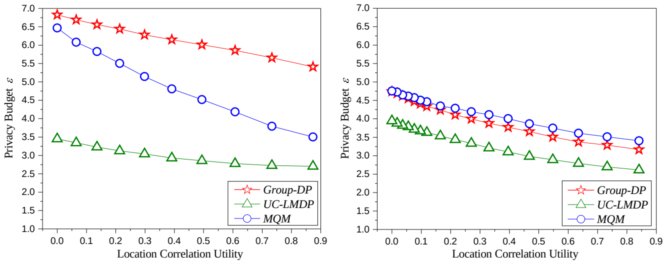
From the definitions of the location utility $\mathcal{U}_d$ and the location correlation utility $\mathcal{U}_c$, larger $\mathcal{U}_d$ or $\mathcal{U}_c$ means poorer data utility. Here, we evaluate $\mathcal{U}_d$ and $\mathcal{U}_c$ for a given privacy budget ε.

First, we evaluate the location utility $\mathcal{U}_d$ for UD-LMDP under the given privacy budget $\varepsilon = 1.5$, and the location correlation utility $\mathcal{U}_c$ for UC-LMDP under the given privacy budget $\varepsilon = 3.5$, for all 62 days. As shown in Fig. 9(a), $\mathcal{U}_d$ of UD-LMDP is smaller than both Group-DP's and MQM's, which means that the data utility of UD-LMDP is better than those of both Group-DP and MQM. Similarly, Fig. 9(b) shows the result of $\mathcal{U}_c$, from which we can see that the data utility of our proposed UC-LMDP is also better.

Then, we evaluate both $\mathcal{U}_d$ and $\mathcal{U}_c$ vs. ε for the two days, *i.e.*, Day #32 with a strong correlation and Day #46 with a weak correlation.

For the location utility $\mathcal{U}_d$: as shown in Fig. 10(a), when the correlation is strong, the $\mathcal{U}_d$ of UD-LMDP is much smaller than those of Group-DP and MQM. Meanwhile, when the correlation is weak, the $\mathcal{U}_d$ of UD-LMDP is a little bit smaller than those of both Group-DP and MQM, as shown in Fig. 10(b). Thus when the correlation is stronger, our proposed UD-LMDP can provide better data utility, compared to both Group-DP and MQM.

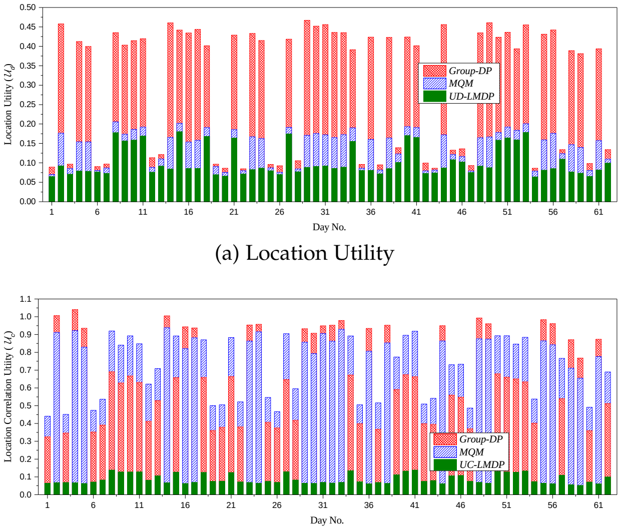
Similarly, for the location correlation utility $\mathcal{U}_c$: as shown



(a) Day #32 ($S \approx 0.99$)

(b) Day #46 ($S \approx 0.44$)

Fig. 8: Privacy budget vs. location correlation utility



(a) Location Utility

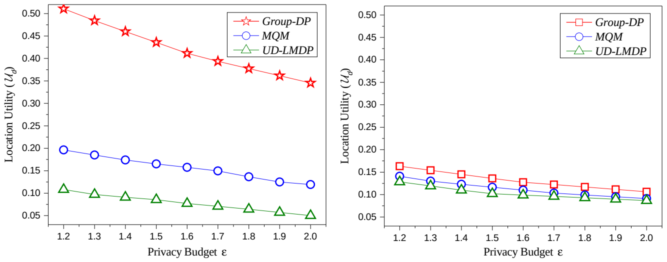
(b) Location correlation utility

Fig. 9: Data utility evaluation of 62 days

in Fig. 11, the $\mathcal{U}_c$ of our proposed UC-LMDP is much smaller than those of Group-DP and MQM; it means that the data utility of our proposed UC-LMDP is much better than those of Group-DP and MQM.

