Differentially Private Map Matching for Mobility Trajectories

Ammar Haydari
University of California, Davis
ahaydari@ucdavis.edu

Chen-Nee Chuah
University of California, Davis
chuah@ucdavis.edu

Michael Zhang
University of California, Davis
hmzhang@ucdavis.edu

Jane Macfarlane
University of California, Berkeley
janemacfarlane@berkeley.edu

Sean Peisert
Lawrence Berkeley National Laboratory
sppeisert@lbl.gov

ABSTRACT

Human mobility trajectories provide valuable information for developing mobility applications, as they contain diverse and rich information about the users. User mobility data is valuable for various applications such as intelligent transportation systems (ITS), commercial business models, and disease-spread models. However, such spatio-temporal traces may pose a threat to user privacy. GPS trajectories in their raw form are not suitable for transportation studies, as they require matching locations with nearest road links — a process called map-matching. This paper presents a differential privacy (DP)-based map-matching algorithm, called DPMM, that generates link-level location trajectories in a privacy-preserving manner to protect users' origin destinations (OD) and travel paths. OD privacy is achieved by injecting Planar Laplace noise to the user OD GPS points. Travel-path privacy is provided with randomized travel path construction using exponential DP mechanism. The injected noise level is selected adaptively, by considering the link density of the location and the functional category of the localized links. For path privacy, our mechanism samples waypoints and selects candidate paths between waypoints. DPMM provides privacy effectively with respect to link density instead of other trajectory samples in the database compared to other privacy mechanisms. Compared to the different baseline models our DP-based privacy model offers closer query responses to the raw data in terms of individual and aggregate trajectory-level statistics with an average 36% at absolute deviation from the baseline for individual statistics on $\epsilon = 1.0$. Beyond individual trajectory statistics, the DPMM outperforms the other benchmark DP-based mechanisms on different aggregate statistics with up to 8x improvement in utility.

KEYWORDS

Differential privacy, Map matching, Data privacy, Trajectory privacy, Mobility dataset

ACM Reference Format:
Ammar Haydari, Chen-Nee Chuah, Michael Zhang, Jane Macfarlane, and Sean Peisert. 2022. Differentially Private Map Matching for Mobility Trajectories. In *Austin '22: ACM Annual Computer Security Applications Conference, December 05–09, 2022, Austin, TX*. ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The pervasive use of location tracking devices and navigation tools generate a huge amount of spatio-temporal data associated with user mobility patterns. These collected user mobilities or trajectory data can be used for variety of purposes, such as advertising, transportation analysis, and personalized recommendation. However, mining such user movement information can reveal sensitive information, hence posing a legitimate privacy threat. Recent studies show that anonymized user trajectories are vulnerable to re-identification attacks even with just a few spatio-temporal points [10].

There have been several proposed privacy mechanisms for trajectory datasets based on two main concepts: indistinguishability and uninformativeness. The former approach via k-anonymity ensures that every trajectory is similar to one another. On the other hand, uninformativeness is achieved via differential privacy, where adversaries cannot retrieve extra information after accessing the dataset [17]. While indistinguishability privacy is achieved through suppression or generalization methods [19, 41], uninformativeness privacy is, in general, achieved by perturbation and noise injection [1, 20, 24, 44]. However, the existing privacy methods result in high utility loss when trajectory queries are performed on the protected mobility data due to several reasons, such as unreasonable location sequences or geospatial mismatches.

Most techniques in the literature protect the privacy of individual user trajectories with respect to other trajectory samples in database [3, 24]. However, this approach cannot guarantee user privacy in low-density datasets. This paper attempts to protect the privacy of every individual trajectory regardless of the rest of the data by masking origin and destinations (OD) with noise injection and protecting travel paths with randomized path selection. Another limitation of existing privacy-preserving methods is the higher mismatches of geospatial location sequences. Discretization of locations through grids or zones does not consider practical implications of the "private location". We propose to incorporate the road segment densities, which intrinsically imply population densities, instead of grid or zone structures in designing our differential privacy mechanism.

Differential privacy (DP) provides statistical privacy protection by applying randomization techniques to the database and masking the personalized identifiers [13]. DP assures that an adversary with background knowledge about the dataset cannot extract private information from the dataset. The goal of this work is to design a DP-based privacy mechanism with deterministic constraints in order to have a lower bound for both location privacy and travel path

ACSAC '22, December 05–09, 2022, Austin, TX
2022. ACM ISBN 978-1-4503-XXXX-X/22/12…$15.00
https://doi.org/XXXXXXX.XXXXXXX

privacy. The proposed scheme outputs a set of privacy-preserved trajectories at the road segment level.

Injecting a fixed level of noise to all geo-spatial positioning (GPS) samples cannot guarantee the privacy of locations. We have achieved promising results applying adaptive noise injection to origin destinations conditioned on the travel intensities of the associated road segments to protect the privacy of aggregated mobility networks [23].

In this study, we propose a two-stage differential privacy method for map-matching, called DPMM, to protect the privacy of individual trajectories. First we apply adaptive noise injection to OD locations. Second we match the GPS points to the road segments privately and select randomized paths between selected road segments to generate private user trajectories. Our contributions are listed below:

- We expanded our prior work [23] to protect user OD location privacy for individual trajectories by injecting Planar Laplace noise to the user OD GPS points.
- We employ the exponential DP mechanism to randomize travel path construction to protect individual user trajectories.
- Both the injected noise level and path selection are adapted based on link density of the location and the functional category of the localized links.
- Our experimental evaluations show that our DPMM scheme can protect user location and trajectory privacy while maintaining high utility by providing accurate query responses compared to raw data.

2 BACKGROUND AND RELATED WORK

The privacy risk associated with trajectory datasets is at every level, including single location sample, whole trajectory level, and set of trajectories (community) level. There are two privacy concerns associated with user mobility data this paper addresses: location privacy and trajectory privacy. Location privacy refers to protecting individual user's true locations at any point in time. On the other hand, trajectory privacy protects the knowledge of specific path or route (a sequence of spatial-temporal samples) taken by a user [4, 9, 26]. Our goal is to apply DP to achieve both location and trajectory privacy without compromising the utility of the dataset (in terms of providing accurate response to a subset of fine-granularity queries).

