

Privacy-preserving Travel Time Prediction with Uncertainty Using GPS Trace Data

Fang Liu*, Dong Wang*, Zhengquan Xu

Abstract—The rapid growth of GPS technology and mobile devices has led to a massive accumulation of location data, bringing considerable benefits to individuals and society. One of the major usages of such data is travel time prediction, a typical service provided by GPS navigation devices and apps. Meanwhile, the constant collection and analysis of the individual location data also pose unprecedented privacy threats. We leverage the notion of geo-indistinguishability, an extension of differential privacy to the location privacy setting, and propose a procedure for privacy-preserving travel time prediction without collecting actual individual GPS trace data. We propose new concepts to examine the impact of geo-indistinguishability sanitization on the usefulness of GPS traces and provide analytical and experimental utility analysis for privacy-preserving travel time prediction. We also propose new metrics to measure the adversary error in learning individual GPS traces from the collected sanitized data. Our experiment results suggest that the proposed procedure provides travel time analysis with satisfactory accuracy at reasonably small privacy costs.

Index Terms—differential privacy, geo-indistinguishability, effective number of mapped full trajectories, usefulness, usable trajectory, continuous positioning degree, average distance

1 INTRODUCTION

1.1 Motivation and Problem

The rapid growth of GPS technology and mobile devices has led to a quick accumulation of massive location data. Analysis and understanding of the data have brought enormous benefits to individuals and society. Meanwhile, collection and processing of location data can easily expose personal behaviors, interests, social relations, or other private information, especially if combined with other data sources. de Montjoye et al. [1] studied 15-month location data from 1.5 million people and found that as little as 4 space-time points can uniquely identify 95% individuals. Meanwhile, users are often not fully aware of privacy risks from sharing their location data with service providers and how their data are used [2, 3].

One important application of GPS data is Travel time Prediction with Uncertainty (TPU). TPU examines how quickly a person arrives at a destination with a certain level of confidence. It is important for transportation and urban planning and a typical route planning service provided by navigation systems and mapping apps. TPU often relies

on continuous collection and processing of users' travel trajectories and thus exposes data contributors to privacy threats from adversaries, honest-but-curious (e.g., the service provider itself) and malicious, as depicted in Fig. 1(a)).

To our best knowledge, there is no work focusing on privacy-preserving TPU (PP-TPU) analysis. We display in Fig. 1(b) and 1(c) two possible strategies for PP-TPU. Fig. 1(b) focuses on sanitizing aggregated statistics calculated from the actual user data, say via a differentially private randomization mechanism such as the Laplace [4] or the Gaussian mechanisms [5, 6]). This strategy mitigates the privacy threats from the adversaries who aim to learn something new about their targets from the released aggregate information, but it cannot manage the privacy risk brought by the adversaries who have access to the original data, such as the service provider itself. In Fig. 1(c), sanitization occurs during data collection; that is, the true individual responses go through a sanitization mechanism locally before being shared with a third party. As a result, the true responses are only known to the users themselves.

We aim to develop a PP-TPU procedure that implements the strategy in Fig. 1(c), leveraging the state-of-the-art notions and sanitization mechanisms in data privacy research as stepping stones to achieve our goal.

1.2 Related Work

Data encryption, anonymization, and obfuscation are common frameworks for controlling the privacy risks incurred by location data collection and sharing. The PP-TPU procedure we propose can be regarded as a data obfuscation approach. Below we provide a brief overview of each framework, analyze their limitations and challenges, and state the rationale for us adopting the data obfuscation framework to develop the PP-TPU procedure.

Location encryption uses cryptographic techniques to mitigate privacy risks in location data [7–13]. This is a common approach for protecting individual privacy but can be costly in terms of computation and resource [14]. Furthermore, data, once decrypted, are no longer private to those who have the authority to access the data; the privacy:utility ratio from the data user perspective is either 100:0% or 0:100%, corresponding to the two states of encryption and decryption, respectively. These two extreme options of data access often do not meet the practical needs for data sharing. Indeed, a non-zero small privacy cost is often acceptable in practice so to create more options between the two extremes and share information with more data users.

*Co-first authors. F. Liu is Professor in the Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN, 46556, USA (corresponding author; e-mail: Fang.Liu.131@nd.edu); D. Wang (e-mail: dwang22@nd.edu) is a doctoral student and Z. Xu (e-mail: xuzq@whu.edu.cn) is Professor in the Department of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, 430079, China.

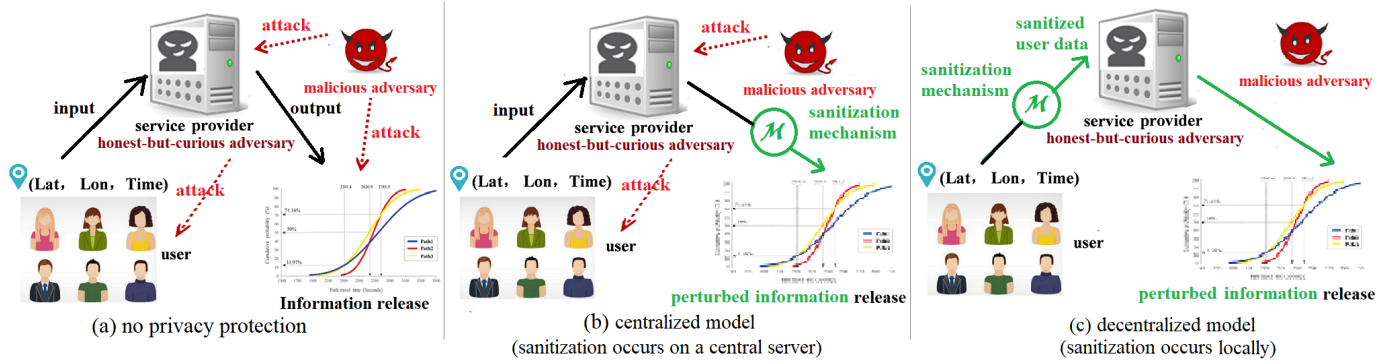


Fig. 1: Privacy Protection Strategies for Analysis using GPS Data

Data anonymization and obfuscation provide options in between the two extremes. These concepts focus on privacy-preserving data processing and analysis via methods such as data coarsening, removal of identifiers, reporting dummy locations via randomization mechanisms, among others. The key issue in these approaches is to strike a good balance between privacy loss and data utility (the higher the privacy loss, the more utility there is in the anonymized or obfuscated data relative to the original data).

Several formal privacy concepts have been developed to attain anonymization for general data, such as k -anonymity [15, 16] and l -diversity [17]. Both concepts have been adapted and applied in the location privacy setting (e.g., [18–29] for k -anonymity, [30–32] for l -diversity). Though the concepts are intuitive, neither k -anonymity nor l -diversity involves randomization; it has been shown that adversaries may still learn sensitive information or re-identify individuals from anonymized location data [1, 33].

Differential privacy (DP) [4] involves randomization and is a mathematically robust conceptual for data obfuscation. DP is proved to be robust against a wide range of adversary attacks [34, 35], processes properties such as privacy loss composability [36] and immunity to post-processing [5] that facilitate practical implementation, and has quickly become the mainstream in privacy research and applications. Big tech companies (e.g., Apple, Google, IBM) and government agencies (Census2020) also adopt DP or its variants to collect or release data. The classical DP concept is also applicable in location data analysis [37–40].

Conceptual extensions of DP for location privacy also exist, among which geo-indistinguishability (GI) [41] is perhaps the most popular. GI has been explored in a wide range of location privacy applications. Andrés et al. [41] propose the planar Laplace mechanism to perturb the 2-dimensional location data. Chatzikokolakis and Palamidessi [42] extend GI by developing an elastic indistinguishability metric that adapts the amount of injected noises according to the area density. Cunha et al. [43] propose a clustering mechanism for continuous location traces clustering application. Shi et al. [44] apply GI to preserve location information of passengers in vehicles of transportation companies. Qian et al. [45] propose a GI task allocation mechanism to preserve location privacy in mobile crowdsensing applications. Qiu et al. [46] design a strategy to minimize the loss in quality of service due to GI obfuscation. Shi et al. [47] present a closed-form relationship between localization accuracy

and the GI privacy level. Takagi et al. [48] propose Geo-Graph-Indistinguishability that extends DP to the setting of location privacy on road networks. Ren and Tang [49] present the “Expanding GI” framework to protect the privacy of vehicle locations by abstracting maps as bitmaps and utilizing linear programming to control information loss.