9 Conclusion

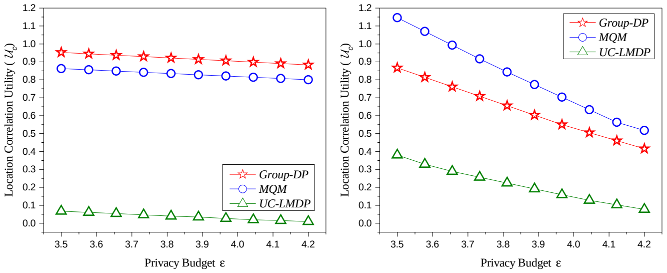
In this paper, we tackle the trajectory correlation privacy problem through the constrained optimization via the LM method. Starting from the joint Gaussian distribution of two correlated trajectories, we first prove that the n-body Laplace framework satisfies ε-differential privacy. Then, we obtain the optimized differential privacy budgets ε for two kinds of utilities, *i.e.*, the location utility for UD-LMDP; and the location correlation utility for UC-LMDP. Also, analytic formulas have been obtained for the n-element Laplace scale parameter set for both UD-LMDP and UC-LMDP. Finally, we have performed experiments of our approaches with real data, compared with the existing state-of-the-art approaches. The experimental results show that our proposed UD-LMDP and UC-LMDP do achieve better privacy for a given data utility or achieve better data utility for a given privacy budget. In the future, temporal correlations within a trajectory could be incorporated into the n-body differentially private framework for better protection of spatio-temporal correlation between two users.



(a) Day #32 ($S \approx 0.99$)

(b) Day #46 ($S \approx 0.44$)

Fig. 10: Location utility vs. privacy budget



(a) Day #32 ($S \approx 0.99$)

(b) Day #46 ($S \approx 0.44$)

Fig. 11: Location correlation utility vs. privacy budget

Acknowledgments

This work is partially supported by Key Project of National Science Foundation of China (Grant No. 61732022); National Science Foundation of China (Grant No. 61472131, 61772191); National Science Foundation (Grant CNS-1745894); and Science and Technology Key Projects of Hunan Province (Grant No. 2015TP1004, 2016JC2012).

References

[1] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in Proc. of KDD, 2011, pp. 1082-1090.

[2] L. Sweeney, "k-anonymity: A model for protecting privacy," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.

[3] R. Chen, C. M. Fung, N. Mohammed, B. C. Desai, and K. Wang, "Privacy-preserving trajectory data publishing by local suppression," Information Science, 2013, 231: 83-97.

[4] J. Hua, Y. Gao and S. Zhong, "Differentially Private Publication of General Time-Serial Trajectory Data," in Proc. of INFOCOM 2015, pp. 549-557.

[5] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in Proc. of 3rd Theory Cryptogr. Conf., 2006, pp. 265284.

[6] K. Jiang, D. Shao, St' ephane Bressan, T. Kister, and K. Tan, "Publishing trajectories with differential privacy guarantees," in Proc. of SSDBM, 2013, pp. 12.

[7] R. Chen, C. M. Fung, and B. C. Desai, "Differentially private trajectory data publication," CoRR, abs/1112.2020, 2011.

[8] R. Chen, G. Acs, and C. Castelluccia, "Differentially private sequential data publication via variable-length n-grams," in ACM Conference on Computer and Communications Security, 2012, pp. 638-649.

[9] S. Wang, and R. O. Sinnott, "Protecting personal trajectories of social media users through differential privacy," COMPUTERS & SECURITY, 2017,67: 142-163.

[10] S. Song, Y. Wang, and K. Chaudhuri, "Pufferfish Privacy Mechanisms for Correlated Data," in Proc. of ACM International Conference on Management of Data, 2017, pp. 1291-1306.

[11] V. Bindschaedler, and R. Shokri, "Synthesizing plausible privacy-preserving location traces," in Proc. of IEEE Symposium on Security & Privacy, 2016: 546-563.