Location Privacy: There are several DP-based location privacy studies in literature. One way of achieving location privacy is perturbation by injecting controlled noise to the location coordinates [1, 15]. Laplace noise [1] and circular noise methods [15] are the two well-known perturbation-based location privacy models in DP community. Another approach for location privacy is forming location grids with lower resolution depending on the density, then sampling fake locations from the private grids [35, 42]. Studies show that sampling fake locations cannot guarantee hiding the true locations due to statistical data correlations [27]. Hence, for OD location privacy, we employ adaptive noise injection methodology from [23] by considering the road segment density with the Laplace noise mechanism presented in [1]. Neither of these prior studies protect trajectory privacy.

We have previously introduced a location privacy mechanism for aggregated mobility datasets [23]. We propose selecting the magnitude of noise for ODs based on the road segment densities and the functional category of roads to form an aggregated mobility network. The noise injection is only applied to a subset of trajectory ODs if the road segment they belong to has less than a set density threshold. This work applies the idea of the adaptive noise injection approach to all trajectory ODs.

Trajectory Privacy: Privacy-preserving trajectory data publishing has been studied in literature extensively [27]. Compared to location privacy, trajectory privacy generally uses generative methods instead of location perturbation. Prefix-tree and human mobility model extraction approaches are the two main directions for trajectory privacy methods for DP. Researchers, in [6], apply DP with a prefix-tree data structure to user trajectory datasets by injecting noise to the count queries. A case study extension of this work with a real public transportation dataset is presented in [7]. More recently, several synthetic trajectory generation methods based on prefix-tree data structures with adaptive generalization approaches have been proposed [24, 45].

Another line of synthetic trajectory generation is based on modeling human movements [20, 32]. This approach extracts features from user trajectories and injects controlled noise to the mobility distributions to make them private. However, human mobility characteristics are highly complicated, and the model-based methods do not capture the real mobility dynamics all the time [33]. Recently, synthetic data generation models with machine learning, especially deep learning, are attracting attention for either lack of available data or privacy concerns [16, 37]. Deep generative models-based privacy mechanisms have been proposed in literature to extract human mobility features with non-linear learners [18].

Instead of trajectory generation, several studies target different directions for the privacy of mobility trajectories. For example, dummy location injection [30], location swapping in the mixed zone [40], location generalization [28], and trajectory reconstruction [8] are some of the proposed approaches for trajectory privacy.

Since dealing with location sequences is challenging in the continuous domain, proposed schemes are generally in the discretized grid domain. However, having a grid-like discrete representation cannot prevent geospatial mismatching. For instance, when a location is randomly sampled from a grid where the road network is sparse, mostly generated sample points to a non-sense location. This restriction practically results in higher utility loss. So instead, DPMM discretizes the locations to road segments, resulting in more realistic trajectories.

3 METHODOLOGY OVERVIEW

Protecting personally identifiable information is crucial before publishing the user mobility data. Differential privacy is a probabilistic approach that provides privacy through noise injection and/or randomized selection. We propose a method for generating differentially private mobility trajectories with map-matching, called DPMM, to protect personal identifiers. This section summarizes the notations and definitions that are required for the proposed DPMM privacy model.

3.1 Notations and Metrics

Let $D(V, E)$ represent the road network as a weighted digraph, where the set of nodes V correspond to a road intersection, the set of edges E to roads, and weights that represent link metrics, such as length of the link or traffic volume. A link $\phi \in E$ connects intersections u and v where specific link attributes, such as the number of lanes, and speed limit, are stored in the link description. We have two sets of trajectories: GPS trajectories and link trajectories. Let us define the GPS trajectories and then link trajectories:

1) GPS Trajectories: A sequence of GPS coordinates with l number of samples $\mathbf{p} \in T = \{\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_l\}$ form a GPS trajectory that reflects the continuous motion of the object. The set of all GPS trajectories are Ψ where $T \in \Psi$.

2) Link Trajectories: Given m number of vehicles $\Phi \in \Lambda = \{\Phi_1, \Phi_1, ..., \Phi_m\}$ on the road network, each vehicle travels between ODs using an ordered link path generating a user travel path known as a *micro-graph* $\Phi_i \in D$. Every link trajectory has n number of links $\phi \in \Phi = \{\phi_1, \phi_2, ..., \phi_i, ..., \phi_n\}$ and $\Phi \subset E$. The raw link trajectory is Λ and the privacy preserved link trajectory is Σ. The goal of our research is to release the privacy preserved link trajectories using the raw trajectories $\Lambda \longrightarrow \Sigma$. Every link in network $D(V, E)$ includes the road characteristics. Each link ϕ is classified into one of five classes in terms of the capacity and functional role of the road, called a functional class. Arterial roads have lower functional classes, rural streets have higher functional classes. Next, we introduce the general concept of map matching algorithms.

3) Map-Matching: GPS coordinates are an estimate of a device's location using satellite broadcast information. However, these locations do not always represent the exact travel path due to several intrinsic and environmental errors such as satellite geometry, signal blockage, tree cover, or urban canyons [2]. Consequently, GPS locations may not match a link on the road network. Map-matching generates an ordered set of road network links describing the user's trajectory considering the road network $D(V, E)$ and GPS points [36].

Map-matching algorithms play an essential role for transportation engineers as part of trajectory processing to minimize trajectory errors [5]. Since most GPS trajectories already require map-matching as a pre-processing before using them in transportation applications, DPMM eases the burden of map matching by generating privacy preserved link trajectories given raw GPS trajectories.

3.2 Differential Privacy

Location privacy and path privacy are the two main notion of privacy for trajectories in this work. We require that the output of a query statistically guarantees the privacy of individual user locations independent of the background knowledge. Differential privacy (DP) [14] guarantees that modifying the single input value has a negligible effect on the output statistical query. In this section, we summarize the general definitions and metrics of DP that are applicable to our problem.

We introduce the privacy concerning data $\mathbf{X} \in \mathcal{X}$ as vehicular mobility information in query $q \in \mathcal{Q}$. The data holder wants a mechanism that hides the sensitive information and reports the privacy preserved version of sensitive information using a randomized algorithm $\mathcal{A} : \mathbf{X} \times \mathcal{Q} \longrightarrow \mathcal{D}$ where $\mathcal{Q}$ is the query space and $\mathcal{D}$ is the output space. DP promises that the algorithm $\mathcal{A}$ is differentially private such that participation or removal of a record results in minimal changes to the output of a query.