In summary, despite the various applications of GI in practical problems, its potential in TPU analysis has not been explored to our knowledge. The reasons for us adopting the GI concept for developing a PP-TPU procedure rather than other privacy protection frameworks for location data are as follows. First, our GI-based PP-TPU procedure avoids the limitations of encryption-based approaches, thanks to the tunable parameter (privacy budget/loss ϵ) in GI, and provides options between the two extremes of zero utility/full protection ($\epsilon \rightarrow 0$) and full utility/zero protection ($\epsilon \rightarrow \infty$). Second, while PP-TPU can leverage the classical DP concept to release aggregate location information, it often requires the collection of actual individual data (Fig. 1(b)). In contrast, GI provides a framework to perform TPU analysis without collecting actual individual GPS records (Fig. 1(c)). This strategy, compared to Fig. 1(b), better protects users’ privacy and boost users’ confidence in data sharing as the data sanitization is performed locally on users’ own devices; in addition, users can pick ϵ themselves depending on how willing they are to share their data. Third, GI follows similar mathematical reasoning as DP and many desirable properties of DP are also applicable to GI such as composability and immunity to post-processing that are important to meeting the specific challenges in the development of the PP-TPU procedure. Specifically, the TPU analysis requires the collection of multiple 3-tuple GPS records (2-dimensional location coordinates and timestamp) from a single trip in a traveler. Due to the sequential composition of privacy loss, the overall privacy loss can become unrealistically high to obtain a useful sanitized trajectory, or the sanitized trajectory is useless if the overall cost is kept reasonable. Besides, the GPS records, after sanitization, have to make sense in the 3-dimensional spatiotemporal space and for the actual road networks on which TPU is performed.

1.3 Our Contribution

Our PP-TPU procedure balances the trade-off between the privacy risk from collecting individual travel trajectories and the utility of sanitized trajectories, while being mindful of its practical feasibility. The conceptual and methodolog-

ical contributions and the potential practical impacts of the procedure are summarized as follows.

- The procedure takes into account privacy loss composability when sanitizing multiple GPS records per trajectory. To limit the overall privacy loss, it curbs the number of GPS records collected per trajectory and leverages public information of maps to filter out unusable trajectories.
- We propose new concepts of *usable set of travel trajectories*, *effective number of mapped full trajectories*, *usefulness*, and different distance deviation measures to quantify the utility of sanitized GPS records and trajectories for PP-TPU.
- We propose new metrics *continuous positioning degree* and *average distance* to quantify the adversary error in learning individual trajectories given sanitized GPS records.
- We examine the feasibility of the proposed PP-TPU procedure by quantifying the trade-off between privacy loss and the utility of sanitized GPS records and traces analytically and empirically, providing insights on choosing privacy loss parameters in different application scenarios.
- The procedure is easy to implement. Service providers and GPS navigation systems and apps may use it to collect user location data and provide TPU service with guaranteed privacy protection.

2 PRELIMINARIES

2.1 Map Matching of GPS Records

TPU starts with the collection of GPS records that contain the spatial-temporal information of a traveler and then matches the GPS locations with physical road network maps. Each GPS record contains the location P_i (latitude and longitude coordinates) and timestamp t_i information. Due to satellite signal blockage, multi-path effects, and other factors that may affect GPS signals, collected GPS location information is not always accurate. Commonly direct projections of GPS coordinates may not correspond to any meaningful real map coordinate, and road mapping algorithms often involve some level of approximation. Fig. 2 shows an example of the shortest path map-matching algorithm [50] that project 3 GPS registration points onto the physical road network.

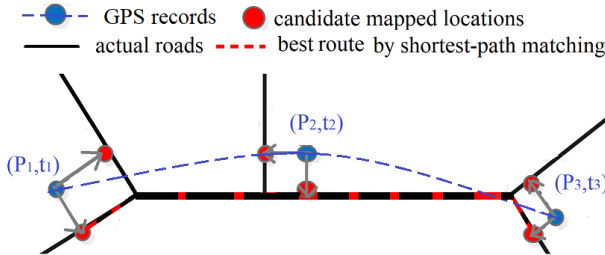


Fig. 2: Shortest path map matching

2.2 Travel Time Prediction with Uncertainty

Various approaches to travel time prediction have been developed. The naïve travel time prediction outputs a single projected travel time value, but the reachable time-space range of a traveler is rather stochastic, due to the dynamic nature of human behaviors, traffics, etc [51]. Studies [52] have shown that individuals, when facing uncertainty in travel time, tend to avoid the risk of lateness and often

reserve some time to ensure that they can arrive on time with a high level of confidence [1]. It is also important for transportation and urban planning to understand the uncertainty around travel time for infrastructure development and designs, among others.

The analysis of Travel time Prediction with Uncertainty (TPU) aims to obtain $f(t)$, the probability density function (pdf) of travel time t spent over a trip with starting point A and destination B. The probability that the destination B can be reached within time b can be easily obtained from the cumulative density function (CDF) of t , that is, $p = F_t(b) = \int_0^b f(t)dt$. For example, suppose $b = 10$ minutes and $p = 0.9$, then there is a 90% chance of arriving at the destination within 10 minutes. In reality, $f(t)$ is unknown and is often estimated by an empirical $\hat{f}(t)$ based on collected individual travel data, such as via the GPS.

2.3 Differential Privacy and Geo-indistinguishability

Differential privacy (DP) is a state-of-the-art privacy protection model that guarantees privacy for released information in mathematically rigorous terms.

Definition 1 (ϵ -differential privacy [4]). A randomization mechanism \mathcal{M} is ϵ -differentially private, if for any pair data sets X and X' that differ by one record and every possible outcome set Ω to a query,

$$\Pr[\mathcal{M}(X) \in \Omega] \leq e^\epsilon \cdot \Pr[\mathcal{M}(X') \in \Omega], \quad (1)$$

where $\epsilon > 0$ is the privacy budget or loss parameter. The smaller ϵ is, the more privacy protection there is on the individuals in the data set. X and X' differ by one record may refer to the case that X and X' are of the same size but differ in the attribute values in exactly one record, or the case that X' is one record more than X or vice versa.

The classical DP in Definition 1 provides a mathematical model for privacy guarantees when releasing aggregate statistics from a group of individuals. The local DP [53, 54] is an extension of the classical DP to a single user's data and can be used to develop mechanisms for sanitizing individual responses rather than aggregate results, with privacy.

Definition 2 (ϵ -local differential privacy [54, 55]). A randomization mechanism \mathcal{M} provides ϵ local DP if

$$\Pr[\mathcal{M}(x) \in \Omega] \leq e^\epsilon \cdot \Pr[\mathcal{M}(x') \in \Omega]. \quad (2)$$

for all pairs of an individual's possible personal data x and x' and all possible output subset Ω from \mathcal{M} .

The local DP implies that even if an adversary has access to the sanitized personal responses from a randomization mechanism that satisfies local DP, the adversary is still unable to learn much new about the user's actual responses.

GI is an extension of DP to location privacy and aims at releasing individual location records. In that sense, GI is more similar to the local DP concept than the classical DP; but all three concepts are based on similar mathematical formulations. The formal definition of GI is given below.

Definition 3 (Geo-indistinguishability [41]). Let $d(P, P')$ denote the Euclidean distance between any two distinct locations P and P' , and ϵ be the unit-distance privacy

loss. A randomization mechanism \mathcal{M} satisfies GI iff for all possible released location P^* , $\gamma > 0$, and any possible pair of P and P' within the radius of γ ,

$$\Pr(P^*|\mathcal{M}(P)) \leq e^{\epsilon\gamma} \cdot \Pr(P^*|\mathcal{M}(P')). \quad (3)$$

In other words, \mathcal{M} in Eq (3) enjoys $(\epsilon\gamma)$ -privacy for any specified γ , and the probability of distinguishing any two locations with a radius of γ , given the released location P^* , is $e^{\epsilon\gamma}$ times the probability when not having P^* . For a fixed ϵ , the larger γ is, the larger the privacy loss $(\epsilon\gamma)$ will be. For example, Tom is standing in Times Square in NYC and looking for a restaurant for lunch. He sends a request to a service provider for a list of restaurants nearby. However, he does not want to disclose his exact location and chooses to release a perturbed location P^* via GI with $\epsilon = 0.1$ per mile. The probability the service provider will identify his true location within a radius of $\gamma = 1$ mile given the perturbed location information is at most 1.1 folds of the probability when not having the information, and at most 148 folds within a radius of $\gamma = 50$ miles. In the latter case, though the probability of distinguishing locations given P^* dramatically increases compared to not having P^* , it is not practically alarming from a privacy perspective as the increase is caused by the large γ rather than a large ϵ . In other words, the service provider will have great confidence that Tom is in NYC given the released P^* , but little confidence in pinpointing exactly where in NYC. If it were the combination of $\epsilon = 50$ and $\gamma = 1$, then the probability the service provider identifying Tom's true location within a radius of 1 mile would increase by 248 folds given the perturbed location information, constituting a disastrous situation in privacy.

The planar Laplace mechanism can be used to achieve ϵ -GI by perturbing location information in polar coordinates.

Definition 4 (polar Laplace mechanism [41]). The sanitized location P^* , given the actual location P with coordinates (x, y) in the Euclidean space, satisfies GI with coordinates

$$(x^*, y^*) = (x + r \cos(\theta), y + r \sin(\theta)), \quad (4)$$

where the joint distribution of R and θ is

$$f(r, \theta) = \epsilon^2 r e^{-\epsilon r} / (2\pi). \quad (5)$$

Eq (5) implies R and θ are independently distributed and

$$r \sim \text{gamma}(2, \epsilon) = r \epsilon^2 e^{-\epsilon r} \quad (6)$$

$$\theta \sim \text{Unif}(0, 2\pi) = 1/(2\pi). \quad (7)$$

In summary, to generate a sanitized location P^* , one may draw R from the gamma distribution with shape 2 and scale ϵ^{-1} and θ from $\text{Unif}(0, 2\pi)$, and then calculate the coordinates of P^* in the Euclidean space per Eq (4).