[12] M. Guo, X. Jin, N. Pissinou, S. Zanlongo, B. Carbunar, and S. S. Iyengar, "In-network trajectory privacy preservation," ACM Computing Surveys, 2015, 48(2): 23.

[13] S. Gao, J. Ma, W. Shi, G. X. Zhan, and C. Sun, "TrPF: A trajectory privacy-preserving framework for participatory sensing," IEEE Transactions on Information Forensics and Security, 2013, 8(6): 874-887.

[14] T. H. You, W. C. Peng, and W. C. Lee, "Protecting moving trajectories with dummies," in Proc. of IEEE International Conference on Mobile Data Management, 2008, pp. 278-282.

[15] M. Terrovitis, G. Poulis, N. Mamoulis, and S. Skiadopoulos, "Local Suppression and Splitting Techniques for Privacy Preserving Publication of Trajectories," IEEE Transactions on Knowledge & Data Engineering, 2017, 29(7): 1466-1479.

[16] Z. Huo, X. Meng, H. Hu, and Y. Huang, "You can walk alone: Trajectory privacy-preserving through significant stays protection," in Proc. of 17th International Conference on Database Systems for Advanced Applications, 2012, pp. 351-366.

[17] S. Gao, J. Ma, C. Sun, and X. H. Li, "Balancing trajectory privacy and data utility using a personalized anonymization model," Journal of Network and Computer Applications, 2014, 38: 125-134.

[18] X. Liu, K. Liu, L. Guo, X. L. Li, and Y. G. Fang, "A game-theoretic approach for achieving k-anonymity in Location Based Services," in Proc. of INFOCOM, 2013, pp. 2985-2993.

[19] A. E. Cicek, M. E. Nergiz, and Y. Saygin, "Ensuring location diversity in privacy-preserving spatio-temporal data publishing," VLDB Journal, 2014, 23(4): 609-625.

[20] Y. Tian, W. Wang, J. Wu, Q. L. Kou, Z. Song, and E. C.-H. Ngai, "Privacy-Preserving Social Tie Discovery Based on Cloaked Human Trajectories," IEEE Transactions on Vehicular Technology, 2017, 66(2): 1619-1630.

[21] M. Li, L. Zhu, Z. Zhang, "Achieving differential privacy of trajectory data publishing in participatory sensing," INFORMATION SCIENCES, 2017, 400: 1-13.

[22] N. Wang, Y. Gu, J. Xu, "Differentially Private Event Histogram Publication on Sequences over Graphs," JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY, 2017, 32(5): 1008-1024.

[23] D. Riboni and C. Bettini, "Differentially-private release of check-in data for venue recommendation," in Proc. of 2014 IEEE International Conference on Pervasive Computing and Communications (PerCom), 2014, pp. 190-198.

[24] D. Quan, L. Yin, and Y. Guo, "Enhancing the Trajectory Privacy with Laplace Mechanism," in Proc. of 2015 IEEE Trustcom/BigDataSE/ISPA, 2015, (1): pp. 218-1223.

[25] Y. Cao and M. Yoshikawa, "Differentially Private Real-Time Data Release over Infinite Trajectory Streams," in Proc. of the 16th IEEE International Conference on Mobile Data Management. 2015, (2): pp. 68-73.

[26] L. Liao, D. J. Patterson, D. Fox, and H. Kautza, "Learning and inferring transportation routines," Artificial Intelligence, 2007, 171(5-6): 311-331.

[27] S. Qiao, C. Tang, H. Jin, T. Long, S. C. Dai, Y. C. Ku, and M. Chau, "Putmode: prediction of uncertain trajectories in moving objects databases," Applied Intelligence, 2010, 33(3): 370-386.

[28] M. Gotz, S. Nath, and J. Gehrke, "Maskit: Privately releasing user context streams for personalized mobile applications," in Proc. of SIGMOD 2012, pp. 55-58.

[29] L. Fan, L. Xiong, and V. S. Sunderam, "Differentially private multidimensional time series release for traffic monitoring," in Proc. of DBSec 2013, pp. 33-48.

[30] R. Chen, B. C. Fung, B. C. Desai, and N. M. Sossou, "Differentially private transit data publication: a case study on the montreal transportation system," in Proc. of KDD, 2012, pp. 213-221.

[31] X. He, G. Cormode, A. Machanavajjhala, C. M. Procopiuc, and D. Srivastava, "DPT: differentially private trajectory synthesis using hierarchical reference systems," in Proc. of the VLDB Endowment, 2015, 8(11): 1.