Let us first define the neighboring datasets:

Definition 1. *[Neighboring Dataset] Considering two databases* $\mathbf{X}$ *and* $\mathbf{X}'$*, if they differ by only one element* $\mathbf{x_i} \longrightarrow \mathbf{x_i}'$ *corresponding to a link trajectory, they are neighboring datasets.*

The above definition formalizes the adjacent or neighboring dataset that plays a crucial role in differential privacy.

Definition 2. *[ϵ-Differential Privacy] Given for every neighboring sets $d \subset \mathcal{D}$, a randomized algorithm $\mathcal{A}$ is ϵ-differentially private if*

$$Pr(\mathcal{A}(X) \in d) \leq e^{\epsilon} Pr(\mathcal{A}(X') \in d) \tag{1}$$

where ϵ is a positive real number and probability comes from the randomness of the algorithm. $\frac{Pr(\mathcal{A}(X) \in d)}{Pr(\mathcal{A}(X') \in d)}$ *is the privacy leakage risk for the randomized algorithm $\mathcal{A}$.*

ϵ-differential privacy is known as randomized response where adding or removing a single element from the database results in a similar probability. The smaller value of ϵ represents higher privacy guarantee and provides in-distinguishability.

An appropriate epsilon, in DP, is typically determined based on the sensitivity of the underlying data. The definition of sensitivity is given in [13] as follows:

Definition 3. *[Sensitivity] For any query function $f : D \longrightarrow R^n$ that maps the dataset D to fixed sized real numbers, the sensitivity of f is defined as*

$$\Delta f = \max_{\mathbf{X}, \mathbf{X}'} \left\| f(\mathbf{X}) - f(\mathbf{X}') \right\|_1 \tag{2}$$

for all neighboring datasets $\mathbf{X}$ *and* $\mathbf{X}'$.

Definition 4. *[Composition] Let a set of randomized algorithms $\mathcal{A}_1, ..., \mathcal{A}_k$ that each $\mathcal{A}_i$ satisfies ϵ_i-DP.*

- *Sequential Composition: Let $\mathcal{A}$ be another randomized mechanism that executes $\mathcal{A}_1, ..., \mathcal{A}_k$ with independent randomness for each $\mathcal{A}_i$, then $\mathcal{A}$ satisfies $(\sum_i \epsilon_i)$-DP.*
- *Parallel Composition: Let dataset* $\mathbf{X}$ *is partitioned deptermin-istically to different subsets* $\mathbf{X}_1, ..., \mathbf{X}_k$ *and executing each $\mathcal{A}_i$ with a different disjoint set* $\mathbf{X}_i$ *satisfies* $\max_i (\epsilon_i)$*-DP.*
- *Post-processing a randomized algorithm $\mathcal{A}$ that satisfies ϵ-DP does not break or consume any privacy budget.*

Given the composition properties and total ϵ privacy budget, the proposed DPMM builds different blocks carefully according to composition properties to achieve a DP satisfied randomized algorithm $\mathcal{A}$.

DP guarantees privacy for both numerical and non-numerical queries. While noise injection is a leading method for numerical queries, exponential mechanism is a mainly used mechanism for non-numerical queries [13, 31].

Input perturbation and output perturbation are the two ways to implement DP. When we want to achieve OD location privacy on trajectories, one way to do is through input perturbation, where noise is injected into the GPS points. Using noise function $L(\epsilon, R)$, the GPS points are perturbed based on the below definition.

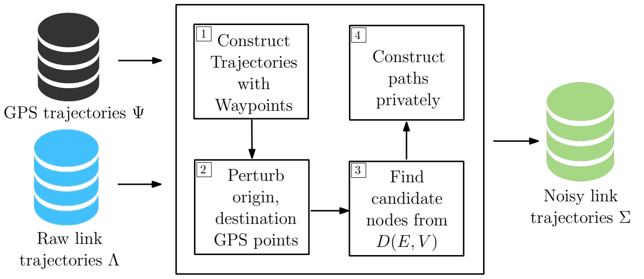

GPS trajectories Ψ
Raw link trajectories Λ
1 Construct Trajectories with Waypoints
2 Perturb origin, destination GPS points
3 Find candidate nodes from $D(E,V)$
4 Construct paths privately
Noisy link trajectories Σ

Figure 1: Differentially private link trajectory generation scheme.

Definition 5. *[Laplace Mechanism] For any function $f : D \longrightarrow R^n$, the mechanism $\mathcal{A}$ gives ϵ-DP as follows:*

$$\mathcal{A}(D) = f(D) + Laplace(\epsilon, R) \tag{3}$$

Noise injection to the input can be done with different noise functions depending on the application requirements. Section 4.2 describes the additive noise method in detail. By injecting noise into the input GPS points, the method guarantees that the OD locations of trajectories are differentially private.

For privacy on non-numerical queries, exponential mechanism selects an output from input domain taking into consideration of a score function $q(\mathbf{X}, r)$ where r is the discrete output from the domain. Exponential mechanism assigns higher probabilities for the higher score to incentivize the higher utility outcomes.

Definition 6. *[Exponential Mechanism] Let $q : (\mathbf{X}, \mathbf{R}) \longrightarrow R$ be a score function for a database $\mathbf{X}$ and domain specific discrete outputs R, the algorithm $\mathcal{A}$,*

$$\mathcal{A}(D, q) = \left\{ return\ r \in R\ with\ probability \propto \exp \frac{\epsilon q(D, r)}{2\Delta q} \right\} \tag{4}$$

satisfies ϵ-DP.

4 PRIVATE MAP MATCHING

This section describes the components of the proposed DP-based map-matching algorithm for trajectory privacy that generates synthetic link-level user trajectories. DPMM guarantees statistical privacy protection for link trajectories with noisy ODs and randomized travel paths. Figure 1 shows the flowchart of our DPMM mechanism. We transform the point-wise GPS trajectories into an ordered series of road network links and enforce privacy on trajectories with the road segment.