3 PRIVACY-PRESERVING TPU WITH GI

Applied to the collection of GPS records, the GI notion can help protect individual privacy on several types of information: an individual location at a given time point, the travel trajectory of an individual over a time period, and any derived information from the collected sanitized trajectories, including TPU.

In what follows, we present a procedure to achieve PP-TPU in the framework of GI, taking into the composability of privacy costs from disclosing multiple location points from a trajectory and leveraging public knowledge of maps and road networks to improve the utility of PP-TPU on a given target road. We also examine the accuracy of sanitized information relative to the original information; analyze the privacy guarantees and indistinguishability of the proposed procedure, along with newly proposed metrics for quantifying adversary errors.

3.1 Proposed PP-TPU Procedure

We propose a new PP-TPU procedure. The PP-TPU procedure sanitizes GPS records locally via the planar Laplace mechanism to guarantee GI before the information is shared with the service provider (the strategy in Fig 1(c)). This approach mitigates the privacy risks of learning new private information about an individual from the collected GPS records for various types of adversaries, as only the users themselves possess the true responses. We also take several measures to improve the accuracy of the proposed PP-TPU procedure and to quantify the utility of sanitized trajectories, detailed below.

First, given a fixed total per-trajectory privacy cost, we limit the number of records to be collected per traveler so that the sanitization of each location record does not inject too much noise to render the sanitized record useless.

Second, we filter out non-usable trajectories for the PP-TPU given a target route R . Due to the sanitization noise injected to satisfy GI at each location, the travel direction between two consecutive time points may be opposite route R , which has a pre-specified direction. Not to bias the total travel distance, we keep the sanitized locations as is, as long as they can be mapped onto route R , but attach a sign to indicate the travel direction consistency with route R , namely, positive distance if the traveling direction is consistent with the direction of route R , negative if opposite, and 0 if the two mapped locations completely overlap. After the complete set of the GPS records from the traveler is mapped, we sum the signed distances on R for the traveler. If the summed distance is negative, then the trajectory is not usable, as defined in Definition 5.

Definition 5 (usable trajectory). A usable trajectory given a target route is a trajectory that satisfies the following two conditions: 1) at least two consecutive locations are mapped onto the target route R ; 2) the total travel distance summed over distance segments calculated from the mapped coordinates on R is non-negative. The set of usable trajectories is the *usable set* \mathcal{U} .

Third, we provide users an option to weigh different trajectories for their various levels of contribution towards the TPU on a given target route R . The motivation behind this is as follows. It is very likely that not all the GPS records will be mapped to route R , even if the traveler stays on R all the time at least for the period of interest, for a few reasons. First, GPS information is not always accurate due to satellite signal blockage and multipath effects, causing difficulty in road matching. Second, road mapping procedures themselves often involve approximation and errors.

Third, with the additional randomness introduced by the GI sanitization, the location accuracy will further decrease. Therefore, each trajectory may have a different number of GPS records mapped onto R , some of which are consecutive in times and others are not. When calculating the travel distance on R for a traveler, it makes sense to only count the distances between the locations at two consecutive time points if both are mapped onto R . One way to formulate the weight is to let it be proportional to how much a sanitized mapped trajectory overlaps with the target route.

Definition 6 (trajectory weight). Denote by d_i^* the travel distance for traveler i on the target route R of length d from the usable set \mathcal{U} . The weight that traveler i carries in the TPU is $w_i = d_i^*/d$.

Algorithm 1 lists the steps of our proposed PP-TPU procedure, with the above three utility-improvement measures implemented in various stages of the procedure. With the

Algorithm 1: The PP-TPU Procedure

input : GPS location coordinates (x_{ij}, y_{ij}) with timestamp τ_{ij} for $i=1, \dots, K$ trajectories and $j=1, \dots, n_i (\leq n$ the maximum records per trajectory); per-trajectory privacy budget ϵ_i ; target route with total distance d .

output: sanitized travel time t^* , trajectory weight w .

```

1 Usable set  $\mathcal{U} \leftarrow \emptyset$ ;
2 for  $i = 1, \dots, K$  do
3    $d_i^* \leftarrow 0$ ;  $\delta t_i \leftarrow 0$ ;
4   for  $j = 1, \dots, n_i$  do
5     Perturb  $P_j = (x_{ij}, y_{ij})$  via the planar Laplace mechanism in Eq (4) with privacy budget  $\epsilon_i/n_i$  to yield  $P_j^* = (x_{ij}^*, y_{ij}^*)$  at time  $\tau_{ij}$ ;
6     Map  $P_j^*$  onto the area map to obtain the map coordinates  $Q_j^*$ ;
7     if  $(Q_{j-1}^*, Q_j^*)$  for  $j > 1$  fall on Route  $R$  then
8       Calculate the signed Euclidean distance  $d_{ij}$  between  $Q_{j-1}^*$  and  $Q_j^*$ ;
9        $d_i^* \leftarrow d_i^* + d_{ij}$ ;  $\delta t_i \leftarrow \delta t_i + (\tau_{ij} - \tau_{i,j-1})$ ;
10    end
11  end
12  if  $d_i^* \geq 0$  then
13     $\mathcal{U} \leftarrow \mathcal{U} \cup i$ ;
14    Calculate speed  $s_i^* = d_i^*/(\delta t_i)$ , predicted travel time  $t_i^* = d/s_i^*$ , and weight  $w_i = d_i^*/d$ .
15  end
16 end

```

output weights w from Algorithm 1, we can calculate the effective number of mapped full trajectories to provide an overall metric on the impact of mapping and sanitation of GPS records on the TPU on a target road.

Definition 7 (effective number of mapped full trajectories). The effective number of mapped full trajectories is $K_{\text{eff}} = \sum_{i \in \mathcal{U}} w_i$.

Since $w_i \in [0, 1]$, $K_{\text{eff}} \leq |\mathcal{U}|$, where $|\mathcal{U}|$ is the number of trajectories in \mathcal{U} . $|\mathcal{U}|$ in turn is $\leq K$, where K is the number of raw GPS trajectories before mapping, sanitation, and filtering out. K_{eff} in a PP-TPU depends on ϵ , the number

of GPS trajectories K before mapping, and the pattern and complexity of the road networks onto which the GPS records are projected. Besides using weights to calculate K_{eff} , we can also incorporate the weights in the TPU by define a weighted version of $f_w^*(t)$. For example, we may sample K_{eff} travel times from set $(t_1^*, \dots, t_{|\mathcal{U}|}^*)$ with the sampling probabilities proportional to $\mathbf{w} = \{w_1, \dots, w_{|\mathcal{U}|}\}$ and obtain an empirical $\hat{f}_w^*(t)$ based on the samples.

3.2 Accuracy of Sanitized Information

As mentioned above, road mapping procedures per se involve approximation and errors, the quantification of which is challenging and case-dependent. As such, we focus on the accuracy of the perturbed GPS records relative to their original, instead of on the mapped coordinates. It is reasonable to assume that if sanitized and original GPS records are close, so are their mapped locations.

We quantify the closeness between a sanitized GPS location vs its original using the “usefulness” definition [41]. A location perturbation mechanism is (α, δ) -usefulness if the distance between the sanitized and original locations is $\leq \alpha$ with a probability of $1 - \delta$, for every original location. For example, for a unit-distance privacy budget $\epsilon = 2$, the probability that a sanitized location via the planar Laplace mechanism is within $\alpha = 1.5$ units of the original location is $1 - \delta = 0.8$, calculated directly from the CDF of gamma(2, 1.5). In other words, the planar Laplace mechanism of $\epsilon = 2$ GI is (1.5, 0.2)-useful for sanitizing locations. We plot the relationships between α and $1 - \delta$ for a range of ϵ values for the planar Laplace mechanism in Fig. 3(a).

In addition, we may assess the accuracy of the distance between two sanitized locations. Denote by (x_j, y_j) and $(x_{j'}, y_{j'})$ the coordinates of two recorded GPS locations at times τ_j and $\tau_{j'}$, respectively. The sanitized coordinates for the two locations via the planar Laplace mechanism in Eq (5) are respectively,

$$\begin{cases} x_j^* = x_j + r \cos(\theta), & y_j^* = y_j + r \sin(\theta) \\ x_{j'}^* = x_{j'} + r' \cos(\theta'), & y_{j'}^* = y_{j'} + r' \sin(\theta') \end{cases}, \quad (8)$$

the distance between which can be calculated by the Euclidean distance

$$\begin{aligned} d_{jj'}^{*2} &= (x_j^* - x_{j'}^*)^2 + (y_j^* - y_{j'}^*)^2 = d_{jj'}^2 + \Delta_{jj'}, \text{ where } (9) \\ \Delta_{jj'} &= r^2 + r'^2 - 2rr'(\cos(\theta_j)\cos(\theta_{j'}) + \sin(\theta)\sin(\theta')) + \\ &\quad 2(x_j - x_{j'})(r' \cos(\theta') - r \cos(\theta)) + \\ &\quad 2(y_j - y_{j'})(r' \sin(\theta') - r \sin(\theta)), \end{aligned} \quad (10)$$

and $d_{jj'}$ is the Euclidean distance between the original GPS records at times τ_j and $\tau_{j'}$. $\Delta_{jj'}$ can be regarded as the bias of the squared sanitized distance from the original distance, $d_{jj'}^*$, conditional on $d_{jj'}$ is a random variable as r, r', θ, θ' are all random variables. We propose two metrics to examine the accuracy of $d_{jj'}^*$ relative to $d_{jj'}$.