[32] X. He, N. Raval, and A. Machanavajjhala, "A demonstration of VisDPT: visual exploration of differentially private trajectories," in Proc. of the VLDB Endowment, 2016, 9(13): 1.

[33] Q. Wang, Y. Zhang, X. Lu, Z. G. Wang, Z. Qin, and K. Ren, "RescueDP: Real-time spatio-temporal crowd-sourced data publishing with differential privacy," in Proc. of IEEE INFOCOM 2016: 1-9.

[34] Q. Wang, Y. Zhang, X. Lu, Z. B. Wang, Z. Qin, and K. Ren, "Real-time and Spatio-temporal Crowd-sourced Social Network Data Publishing with Differential Privacy," IEEE Transactions on Dependable and Secure Computing, 2016, PP(99): 1-1, DOI: 10.1109/TDSC.2016.2599873.

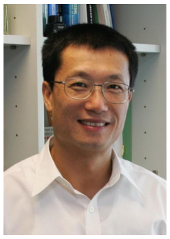
[35] Y. Chon, E. Talipov, H. Shin, and H. Cha, "Mobility prediction-based smart phone energy optimization for everyday location monitoring," in Proc. of SenSys, 2011, pp. 82-95.

Lu Ou received the B.S. degree from Changsha University of Science and Technology in computer science and the M.S. degree from Hunan University in software engineering, in 2009 and 2012, respectively. Currently, she is pursuing the Ph.D. degree in the College of Computer Science and Electronic Engineering at Hunan University, China. She is a student member of IEEE. Her research focuses on security, privacy, optimization and big data.

Zheng Qin received the Ph.D. degree in computer software and theory from Chongqing University, China, in 2001. From 2010 to 2011, he served as a Visiting Scholar at the Department of Computer Science, Michigan University. He is a Professor in the College of Computer Science and Electronic Engineering, Hunan University, where he serves as the vice dean. He also serves as the director of the Hunan Key Laboratory of Big Data Research and Application, the vice director of Hunan Engineering Laboratory of Authentication and Data Security. He is a member of China Computer Federation (CCF) and IEEE, respectively. His main interests are network and data security, privacy, data analytics and applications, machine learning, and applied cryptography.

Shaolin Liao received his Ph.D. in Electrical Engineering from the University of Wisconsin, Madison, USA, in 2008. Dr. Liao is a R&D Staff at Argonne National Laboratory, and an Adjunct Faculty at the Department of Electrical and Computer Engineering of Illinois Institute of Technology, Chicago, IL, USA. Dr. Liao was a Postdoctoral Fellow at the Department of Physics, the City University of New York (CUNY, 2008-2010). Dr. Liao received his B.S. in Materials Science and Engineering, Tsinghua University, Beijing, China, in 2000. Dr. Liao is a Senior Member of IEEE and an Associate Editor of IEEE Access. Dr. Liao's interests span the multidisciplinary areas of privacy and machine learning of big data, Simulation, Algorithms and Modeling (SAM) in signal processing, as well as novel methods in Computational Electromagnetics (CEM).

Yuan Hong received his Ph.D. degree in Information Technology from Rutgers, the State University of New Jersey. He is an Assistant Professor in the Department of Computer Science at Illinois Institute of Technology. His research interests primarily lie at the intersection of privacy, security, optimization, and data mining. His research is supported by the National Science Foundation. He is a Senior Member of IEEE.



Xiaohua Jia received his BSc (1984) and MEng (1987) from University of Science and Technology of China, and DSc (1991) in Information Science from University of Tokyo. He is currently Chair Professor with Dept of Computer Science at City University of Hong Kong. His research interests include cloud computing and distributed systems, data security and privacy, computer networks and mobile computing. Prof. Jia is an editor of IEEE Internet of Things, IEEE Trans. on Parallel and Distributed Systems (2006-2009), Wireless Networks, Journal of World Wide Web, Journal of Combinatorial Optimization, *etc.*He is the General Chair of ACM MobiHoc 2008, TPC Co-Chair of IEEE GlobeCom 2010 - Ad Hoc and Sensor Networking Symp, Area-Chair of IEEE INFOCOM 2010, 2015-2017. He is a Fellow of IEEE.