4.1 Trajectories with Waypoints

Trajectories are time-ordered sequential location samples, and the sampling rate varies depending on the device. Before private path construction, we represent the trajectories with fewer GPS waypoints that retain the movement characteristics. This waypoint approach only preserves the critical locations enough for movement representation by removing insignificant locations. For example, in a higher sampling rate trajectory, sequential path construction may result in redundant extra paths due to frequent path findings (see Section 4.4 for more details). Furthermore, the frequent path selection also consumes more privacy budget ϵ. In summary, the waypoint representation enhances the path quality and decreases the computational complexity by dealing with fewer location pairs.

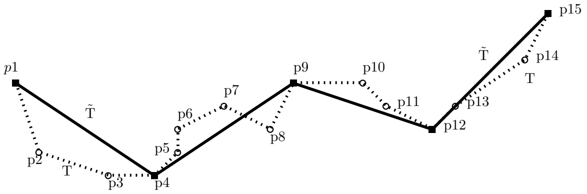


Figure 2: Trajectory Simplification

For a trajectory T, let n coordinates be $p_1, p_2, ..., p_n$ where every p_i is represented with (x_i, y_i) and $n-1$ line segments be $\overline{p_1p_2}, ..., \overline{p_{n-1}p_n}$. Figure 2 shows a toy trajectory simplification example from the original trajectory T to simplified trajectory $\tilde{T}$. Original trajectory has 15 coordinate points $p_1, p_2, ..., p_{15}$. Using trajectory simplification, we can represent trajectory T with waypoints $p_1, p_4, p_9, p_{12}, p_{15}$, which allow us to find approximate paths between distant points. The first step of the proposed DP mechanism is to represent trajectory with fewer waypoints.

In literature, there are several algorithms to decimate curves that are composed of line segments as we have in trajectories. We consider non-parametric Ramer–Douglas–Peucker (RDP) algorithm for representing higher sampling rate trajectories with sample waypoints [12]. RDP is a heuristic method that we attached to the DPMM to retain important GPS waypoints in the randomization process and help generate more practical travel paths.

RDP recursively approximates the whole trajectory to fewer points representation starting from $\overline{p_1p_n}$ line segment and an error bound σ, which also known as simplification error. RDP then calculates distance offset of each point coordinate from $p2$ to p_{n-1} with perpendicular distance. Let p_k be the point with maximum of perpendicular distances from $\overline{p_1p_n}$. If $\sigma_k > \sigma$, RDP splits the line segment into two sub-segments $\overline{p_1p_k}$ and $\overline{p_kp_n}$ where σ_k is the offset distance from p_k to $\overline{p_1p_n}$. The simplification continues recursively for $\overline{p_1p_k}$ and $\overline{p_kp_n}$. The RDP terminates if $\sigma_k \leq \sigma$ or $\overline{p_ip_j}$ is a consecutive segment with $j - i = 1$. It worth mentioning that RDP only removes the unnecessary middle points of trajectories by keeping the OD points in $\tilde{T}$. The time complexity of RDP is $\mathcal{O}(n^2)$.

4.2 Private Origin-Destinations

Traveling from one geographical location called the origin to another geographical location called the destination is sensitive information that must be protected. The map-matching algorithm infers the ordered set of road segments (links) using GPS locations from the $D(V, E)$ network and finds paths for each pair of user's GPS points.

We recently propose an adaptive noise injection model for location privacy on aggregated mobility networks in [23]. DPMM employs the previous noise injection methodology for trajectory privacy on ODs. This method injects adaptive Planar Laplace noise to the GPS points before matching them with an appropriate link to provide OD privacy on the map-matching algorithm.

The OD GPS points are obfuscated based on the network density with noise injection and they are matched with a new link. The two key parameters used for noise sampling are ϵ and R. While ϵ

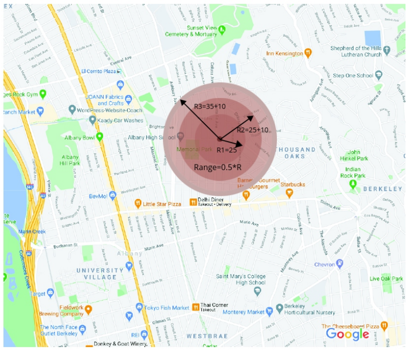


Figure 3: Buffer range for determining link density

is responsible for the noise level, R is the distance parameter for moving the center of the noise in the geospatial domain. The output of the noise function is a new randomized GPS location in the same space.

Geo-indistinguishability is one of the noise injection methods for hiding GPS locations [1]. The Laplace noise is sampled from a bounded probability density function on polar coordinate systems instead of Cartesian space as follows:

$$D_\epsilon(r,\theta) = \frac{\epsilon^2}{2\pi} r e^{-\epsilon r} \tag{5}$$

where r is the distance of $\mathbf{x}$ from $\mathbf{x_0}$, θ is the angle and $\frac{\epsilon^2}{2\pi}$ is a normalization factor. While θ is random uniformly drawn from $[0, 2\pi)$, ϵ is direct user input and r is scaled with given radius R from the input. We refer readers to the original study for the details of noise sampling [1].

The experimented geo-indistinguishability provides location privacy by adding controlled noise $L(\epsilon, R)$ to the OD GPS points x_i within a certain range R in order to mask the actual locations using density based noise range selection method R.

ALGORITHM 1: Adaptive noise magnitude selection

```
Input $h_1$ for the number of links in the buffer range ;
Input $h_2$ for the number of links in the buffer range that
  belong to the same functional class ;
Input $Z$ initial buffer range ;
$LinkSet$ ← empty set ;
$LinkSetFC$ ← empty set ;
while Size of $LinkSet \leq h_1$ or Size of $LinkSetFC \leq h_2$ do
    Find the $LinkSet$ links within buffer zone ;
    Find the $LinkSetFC$ links within buffer zone ;
    Z+=10 meters ;
end
$R \leftarrow \frac{1}{2}Z$ ;
return $R$ and $LinkSetFC$
```

Density-Aware Noise Injection: This section explains the density-aware structure of the noise injection approach using Planar Laplace noise proposed in [1]. Randomly injecting noise without considering the network's density would not achieve the desired privacy level all the time. The DPMM provides location privacy for trajectories even for outliers by selecting noise level adaptively with respect to the link density of network $D(V, E)$

The ODs of trajectories are the most vulnerable parts due to revealing users' start and end locations, such as home or office addresses. Therefore, providing privacy for ODs requires much attention. In this work, we consider the link density around the OD of trajectories to define the level of noise that needs to be injected. As we mentioned earlier, every link has functional class information, and DPMM moves the GPS point to a place that matches a new link with the same functional class of the original link.