For the first metric, we define (d, α, δ) -usefulness for sanitized distances, in a similar manner to the (α, δ) -usefulness in general [56] and for sanitized locations [41].

Definition 8 ((d, α, δ)-usefulness of sanitized distance). A randomization mechanism is (d, α, δ) -useful, if there is a probability of $1 - \delta$ that the sanitized distance d^* satisfies

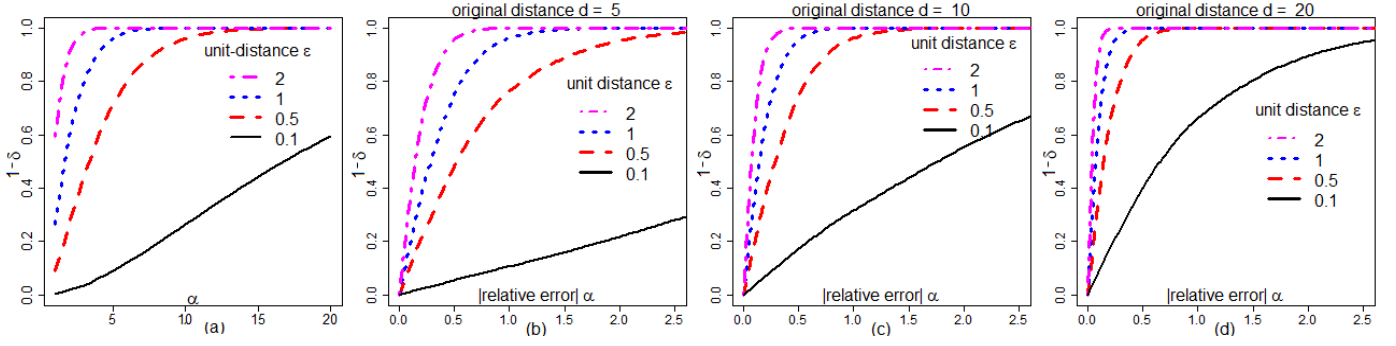


Fig. 3: Usefulness analysis on perturbed GPS location (a) and distances (b) to (d)

$|d^*/d - 1| < \alpha$ for every pair of locations with a distance of at least d .

α is the relative error of the sanitized d^* to the original d . The smaller α and the larger δ are for a given d , the more useful the mechanism is in terms of distance preservation. Figs. 3(b) to 3(d) depict the relationship between α and δ when the original distance d is 5, 10, and 20 at different levels of unit-distance privacy cost ϵ . As d increases, δ decreases for the same α . From the plots, we can claim that there is a 80% probability that the distance d^* between the perturbed locations via the planar Laplace mechanism of $\epsilon = 1$ GI is $\pm 25\%$ of $d \geq 10$; in other words, the mechanism is (10, 0.25, 0.2)-useful at $\epsilon = 1$. Similarly, we may also claim the mechanism is (5, 0.5, 0.2)-useful for $\epsilon = 1$, and (5, 1.0, 0.3)-useful for $\epsilon = 0.5$, etc.

For the second utility metric on sanitized distances, we calculate the expected %deviation $\mathbb{E}(d_{jj'}^*/d_{jj'} - 1)$ and the % root mean squared deviation (%RMSD) $\sqrt{\mathbb{E}(d_{ij}^*/d_{ij} - 1)^2}$ of the sanitized distance from the original distance, respectively. Eqs (9) and (10) suggest there is no closed-form expression for either of them; but we can always examine the numerical deviations for a given scenario. Table 1 lists the expected %deviation and %RMSD in distance for different scenarios of ϵ and d . As expected, the larger ϵ or the larger d is, the smaller the %deviation is. Also listed in the table is the expected %deviation in squared distance $\mathbb{E}(d^{*2}/d^2) - 1$, which has a closed-form solution. Specifically, $\mathbb{E}(r^2) = \mathbb{E}^2(r) + \mathbb{V}(r) = 6\epsilon^{-2}$ (\mathbb{V} denotes variance), so is $\mathbb{E}(r'^2)$; since $2(x_j - x_{j'}) (\mathbb{E}(r') \mathbb{E}(\cos(\theta')) - \mathbb{E}(r) \mathbb{E}(\cos(\theta))) + 2(y_j - y_{j'}) (\mathbb{E}(r') \mathbb{E}(\sin(\theta')) - \mathbb{E}(r) \mathbb{E}(\sin(\theta))) = 0$, then

$$\mathbb{E}(d_{jj'}^{*2} - d_{jj'}^2) = 12\epsilon^{-2} = O(\epsilon^{-2}); \quad (11)$$

$$\text{and } \mathbb{E}(d_{jj'}^{*2}/d_{jj'}^2 - 1) = 12/(d_{jj'}\epsilon)^2. \quad (12)$$

Eq (11) indicates that, in expectation, the squared distance between two sanitized GPS locations always deviates from the squared original distance by a constant $12\epsilon^{-2}$ for a given ϵ , regardless of $d_{jj'}$; however, Eq (12) implies that the deviation is not meaningful for large $d_{jj'}$.

3.3 Privacy Guarantee and Adversary Error

As illustrated in Fig 1(c), the proposed PP-TPU procedure is based on sanitized GPS trajectory data, mitigating the privacy risk from both the honest-and-curious and malicious adversaries. The employed privacy model, GI, is an extension of the notion of DP to location settings with a similar mathematical concept for controlling privacy loss

TABLE 1: Expected %deviation and %RMSD in distance, and expected %deviation in squared distance

	$\left(\frac{\mathbb{E}(d^*)}{d} - 1\right) (\%)^\dagger$			$\sqrt{\mathbb{E}(\frac{d^*}{d} - 1)^2} (\%)^\dagger$			$\left(\frac{\mathbb{E}(d^{*2})}{d^2} - 1\right) (\%)^\ddagger$		
d	50	100	200	50	100	200	50	100	200
$\epsilon = 0.01$	5.00	2.09	0.75	6.17	2.80	1.23	48	12	3
$\epsilon = 0.05$	0.51	0.13	0.03	0.95	0.47	0.24	1.92	0.48	0.12
$\epsilon = 0.25$	0.02	0.00	0.00	0.19	0.10	0.05	0.0768	0.0192	0.0048

[†] numerical results; [‡] analytical results via Eq (12).

when sharing information. DP is known to provide “provable privacy protection against a wide range of potential attacks, including those currently unforeseen” [34, 35]. The proposed PP-TPU procedure in Sec 3.1 protects several types of spatial-temporal information: the location of a traveler at a given time point, a travel trajectory of the traveler for a given time period, any calculated statistics from the trajectory (e.g, travel distance, travel speed) per the immunity property of DP and GI against post-processing. We examine each yielded privacy protection type below in detail, especially in the case of a travel trajectory.

First, per the definition of GI in Definition 3, the probability of distinguishing the true location P from any other locations with a radius of γ , given the released perturbed location P^* increases by $e^{\epsilon\gamma} - 1$ folds compared to the probability when not having P^* . In other words, the same privacy guarantees and indistinguishability as illustrated in Definition 3 apply to the GPS records collected at each timestamp for the PP-TPU.

Second, the proposed PP-TPU procedure protects the privacy of a collected travel trajectory over a time period. Though each of the location records on the trajectory is perturbed via the planar Laplace mechanism has a straightforward interpretation on indistinguishability as presented above, how to quantify the adversary error in learning about the original trajectory based on the released sanitized trajectory is less studied. Below we propose two metrics – the *average distance (AD)* and the *consecutive positioning rate (CPD)* – to quantify the adversary error and assess the effectiveness of a randomization procedure in protecting travel trajectory privacy. We apply both metrics to examine the adversary error in the experiments in Sec. 4.

Definition 9 (average distance). The average distance (AD) between the sanitized and original mapped travel trajectories on a road network is the averaged distance between the two sets of mapped locations at the same set of timestamps from the two trajectories.

We may calculate the AD empirically for a set of K mapped trajectories as follows. The original and sanitized coordinates of the mapped trajectories i ($i = 1, \dots, K$) are $\{(x_{ij}, y_{ij})\}$ and $\{(x_{ij}^*, y_{ij}^*)\}$, respectively, at time $\{\tau_{ij}\}$ for $j = 1, \dots, n_i$. The AD is given by

$$K^{-1} \sum_{i=1}^K \left(n_i^{-1} \sum_{j=1}^{n_i} d((x_{ij}, y_{ij}), (x_{ij}^*, y_{ij}^*)) \right), \quad (13)$$

Given a set $\{n_i\}$ for $i = 1, \dots, K$, the larger AD is, the larger the adversary error and the more difficult it is to recover the original trajectory from the sanitized trajectory.

Definition 10 (consecutive positioning degree). The consecutive positioning degree (CPD) $p(l)$ is a probability distribution of correctly identified l consecutive locations on a trajectory based on the released sanitized trajectory with n GPS records, for $l = 0, \dots, n$. The expected value of correctly identified positions out of n is $n_c = \sum_{l=0}^n l \times p(l)$.