Link density in the road network quantifies the populations in general. While central areas have more streets and intersections, which implies more population, the rural places have fewer road segment connections due to limited populations. Therefore, it is easier to provide privacy for the people who live in central areas. On the other hand, it is hard for the rural areas since location traces are unique in the outskirts of the communities. We define a density function for noise injection as follows:

Definition 7. *[Density Function] Given the ϵ value, radius R of the noise function $L(\epsilon, R)$ is selected with respect to $R = f(\theta)$ where θ is the network density in terms of the number of links (road segments).*

For each trajectory, OD GPS points are perturbed with the noise injection model. The DPMM adjusts the noise using the link density around the GPS point with respect to a cloaking region (see Fig. 3). To do so, starting from an initial radius Z, the proposed mechanism increases the radius Z until finding a certain number of links and the same functional class links. The number of all links, h_1, and the number of the same functional class links, h_2, are user-defined parameters based on the geographical region and density of the network $D(V, E)$. Sparse vs dense structured regions or shapely vs end-to-end intersection-based network would require different hyperparameters. For example, this project considers shapely road network, which divides the end-to-end intersection road link to the small links based on the road curves and doing map-matching with a different road network requires different h_1 and h_2 hyper-parameters. Once the number of all links and the same functional class links reach the thresholds h_1 and h_2, the center of the final radius Z, which satisfies the two thresholds, is selected as the input for noise function $L(\epsilon, R)$ where $R = \frac{1}{2}Z$. As the Laplace noise is 2-dimensional, we select the half of distance value Z for this noise model and sample a GPS point with given parameters (see Algorithm 1). After the noisy GPS point is returned from $L(\epsilon, R)$, all the nodes belonging to the same function class links are selected as candidate nodes for path construction.

4.3 Candidate Nodes

To construct the path of a trajectory using the network $D(V, E)$, map-matching first needs to have candidate nodes for each GPS point. However, due to geospatial constraints, selecting a single candidate node given the GPS point does not guarantee to match with the correct node. For instance, if the GPS is close to a one-way road and a two-way street with a similar distance, the GPS point may belong to both. Selecting the best node depends on the direction and the next GPS point. To mitigate the geospatial constraints, we propose to choose a set of candidate nodes to find paths. Besides, selecting a travel path randomly using multiple candidate paths increases the privacy (see Section 3.2).

While we select candidate nodes for OD GPS points from the same functional class links using threshold h_2, for waypoints, we find candidate nodes from all the links using the threshold h_1. For every waypoint, we follow the same cloaking-region approach we followed for noise injection to find candidate nodes (see Fig. 3). However, we do not restrict candidate links to have the same functional class criteria for waypoints to increase the randomization in the path construction. The cloaking region method takes the following inputs for each waypoint from $\tilde{T}$: threshold h_1 for searching the number of links, initial radius Z, and road network $D(V, E)$. The output is h_1 number of links, and the nodes that belongs to those links are collected as a candidate node-set. Candidate nodes for each waypoint are stored in separate containers. For our experiments, we prefer to use the same threshold h_1 for ODs and waypoints. Still, the parameters can be adjusted depending on the geographical region and network $D(V, E)$ structure. Algorithm 2 summarises the candidate node selection.

ALGORITHM 2: Candidate node selection

```
Input $h_1$ for the number of links in the buffer range;
Input $Z$ initial buffer range;
$LinkSet \leftarrow$ empty set;
while $Size\ of\ LinkSet \leq h_1$ do
    Find the $LinkSet$ links within buffer zone Z;
    Z+=10 meters;
end
return Candidate Nodes from LinkSet
```

4.4 Private Paths

A user's travel path could allow an adversary to infer further information about the user's identity by linking other available information to the user path. For instance, the adversary may know several locations of a user, such as home or office location and specific automatic toll booths that the user passed through. If a path matches with known locations, the adversary may identify the user. Since locations are sensitive and easy to link with user identities, minimizing the actual travel paths of a user reduces the risk of re-identification by adversaries. Instead of revealing the true path of a trajectory, randomizing the paths in the same trajectory direction using waypoints limits the true travel path while providing similar travel within the same vicinity.

DPMM selects the travel routes for the sequence of waypoints randomly to construct privacy-preserving paths. First, for a simplified trajectory $\tilde{T}$, the proposed method finds candidate paths between waypoints using the candidate node-set, which is constructed using Algorithm 2. Then, since we do not intend to find the shortest path, we implement A^* path finding algorithm with the euclidean distance between nodes as heuristics [21]. A^* algorithm combines the Dijkstra shortest path algorithm with greedy search methods [11] and finds reasonable paths by using heuristics to guide the path finding direction.

For every pair of points p_i and p_j in $\tilde{T}$, candidate paths are stored with the corresponding travel distance. The proposed DPMM selects a travel path randomly with probability proportional to the travel distance. The shorter travel distance has a higher chance of being traveled by the user. Therefore, the selection mechanism assigns a higher probability to the shorter travel distance path. To achieve this, we inversely normalized the distances between 0 and 1. Then, we select the path privately using DP exponential mechanism. Note that our model's sensitivity Δp is 1 because maximum travel distance is always bounded to 1 due to normalization. DPMM follows this procedure sequentially, and the final private link trajectory protects the user travel paths along with OD privacy.

4.5 Travel Path Adjustment

The network used for map-matching is a directed graph. Depending on the road traffic direction, network $D(V, E)$ has separate links for incoming and outgoing links. Randomized path selection sometimes may result in unreasonable travel paths that go reverse and make a u-turn or o-turn reaching the same node visited before. We remove the travel loops after private path selection to prevent redundant paths taken by the private map-matching. Our experimental analysis shows that the loops on raw trajectories are less than 1% in our dataset; removing the loops after private path selection decreases utility loss. Note that the ϵ-DP privacy guarantee still holds with post-processing.

4.6 Complete Trajectory Construction

The proposed private map-matching algorithm combines noise injection and private selection DP methods, as we discussed in separate sections above. Algorithm 3 summarizes the privacy protection mechanism. First, the algorithm creates a waypoints trajectory $\tilde{T}$ by keeping the OD as it is. Next, it injects the proposed adaptive Laplace noise to the OD GPS points and forms candidate nodes from the same functional class links. The third step of the proposed algorithm finds candidate nodes for every waypoint in $\tilde{T}$. In the fourth step, the proposed mechanism finds candidate paths between every consecutive node-set using A^* routing algorithm. Then, it selects paths privately from the candidate paths using the exponential-DP method. Finally, it connects selected candidate paths and removes the travel loops. The algorithm terminates after generating all the private link trajectories from GPS trajectories.

5 SYSTEM ANALYSIS

5.1 Differential Privacy Analysis

The DPMM distributes the privacy budget ϵ evenly to the sub-processes while guaranteeing ϵ-DP. Representing the raw GPS trajectory with $s + 1$ waypoints including ODs results in s paths that needs to be private. Total ϵ budget divided to ϵ_i for OD noise injection and number of waypoints such that $\sum_{i=1}^{2+s} \epsilon_i$. While OD noise injection provides privacy with the property of parallel composition, private path construction provides privacy with sequential composition. Post-processing on map-matched trajectories, such as removing the travel loops, does not violate the ϵ-DP privacy.

The smaller value of ϵ represents higher privacy and indistinguishability, whereas higher ϵ gives more accuracy to the output trajectory. Due to the geospatial and temporal nature of user movements, it is also essential to preserve the accuracy of the generated trajectories while achieving a reasonable privacy guarantee. The

ALGORITHM 3: Privacy Preserving Map-Matching

```
Input $\Lambda, \Psi, D(V, E)$, $h_1$ for number of links in the buffer
  range, $h_2$ for number of same functional class links in the
  buffer range, $Z$ initial buffer range ;
for $T \in \Psi$ do
    Build waypoints trajectory $\tilde{T}$ from $T$ using RDP ;
    for $p \in \tilde{T}$ do
        if $p$ is Origin or Destination then
            Select $R$ and $BufferSetFC$ using Algorithm 1 ;
            Inject adaptive noise to the GPS point $p$ using
              $L(\epsilon, R)$ ;
            Form candidate nodes set from $BufferSetFC$
              links ;
        end
        else
            Form candidate nodes from Algorithm 2 ;
        end
    end
    Find candidate paths with $A^*$ for candidate nodes ;
    Select private paths with exponential-DP mechanism ;
    Connect privately selected paths ;
    Remove the node loops as in Section 4.5 ;
    Build the noisy link matched trajectory with connected
      links ;
end
return Noisy link trajectories $\Sigma$
```

data owner can adjust the privacy budget with respect to the sensitivity for both OD and path privacy. If the data owner wants to hide the ODs more, he/she can select a smaller ϵ value for noise injection to the ODs, which increases the perturbation. The same analogy can be applied to path privacy too. In summary, we left privacy budged distribution to the data owner, and this aspect is out of the scope of this work.

5.2 Attack Resilience

Outlier Leakage. A trajectory may have OD points that are unique in a sense and reveal vulnerable information about user identity [20]. Threat on outlier trajectories mainly applies to rural areas, such as travel between a hospital and a farmhouse. Injecting the same noise magnitude to all GPS points cannot provide privacy for every GPS point. Moving GPS points slightly can provide privacy in central locations. However, repositioning locations in an outlier area at the same level as in central areas may not offer the same privacy. The proposed privacy mechanism deals with outlier trajectory ODs by perturbing them adaptively with respect to road segment density.

Partial Sniffing. An adversary may have access to a sub-trajectory of a user that participated in the trajectory dataset through physical tracking or social networking. Then, an adversary may try to infer the rest of the user travel that passes through the locations in the sub-trajectory. Let a user's sub-trajectory T_{sub} be known by the adversary; there is a high chance to reveal the user's rest of the travel if the adversary can find a matching T from trajectory database [20]. DPMM prevents adversaries from making such inferences with two concepts: OD privacy and path privacy. For example, a true trajectory may travel from a local street to a hospital. When an adversary gets access to a partial trajectory T_{sub} of user trajectory T, he/she may try to infer the home address and the purpose of the travel. However, since the proposed privacy mechanism does not disclose the true ODs and travel path, the adversary cannot correctly identify the user information from the privacy preserved trajectory $T_p \in \Sigma$.

6 EVALUATION

6.1 Dataset Description

This project uses a real-world dataset collected in the San Francisco Bay area in California with fleet and consumer GPS trajectories. We process one day one hour of trajectories (between 1 pm and 2 pm) from the city of San Francisco. In total, the experiments apply DPMM to 833 user trajectories. The dataset is created from various location-sharing applications and GPS tracking devices. When the tracking device is active, location (lat, lon), speed, and heading are collected along with a unique device identifier. Trajectories have varying sampling rates due to being collected from different sources. However, most of the trajectories have sampling rates of less than 1-minute.

6.2 Comparisons to Alternate Approaches

We compared the proposed privacy mechanism with two well-known DP-based private trajectory generators: AdaTrace [20], and DPT [24]. While AdaTrace generates synthetic trajectories by learning the mobility patterns, DPT constructs prefix-tree to generate private user trajectories. We acquired the original implementations from respective authors. Both AdaTrace and DPT models generate synthetic GPS trajectories instead of link trajectories. For a fair comparison between the proposed DPMM and benchmark models, we applied map matching to AdaTrace's and DPT's GPS trajectories to generate equivalent link-level trajectories for our analysis. This is referred to as DP-free version of our map-matching algorithm.

The utility is closely related to the size of the database for benchmark AdaTrace [20] and DPT [24] models. However, the utility for DPMM is bounded by the density of the road network. Therefore, to achieve better utility for the benchmark models, we trained their respective implementations with a whole day of trajectories within the region. The total number of GPS trajectories for one day in San Francisco city in our database is 9249.

Along with other studies in the literature, we also compare DPMM method with different variants:

- **DPMM-No-WP:** This version performs the same privacy mechanism for OD while selecting paths from trajectory T without waypoint sampling and not experimenting trajectory post-processing (removing the loops).
- **DPMM-A*-WP:** Based on DPMM-No-WP, DPMM-A*-Way adds waypoint sampling to base model in order to provide more privacy and less utility loss. This version does not perform trajectory post-processing (removing the loops).
- **DPMM-D-WP:** Following the DPMM-A*-Way, this method uses Dijkstra path finding algorithm instead of A^*. Dijkstra guarantees to find the shortest path while A^* does not.