We choose to examine $p(l)$, the distribution of correctly identified consecutive positions instead of correctly identified positions $p(m)$ for $m = 0, \dots, n$ (regardless of whether they are consecutive or not) because the former would be regarded by many as more revealing of travel trajectory and carrying more privacy concern than latter. How to define “correctly identified positions” is up to the user. One approach is hard-thresholding. Specifically, we choose a clip radius C . If the sanitized location falls within the circle of radius C centered at the original location, then it is claimed as correct positioning. The smaller C is, the harder it is to meet the criterion, but the more meaningful “correct” is. Since each location on a trajectory is perturbed independently via the polar Laplace mechanism, with the hard-thresholding rule, the probability of correctly identifying a location can be determined analytically, which is $p = F(C; 2, \epsilon/n)$, where n is the number of recorded positions on a GPS trajectory and F is the CDF of gamma(2, ϵ/n).

The number of correctly identified positions m given p follows $m \sim \text{Binomial}(n, p)$. As for the distribution of CPD l , we can leverage $\text{Binomial}(n, p)$ to express $p(l)$ analytically when n is small, but $p(l)$ for $1 \leq l < n - k$ with small $k \geq 0$ becomes less tractable as n increases considering that a trajectory may contain multiple location strings of different l . For example, a GPS trace with $n = 10$ records may have 2 occurrences of $l = 1$, 1 occurrence of $l = 2$, and 1 occurrence of $l = 3$. For cases where analytical calculation of $p(l)$ becomes difficult, we design Algorithm 2 that uses Monte Carlo (MC) simulations to calculate $p(l)$. Though the algorithm is presented with the hard-thresholding rule for correct positioning (line 4), the steps are applicable to other definitions of correct positioning. $n_i^{(l)}$ in the algorithm refers to the frequency distribution l in trajectory i , its average over K trajectories gives the MC estimate $p(l)$. The algorithm also outputs \bar{n}_c , the MC estimate of the expected value of correctly identified positions n_c in Definition 10.

4 EXPERIMENTS

We conduct four experiments to investigate empirically the impact of sanitization of GPS trajectories on the utility of TPU in four road network scenarios. In each experiment, there is a pre-specified target route on which the TPU

Algorithm 2: Calculation of CPD $p(l)$

input : K GPS trajectories and their sanitized counterparts with n records per trajectory; clip radius C

output: $n_i^{(l)}$ for $i = 1, \dots, K$;
 $p(l) = \sum_{i=1}^K n_i^{(l)} \left(\sum_{l=0}^n \sum_{i=1}^K n_i^{(l)} \right)^{-1}$;
 $\bar{n}_c = K^{-1} \sum_{i=1}^K \sum_{l=0}^n (n_i^{(l)} \times l)$.

```

1 for  $i = 1, \dots, K$  do
2   for  $j = 1, \dots, n$  do
3     Calculate the distance  $d_{ij}$  between sanitized
       location  $P_{ij}^*$  and original location  $P_{ij}$ ;
4     If  $d_{ij} \leq C$ , then  $e_{ij} = 1$ ; else  $e_{ij} = 0$ ;
5     Let  $e_{i0} = 0$  and  $e_{i,n+1} = 0$ ;
6   end
7   If  $e_{ij'} = 0 \forall j' = 1, \dots, n$ , then  $n_i^{(0)} = 1$ ; else
      $n_i^{(0)} = 0$ ;
8   If  $e_{ij'} = 1 \forall j' = 1, \dots, n$ , then  $n_i^{(n)} = 1$ ; else
      $n_i^{(n)} = 0$ ;
9   for  $l = 1, \dots, n-1$  do
10     $n_i^{(l)} \leftarrow 0$ ;
11    for  $j = 1, \dots, n-l+1$  do
12      if  $(e_{ij'} = 1 \forall j' = j, \dots, j+l-1) \& (e_{i,j-1} = 0) \& (e_{i,j+l} = 0)$  then
13         $n_i^{(l)} \leftarrow n_i^{(l)} + 1$ 
14      end
15    end
16  end
17 end
```

analysis is performed. We examine the utility of PP-TPU for a range ϵ values and assess the adversary error in learning individual trajectories. Though a privacy-preserving travel time distribution may also be obtained by sanitizing the original empirical distribution via a DP mechanism, as illustrated in Fig. 1(b), the server needs to collect the actual individual GPS locations, and the sanitization is processed on the server. Therefore, this approach does not provide the same privacy guarantees as the decentralized and local approach (Fig. 1(c)) taken by Algorithm 1. Since it is impossible to match the level of privacy protection between the two approaches, the utility comparison would not be as meaningful; therefore, we choose not to compare our PP-TPU approach with the DP-based approach in the experiments.

4.1 Experiment Settings

In Experiment 1, the simulated road network contains a single road. In Experiments 2, the simulated road network contains three parallel roads with one being the target route. In Experiment 3, the road network is around a large roundabout in the town of Creteil in France (Fig 4(a)); the target route AB is about 1.5 kilometers long. In Experiment 4, we examine a region in the San Francisco Bay Area (Fig 4(b)); the target road AB is about 50 kilometers long.

The GPS trajectory data in Experiments 1 (1,000 trips) and 2 (1,000 trips on the target road) are simulated as follows. We first simulated speeds from the inverse Weibull distribution with mean $\mu = 24$ meter per second and variance $\sigma^2 = 8$ (the values are chosen to mimic some common real traffic

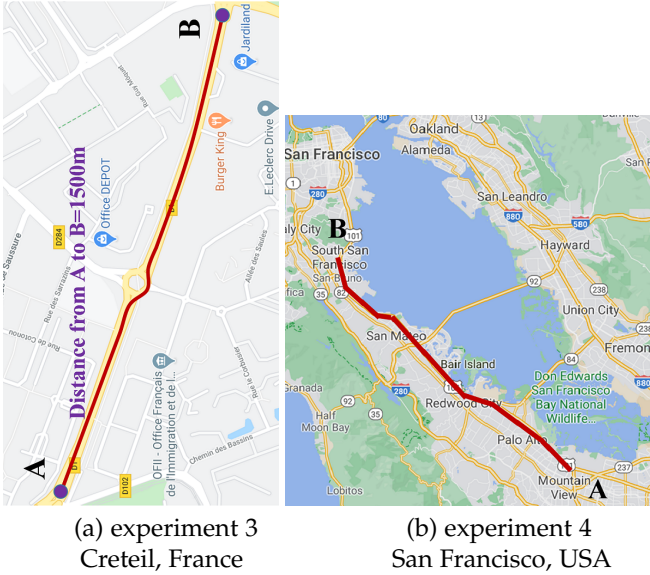


Fig. 4: Area maps in Experiments 3 and 4 (source: Google Map). AB is the target route for TPU in each experiment.

speed distributions). Each simulated speed corresponds to one trip, on which 10 location records were generated at a fixed timestamp of every $\tau = 20$ seconds, leading to travel trajectories of different lengths, depending on the speed. The vehicular mobility trace data in Experiment 3 [57] are downloadable from <http://vehicular-mobility-trace.github.io/> and contains 857,136 sets of location coordinates per second from around 5102 trips during the morning rush hour (7 to 9 AM), simulated based on real data. We randomly chose 1,000 trips within the rectangle bounded by the coordinates of the ends points A and B of the target route. The dataset in Experiment 4 [58] contains real mobility traces of taxi cabs and is downloadable from <http://crawdad.org/epfl/mobility/20090224/index.html>. It contains the GPS coordinates of approximately 500 taxis over 30 days. For this experiment, we used a subset of 30,900 location-time GPS records over the morning rush hours (8 to 9 AM) from 419 trips. In Experiments 3 and 4, we set the maximum number of GPS records per trip at 10 so to control the privacy loss per traveler. If a traveler has ≤ 10 records, we used all of them; otherwise, we randomly sampled 10 records or had 10 records spaced equally over the trajectory if there were enough records to allow that.

4.2 Sanitization and Implementation Details

The GPS records were sanitized via the planar Laplace mechanism and projected into the road map in each experiment using the shortest path algorithm. The PP-TPU was then conducted via algorithm 1 in each experiment. For the GI sanitization, we set the per-location per-meter privacy loss at 0.005, 0.01, 0.03, 0.05, and 0.08 in all 4 experiments. Since the maximum of GPS records per trip is 10, the total privacy cost for releasing a trajectory is $\leq 0.05, 0.1, 0.3, 0.5, 0.8$, respectively.

Fig 5 presents some examples of the sanitized GPS records, and the mapped travel trajectory on road networks given the GPS records. Take Experiment 2 as an example, which plots the GPS records and mapped trajectories from three different travelers. The target road is road 1 for which the TPU analysis is performed. If there was no privacy concern,

the three travelers would share their GPS records (blue circles) with the service provider who would then project the records using a mapping algorithm onto the road network and use the usable travel trajectory on road 1 to calculate travel time and carry out TPU. In this example, the mapped trajectories (cyan lines) fall on the target road for all three travelers. For PP-TPU, the service provider does not collect the actual GPS records, but only their sanitized versions (red squares). The mapping procedure and TPU analysis are the same as in the non-private setting. Since the sanitized GPS records deviate from their original counterparts, it is almost certain the mapped trajectories also deviate from the non-private trajectories. For traveler 1 (the leftmost plot), all ten sanitized GPS records are mapped onto road 1 and can be used for the subsequent travel time calculation. For traveler 2, eight out of the ten sanitized GPS records are mapped on road 1 and two on the nearby road 2. The eight records on road 1 form two location strings of length $l = 4$ and $l = 2$, respectively, that are used for the subsequent PP-TPU analysis. For traveler 3, three out of the ten sanitized GPS records are mapped onto road 1 but none of the two are consecutive in time, so traveler 3 does not contribute toward the PP-TPU. In summary, out of the sanitized trajectories from the three travelers, only those from travelers 1 and 2 contribute to \mathcal{U} .

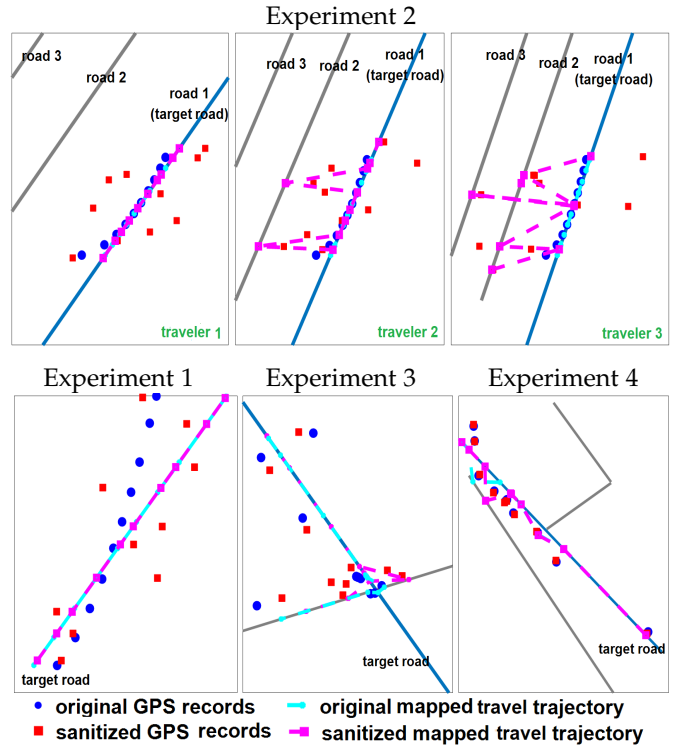


Fig. 5: Examples of sanitized GPS records and mapped travel trajectories at per-trajectory privacy cost of $\epsilon = 0.1$

4.3 Utility and PP-TPU Results

Fig 6 presents the empirical CDFs of the privacy-preserving travel times in the four experiments. As expected, the sanitization deviates the travel time distribution $\hat{f}^*(t)$ from the original $\hat{f}(t)$; the smaller per-trajectory privacy cost ϵ is, the more deviation there is. At $\epsilon \geq 0.3$, $\hat{f}^*(t)$ is close to $\hat{f}(t)$ and satisfactory utility can be reached for PP-TPU in all experiments. From the CDF curves, we can read how quickly

a traveler arrives at the destination with a certain level of confidence, and vice versa. For example, in Experiment 4, there is an 80% probability that a traveler finishes the trip AB within 100 minutes if $\epsilon = 0.5$ is used. In addition to the unweighted TPU in Fig 6, we also performed the weighted TPU analysis; the results are presented in Fig 7. A similar overall trend across ϵ is observed as in the non-weighted setting. In experiments 1 and 2, the weighting seems to affect $\hat{f}^*(t)$ more for smaller ϵ , and the left tail of $\hat{f}^*(t)$ (smaller t) is more sensitive to the weighting than the right tail. In experiments 3 and 4, the weighted distributions are similar to the unweighted version across all ϵ .

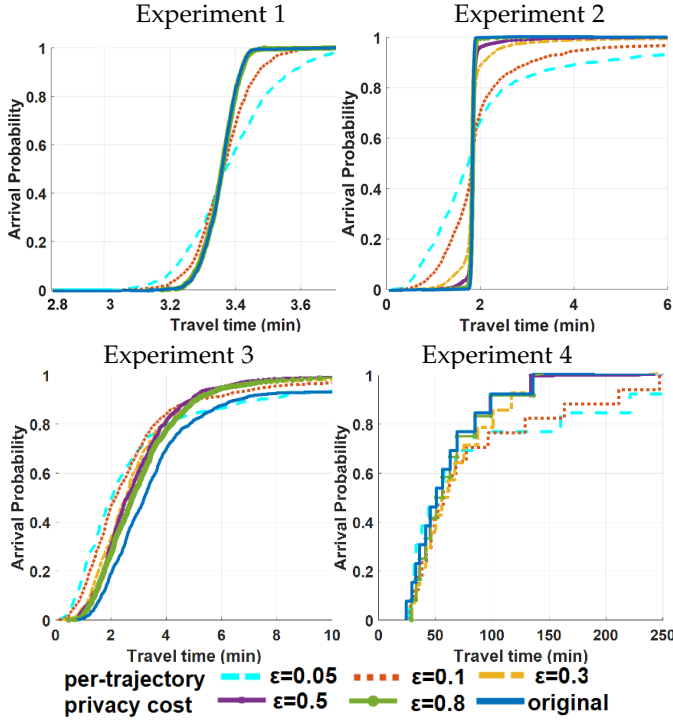


Fig. 6: PP-TPU

Table 2 presents the effective number of mapped full trajectories K_{eff} . Due to the inherent error of the mapping algorithm, not every GPS record can be mapped onto the actual route where the traveler is on, or give a sensible trajectory after mapping. Therefore, K_{eff} is smaller the number of trips even without any GI sanitization. With the GI sanitization and as ϵ decreases, K_{eff} decreases, as expected.

TABLE 2: Effective number of mapped full trajectories K_{eff}

	experiment	ϵ					original (no sanitization)	# trips
		0.05	0.1	0.3	0.5	0.8		
K_{eff}	1	792	853	873	889	892	901	1,000
	2	682	721	820	834	845	876	1,000
	3	229	314	435	460	478	513	1,000
	4	45	49	52	52	53	53	419

In summary, we can draw the following conclusions from the utility analysis in this subsection. (1) The quality of the PP-TPU analysis relates to the type and structure of the road network onto which the GPS records are mapped; some road networks are more sensitive to ϵ than others in the utility of sanitized trajectories. (2) The difference between the unweighted and weighted TPU analysis diminishes as ϵ increases. (3) The CDFs of the privacy-preserving travel time

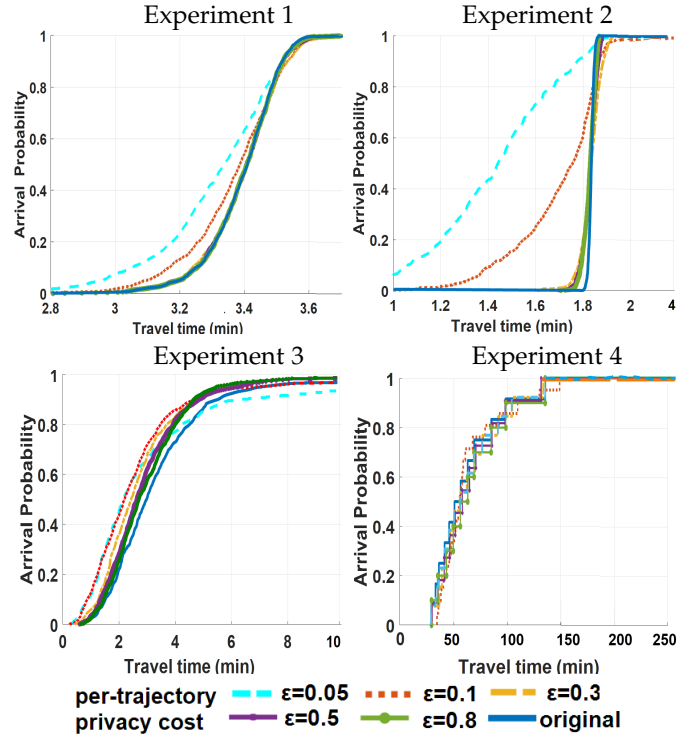


Fig. 7: Weighted PP-TPU

in the 4 experiments are similar to the original CDFs with the per-trajectory ϵ as small as 0.3, so is the effective number of mapped full trajectories, implying useful TPU analysis can be achieved with satisfactory privacy guarantees.

4.4 Adversary Error

Table 3 shows the expected AD between a sanitized and its original mapped trajectories calculated via Eq (13). Note that the 100 repeats were generated differently for experiments 1 and 2 vs. experiments 3 and 4 because the former two are synthetic data while the latter two are quasi-real and real open-source datasets, respectively. Specifically, in experiments 1 and 2, we generated 100 GPS data sets per the simulation setting in Sec 4.1; in experiments 3 and 4, the 100 repeats were obtained by performing 100 sets of sanitization on a fixed GPS dataset in each experiment. As a result, the variability of AD comes from two sources – sampling error and sanitation error – in experiments 1 and 2 and contains only the sanitization error in experiments 3 and 4.

The first observation is that the smaller ϵ is, the larger the distance is, as expected. Second, the AD value varies across the experiments for the same ϵ , which makes sense, as the AD works with the distance between a pair of locations on a map and the road network matters. Given that the road networks differ in the four experiments, it is not surprising that the AD varies by experiment. Third, the adversary error measured by the AD at $\epsilon \leq 0.3$ is sufficiently large per location on a trajectory for each experiment (≥ 30 meters).