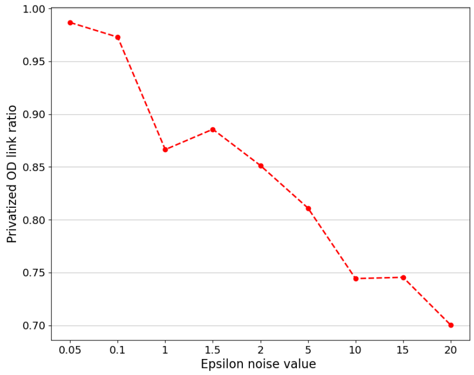


Figure 4: Comparison of different ϵ values and the change of OD-links for different for 1 hour period of trajectories between 1pm and 2pm.

6.3 Utility Metrics

We have chosen utility metrics commonly used in transportation studies including individual trajectory level and aggregated level queries. In this section, we explain the importance of the utility metrics and present some use-case examples as we define the metrics. The goal is to have a higher similarity in the utility metrics between original and privacy-preserved trajectories given the same level of privacy protection.

Individual Utility Metrics: Mobility trajectories are complicated, and evaluating the quality of privacy-preserved trajectories with aggregated statistics alone is not sufficient. For example, the OD Similarity metric for AdaTrace when compared to original trajectories (Table 1) indicates high level of OD similarity between the two. However, their respective actual trajectories show distinct differences (as shown in Figure 8a). Since the proposed DPMM perturbs only the OD GPS points, its distortion on the trajectory and geographical mismatch is limited.

We evaluate the utility of the proposed DPMM model at the individual trajectory level with different variants of the DPMM mechanisms. The relative trip length change of the link trajectories before and after applying DPMM is proposed as a utility metric in this study. Without DPMM, the base map-matching algorithm matches the GPS points with the nearest links and connects such links with the shortest path algorithms. Using the same relative change formulation, we compared the change of the privacy preserved trajectories with clean link trajectories and GPS trajectories. The trip length of the GPS trajectories is calculated using the euclidean distance between the sequence of the GPS points.

Aggregated Utility Metrics: Spacial density analysis plays a key role in understanding human mobility [22]. Our first aggregated utility metric, mainly used for graph data, is the query error that quantifies the error in the characteristics of most visited places. Minimizing the query error makes output privatized data more useful [6, 20, 43]. For this metric, 500 road links are sampled uniformly across all regions from the network $D(V, E)$. Then, the normalized absolute difference between the number of real and synthetic trajectories passing through each link is computed by the following:

$$\text{error}(Q(\Sigma)) = \frac{|Q(\Psi) - Q(\Sigma)|}{\max\{Q(\Psi), s\}}, \tag{6}$$

where $Q(\Psi)$ and $Q(\Sigma)$ are the number of trajectories that pass the certain links for the set of original trajectories vs privacy preserved trajectories, respectively, and s is sanity bound for mitigating the effect of the extremely small selective queries. We specified the sanity bound s as 1% of the users.

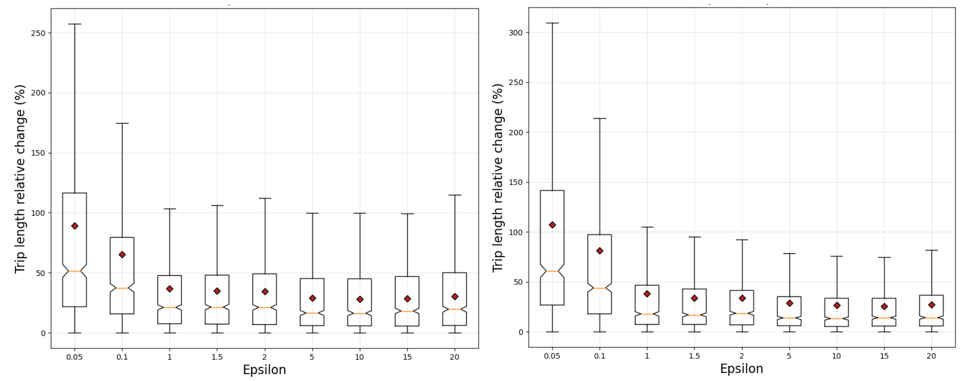
The travel characteristics of moving objects, such as personal vehicles and public transportation for spatio-temporal analysis can provide valuable insights for transportation analysts [34, 39]. The second aggregated utility metrics measures the similarity of the OD distributions, called OD Similarity. This metric evaluates how much the overall characteristics are preserved in terms of OD links. Jensen-Shannon divergence (JSD) is a well-known similarity metric mainly used for measuring the similarity of two probability distributions [29]. We employ JSD for OD similarity.

The third metric measures the changes of the Vehicle Miles Traveled (VMT), which can be useful for different purposes, such as ride-sharing [25] and land use [38], for link trajectories, called VMT Change. Link count refers to the number of times a link occurs on the aggregated link trajectory network. The last utility metric compares link count distribution between original and privacy preserved link trajectories.

6.4 Numerical Results

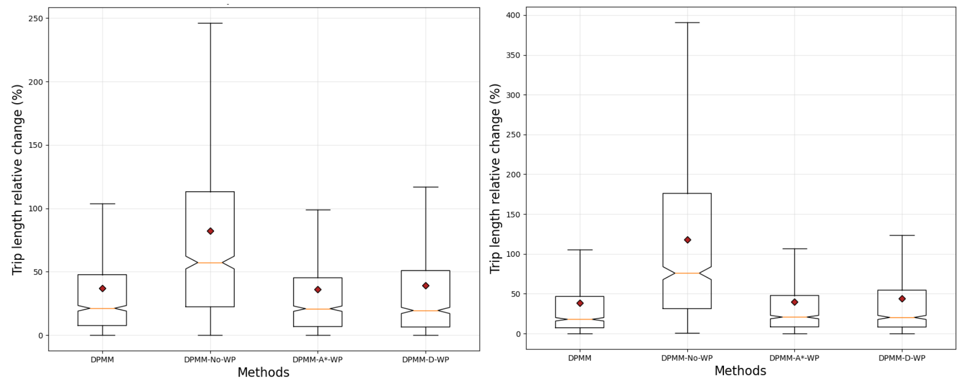
We evaluate the performance of DPMM with benchmark studies and other DPMM variants from two different aspects: change in the privacy preserved trajectories at the individual level and aggregated level. When we apply the DPMM method to the trajectory database, depending on the privacy level ϵ, the utility varies in terms of the privatized OD link ratio and trip lengths. In addition, the experiments quantify the query similarity metrics at an aggregated level with respect to other trajectory privacy methods and compare the results with other studies in the literature. We use a range of ϵ values between 0.05 and 20 to evaluate the performance of the DPMM. The ϵ values are selected to reflect the lower and upper limits of the impact of the DPMM privacy mechanism.