Fig. 8 presents the probability distributions of CPD l and the correctly identified positions m (whether consecutive or not) for three different clip radius C (20, 40, and 80 meters) when the number of records per trajectory $n = 10$ for different ϵ . Since all 4 experiments used the same n and ϵ value, the results in Fig. 8 apply to all four experiments. As expected, as C increases (the criterion for claiming correct positioning

TABLE 3: Mean (SD) average distance between mapped locations on sanitized and original trajectories (100 repeats)

	experiment	ϵ				
		0.05	0.1	0.3	0.5	0.8
AD [†] (meters)	1	180 (6.7)	87 (2.0)	30 (0.9)	18 (0.3)	11 (0.1)
	2	1189 (16.4)	814 (6.8)	435 (4.9)	341 (4.5)	296 (2.4)
	3	739 (24.1)	438 (32.3)	185 (31.4)	121 (5.4)	99 (2.1)
	4	214 (21.5)	108 (11.4)	38 (3.7)	23 (2.1)	13 (1.4)

loosens) or as per-trajectory ϵ increases, the adversary's accuracy for correctly identifying more positions and more consecutive positions increase. In the case of $C = 80$ meters – a rather relaxed criterion for correct identification, the probability of identifying 10 positions out of 10 is $> 80\%$. The probability decreases to $\sim 10\%$ for $C = 40$ meters and $\sim 0\%$ for $C = 20$ meters. The plots also illustrate the differences between CPD l and the number of correctly identified locations m . For example, for $C = 20$, $\text{Pr}(l=6)$ is close to 0% , but $\text{Pr}(m=6)$ is $\sim 20\%$, regardless of whether the 6 positions are consecutive or not.

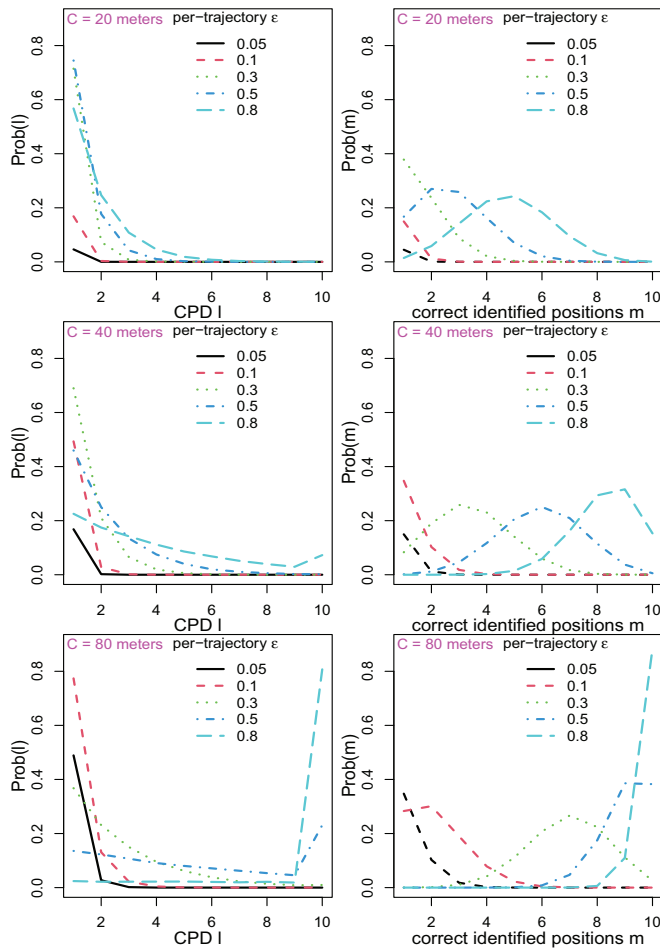


Fig. 8: Probability distributions of CPD l (left column) and correctly identified positions m (right column)

In summary, we can draw the following conclusions from the adversary error analysis in this subsection. (1) The magnitude of the adversary error closely relates to the road network type and structure. (2) The adversary error in reconstructing a trajectory from the sanitized trajectory around $\epsilon \leq 0.3$ is sufficiently large per the measures of AD and CPD to not pose serious privacy threats. (3) Taken together with the observations in the utility analysis, a good trade-off between the PP-TPU utility and privacy protection

can be achieved at per-trajectory $\epsilon \approx 0.3$ with ≤ 10 GPS records per trajectory in these 4 experiments. Users of the PP-TPU procedure can run similar analysis and choose an ϵ that leads to a good balance between utility and privacy protection for their specific problems.

5 CONCLUSIONS

This paper addresses privacy-preserving TPU analyses. We employ the notation of GI to protect individual GPS spatial-temporal records and the subsequent TPU analysis. The proposed PP-TPU procedure can be adopted by service providers (e.g., mobile phone companies, GPS navigator apps) at the GPS data collection stage. We define the effective number of mapped full trajectories, the usefulness concept, and different types of deviations in distance measures based on sanitized GPS records to quantify the utility of the sanitized trajectories. We also propose the concepts of average distance and consecutive positioning degree to assess the adversary error based on released GPS trajectory records. Our analytical results and empirical studies suggest that it is feasible to employ the GI concept to collect and release GPS information for TPU analysis while guaranteeing location privacy for the individuals who contribute their GPS data. Our future work will look into incorporating the dependency among the location points on the same travel trajectory and better utilizing the public road network maps to develop new randomization mechanisms of better utility without comprising privacy.

ACKNOWLEDGMENTS

Fang Liu is supported by NSF Grant #1717417 and Dong Wang is supported by the China Scholarships Council program (NO. 201906270230) and NSFC Grant #41971407. We also thank the editor, associate editor, and five reviewers for their useful comments and suggestions on the manuscript.

REFERENCES

- [1] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific reports*, vol. 3, p. 1376, 2013.
- [2] P. Coppens, L. Claeys, C. Veeckman, and J. Pierson, "Privacy in location-based social networks: Researching the interrelatedness of scripts and usage," in *Proceedings of the Symposium on Usable Privacy and Security*, 2014.
- [3] C. Kessler and G. McKenzie, "A geoprivacy manifesto," *Transactions in GIS*, vol. 22, no. 1, pp. 3–19, 2018.
- [4] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of cryptography conference*. Springer, 2006, pp. 265–284.
- [5] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, 2014.
- [6] F. Liu, "Generalized gaussian mechanism for differential privacy," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 4, pp. 747–756, 2018.
- [7] G. Zhong, I. Goldberg, and U. Hengartner, "Louis, lester and pierre: Three protocols for location privacy,"

- in *International Workshop on Privacy Enhancing Technologies*. Springer, 2007, pp. 62–76.
- [8] R. A. Popa, A. J. Blumberg, H. Balakrishnan, and F. H. Li, “Privacy and accountability for location-based aggregate statistics,” in *Proceedings of the 18th ACM conference on Computer and communications security*, 2011, pp. 653–666.
- [9] Z. Zhu and G. Cao, “Applaus: A privacy-preserving location proof updating system for location-based services,” in *2011 Proceedings IEEE INFOCOM*. IEEE, 2011, pp. 1889–1897.
- [10] S. Mascetti, D. Freni, C. Bettini, S. Wang, and S. Jajodia, “Privacy in geo-social networks: proximity notification with untrusted service providers and curious buddies,” *The VLDB journal*, vol. 20, no. 4, pp. 541–566, 2011.
- [11] M. Li, L. Zhu, and X. Lin, “Privacy-preserving traffic monitoring with false report filtering via fog-assisted vehicular crowdsensing,” *IEEE Transactions on Services Computing*, 2019.
- [12] L. Zhu, C. Zhang, C. Xu, X. Du, N. Guizani, and K. Sharif, “Traffic monitoring in self-organizing vanets: A privacy-preserving mechanism for speed collection and analysis,” *IEEE Wireless Communications*, vol. 26, no. 6, pp. 18–23, 2019.
- [13] C. Zhang, L. Zhu, C. Xu, X. Du, and M. Guizani, “A privacy-preserving traffic monitoring scheme via vehicular crowdsourcing,” *Sensors*, vol. 19, no. 6, p. 1274, 2019.
- [14] P. Zhou, W. Chen, S. Ji, H. Jiang, L. Yu, and D. Wu, “Privacy-preserving online task allocation in edge-computing-enabled massive crowdsensing,” *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7773–7787, 2019.
- [15] P. Samarati, “Protecting respondents identities in microdata release,” *IEEE transactions on Knowledge and Data Engineering*, vol. 13, no. 6, pp. 1010–1027, 2001.
- [16] L. Sweeney, “k-anonymity: A model for protecting privacy,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.
- [17] A. Machanavajhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, “l-diversity: Privacy beyond k-anonymity,” *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, pp. 3–es, 2007.
- [18] M. Gruteser and D. Grunwald, “Anonymous usage of location-based services through spatial and temporal cloaking,” in *Proceedings of the 1st international MobiSys conference*, 2003, pp. 31–42.
- [19] H. Kido, Y. Yanagisawa, and T. Satoh, “Protection of location privacy using dummies for location-based services,” in *21st International Conference on Data Engineering Workshops (ICDEW’05)*. IEEE, 2005, pp. 1248–1248.
- [20] B. Gedik and L. Liu, “Protecting location privacy with personalized k-anonymity: Architecture and algorithms,” *IEEE Transactions on Mobile Computing*, vol. 7, no. 1, pp. 1–18, 2007.
- [21] G. Ghinita, P. Kalnis, and S. Skiadopoulos, “Prive: anonymous location-based queries in distributed mobile systems,” in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 371–380.
- [22] H. Lu, C. S. Jensen, and M. L. Yiu, “Pad: privacy-area aware, dummy-based location privacy in mobile services,” in *Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*, 2008, pp. 16–23.
- [23] C.-Y. Chow, M. F. Mokbel, and W. G. Aref, “Casper* query processing for location services without compromising privacy,” *ACM Transactions on Database Systems (TODS)*, vol. 34, no. 4, pp. 1–48, 2009.
- [24] A. Gkoulalas-Divanis, P. Kalnis, and V. S. Verykios, “Providing k-anonymity in location based services,” *ACM SIGKDD explorations newsletter*, vol. 12, no. 1, pp. 3–10, 2010.
- [25] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, “Achieving k-anonymity in privacy-aware location-based services,” in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*. IEEE, 2014, pp. 754–762.
- [26] F. Fei, S. Li, H. Dai, C. Hu, W. Dou, and Q. Ni, “A k-anonymity based schema for location privacy preservation,” *IEEE Transactions on Sustainable Computing*, vol. 4, no. 2, pp. 156–167, 2017.
- [27] P. Zhao, J. Li, F. Zeng, F. Xiao, C. Wang, and H. Jiang, “Illia: Enabling k-anonymity-based privacy preserving against location injection attacks in continuous lbs queries,” *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1033–1042, 2018.
- [28] J. Wang, Z. Cai, and J. Yu, “Achieving personalized k-anonymity-based content privacy for autonomous vehicles in cps,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4242–4251, 2019.
- [29] S. Zhang, X. Li, Z. Tan, T. Peng, and G. Wang, “A caching and spatial k-anonymity driven privacy enhancement scheme in continuous location-based services,” *Future Generation Computer Systems*, vol. 94, pp. 40–50, 2019.
- [30] M. Terrovitis and N. Mamoulis, “Privacy preservation in the publication of trajectories,” in *The Ninth International Conference on Mobile Data Management (mdm 2008)*. IEEE, 2008, pp. 65–72.
- [31] F. Liu, K. A. Hua, and Y. Cai, “Query l-diversity in location-based services,” in *2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*. IEEE, 2009, pp. 436–442.
- [32] Z. Tu, K. Zhao, F. Xu, Y. Li, L. Su, and D. Jin, “Protecting trajectory from semantic attack considering k-anonymity, l-diversity, and t-closeness,” *IEEE Transactions on Network and Service Management*, vol. 16, no. 1, pp. 264–278, 2018.
- [33] S. Wang, Q. Hu, Y. Sun, and J. Huang, “Privacy preservation in location-based services,” *IEEE Communications Magazine*, vol. 56, no. 3, pp. 134–140, 2018.
- [34] K. Nissim, T. Steinke, A. Wood, M. Altman, A. Beimbenek, M. Bun, M. Gaboardi, D. R. O’Brien, and S. Vadhan, “Differential privacy: A primer for a non-technical audience,” in *Privacy Law Scholars Conf*, vol. 3, 2017.
- [35] C. Dwork, A. Smith, T. Steinke, and J. Ullman, “Exposed! a survey of attacks on private data,” *Annual Review of Statistics and Its Application*, vol. 4, pp. 61–84, 2017.
- [36] F. McSherry and K. Talwar, “Mechanism design via differential privacy,” in *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS’07)*. IEEE,

- 2007, pp. 94–103.
- [37] M. Asada, M. Yoshikawa, and Y. Cao, ““when and where do you want to hide?”—recommendation of location privacy preferences with local differential privacy,” in *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, 2019, pp. 164–176.
- [38] S. Bavadekar, A. Dai, J. Davis, D. Desfontaines, I. Eckstein, K. Everett, A. Fabrikant, G. Flores, E. Gabrilovich, and K. Gadepalli, “Google covid-19 search trends symptoms dataset: Anonymization process description (v1.0),” *arXiv preprint arXiv:2009.01265*, 2020.
- [39] A. Aktay, S. Bavadekar, G. Cossoul, J. Davis, D. Desfontaines, A. Fabrikant, E. Gabrilovich, K. Gadepalli, B. Gipson, and M. Guevara, “Google covid-19 community mobility reports: Anonymization process description (v1.0),” *arXiv preprint arXiv:2004.04145*, 2020.
- [40] Y. Xiao and L. Xiong, “Protecting locations with differential privacy under temporal correlations,” in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015, pp. 1298–1309.
- [41] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, “Geo-indistinguishability: Differential privacy for location-based systems,” in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, 2013, pp. 901–914.
- [42] K. Chatzikokolakis and C. Palamidessi, “Constructing elastic distinguishability metrics for location privacy,” *Proceedings on Privacy Enhancing Technologies*, vol. 2015, no. 2, pp. 156–170, 2015.
- [43] M. Cunha, R. Mendes, and J. P. Vilela, “Clustering geo-indistinguishability for privacy of continuous location traces,” in *2019 4th International Conference on Computing, Communications and Security*. IEEE, 2019, pp. 1–8.
- [44] D. Shi, J. Ding, S. M. Errapotu, H. Yue, W. Xu, X. Zhou, and M. Pan, “Deep Q-network-based route scheduling for tnc vehicles with passengers’ location differential privacy,” *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7681–7692, 2019.
- [45] Y. Qian, Y. Jiang, M. S. Hossain, L. Hu, G. Muhammad, and S. U. Amin, “Privacy-preserving based task allocation with mobile edge clouds,” *Information Sciences*, vol. 507, pp. 288–297, 2020.
- [46] C. Qiu, A. C. Squicciarini, C. Pang, N. Wang, and B. Wu, “Location privacy protection in vehicle-based spatial crowdsourcing via geo-indistinguishability,” *IEEE Transactions on Mobile Computing*, 2020.
- [47] X. Shi, D. Yu, and W.-A. Zhang, “Quantitative relationship between localization accuracy and location privacy level in wireless localization system,” *IEEE Signal Processing Letters*, vol. 27, pp. 1055–1059, 2020.
- [48] S. Takagi, Y. Cao, Y. Asano, and M. Yoshikawa, “Poster: Protecting location privacy on road networks,” in *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*, 2020, pp. 913–915.
- [49] W. Ren and S. Tang, “Egeoindis: An effective and efficient location privacy protection framework in traffic density detection,” *Vehicular Communications*, vol. 21, p. 100187, 2020.
- [50] A. Orda and R. Rom, “Shortest-path and minimum-delay algorithms in networks with time-dependent edge-length,” *Journal of the ACM (JACM)*, vol. 37, no. 3, pp. 607–625, 1990.
- [51] R. Noland and K. A. Small, “Travel-time uncertainty, departure time choice, and the cost of morning commutes,” *Transportation research record*, no. 1493, pp. 150–158, 1995.
- [52] B. Y. Chen, C. Shi, J. Zhang, W. H. Lam, Q. Li, and S. Xiang, “Most reliable path-finding algorithm for maximizing on-time arrival probability,” *Transportmetrica B: Transport Dynamics*, vol. 5, pp. 248–264, 2017.
- [53] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, “What can we learn privately?” *SIAM Journal on Computing*, vol. 40, no. 3, pp. 793–826, 2011.
- [54] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, “Local privacy and statistical minimax rates,” in *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 2013, pp. 429–438.
- [55] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, “What can we learn privately?” *SIAM Journal on Computing*, vol. 40, no. 3, pp. 793–826, 2011.
- [56] A. Blum, K. Ligett, and A. Roth, “A learning theory approach to non-interactive database privacy,” in *Proc. of STOC*. ACM, 2008, p. 609–618.
- [57] F. L. M. Marie-Ange L  bre, “Microscopic vehicular mobility trace of europarc roundabout.” [Online]. Available: <http://vehicular-mobility-trace.github.io/>
- [58] M. Piorkowski, N. S. Djukic, and M. Grossglauser, “Dataset of mobility traces of taxi cabs in san francisco, usa,” 009-02-2. [Online]. Available: <http://crawdad.org/epfl/mobility/20090224/index.html>



Fang Liu Fang Liu is Professor in Applied and Computational Mathematics and Statistics at the University of Notre Dame, Notre Dame, IN. Prof. Liu obtained her Ph.D. from the University of Michigan, Ann Arbor. Her research interests include data privacy, differential privacy, statistical machine learning, model regularization, Bayesian statistics, and analysis of missing data.



Dong Wang Dong Wang is currently working toward the Ph.D. degree of the State Key Laboratory of Engineering in Surveying, Mapping and Remote Sensing of Wuhan University, China. Now she as a visiting Ph.D. student studied at the University of Note Dame. Her research interests include data privacy and data mining.



Zhengquan Xu Zhengquan Xu is Professor of the State Key Laboratory of Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China. Prof. Xu obtained his Ph.D.degree from Tsinghua University. His research interests include multimedia security, privacy, network communications, and spatial data storage.