Individual Trajectory-level Analysis: Regardless of the other user's movements, every OD link may have privacy concerns, and matching an OD with a different link hides the true end location of the user. We inspect the fraction of OD links that are different from the original raw trajectory after the proposed noise injection, which we refer to as the *privatized link ratio*. Depending on the noise level and the road network density $D(V, E)$, DPMM may still match the links with the same link after the noise injection. To quantify the privacy of our method, we inspect the privatized link ratio over the total number of OD links with respect to different ϵ values. For 833 trajectories, we have 1666 OD links. Fig. 4 shows the performance of DPMM in terms of OD privacy. The goal of the proposed mechanism is to move OD links to different links. Therefore the output is expected to have higher ratios for the lower level of ϵ. The highest level of the privatized link ratio is observed with the lowest $\epsilon = 0.05$ with an average of 98.7%.



(a) DPMM vs original link trajectories **(b) DPMM vs GPS trajectories**

Figure 5: Performance of DPMM is compared with the different ϵ values with respect to original link and GPS trajectories.



(a) DPMM vs original link trajectories **(b) DPMM vs GPS trajectories**

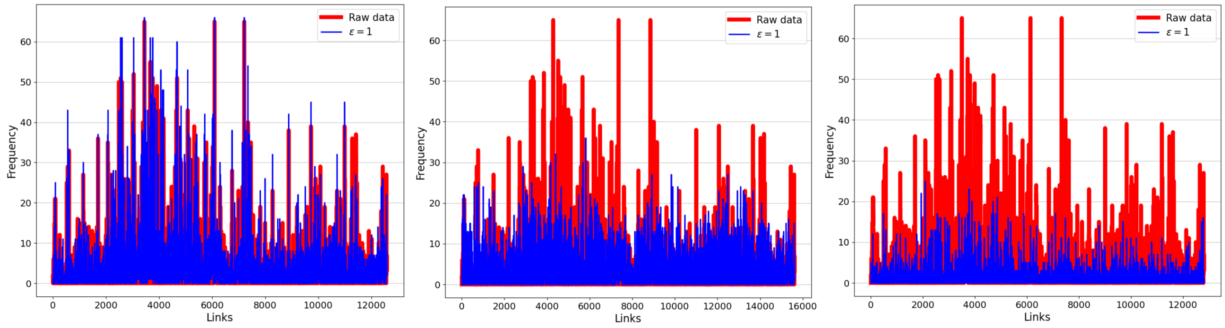
Figure 6: Performance of DPMM is compared with different methods with respect to original link and GPS trajectories on $\epsilon = 1.0$.

Next, Figure 5 quantifies the absolute trip length change with different ϵ values at the trajectory level. Figure 5a and Figure 5b illustrate the comparison of DPMM link trajectories with original link trajectories and GPS trajectories, respectively. The goal is to retain higher utility with a lower ϵ value. The lowest value of $\epsilon = 0.05$ generates the highest dissimilarity between privacy-preserved DPMM link trajectories and original link and GPS trajectories. The distortion in DPMM link trajectories is sensitive to the geographical region, road link density, and the link functional class. The average trip length change varies between 89% and 30% for original link trajectories and between 107% and 27% for GPS trajectories on different ϵ values. For instance, for $\epsilon = 1$, the absolute average distortion on trip lengths is 36.8% and 38.1% for original link trajectories and GPS trajectories, respectively. Since the sequence of GPS trajectories do not reflect the actual trip length, having a higher trip length error regarding link trajectory is expected. Increasing the ϵ noise value decreases distortion and the level of privacy that DPMM guarantees.

The last individual-level utility metric compares the proposed DPMM with other variants in terms of the change in the trajectory trip length (see Figure 6). The ϵ is selected as 1.0 for this part of the experiments. Figure 6a and Figure 6b illustrates the change in the trip length of different DPMM variants where DPMM outperforms the others with the lowest mean difference. The results clearly show the impact of the trajectory sampling from T to $\tilde{T}$ and removing the loop in the trips regarding the utility with the same privacy level. Another interesting observation is that the Dijkstra algorithm has very close similarity with A^* routing algorithm. While A^* does not guarantee the shortest path, the Dijkstra algorithm guarantees to find the shortest path. Since the users do not take the shortest path all the time, selecting a candidate path using A^* routing algorithm is a more reasonable choice due to the unpredictability of user behaviors.

Table 1: Comparison of the aggregated utility metrics with benchmark studies for $\epsilon = 1$. The lower value is the better for Querry Error and OD Similarity metrics. For VMT Error, value closer to zero is better. The bold and green results show the best performance and the second best performance, respectively.

	DPMM	AdaTrace	DPT
Query Error	**0.146**	0.353	0.264
OD Similarity	**0.065**	0.081	0.068
VMT Change	**−0.072**	0.164	−0.641



(a) DPMM

(b) AdaTrace

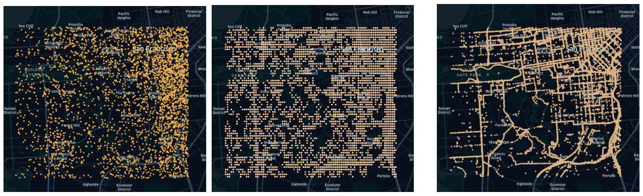
(c) DPT

Figure 7: Link count densities on aggregated network level for DPMM with baseline comparisons using $\epsilon = 1$

Aggregate-level Analysis: Table 1 presents the aggregated utility results of different metrics for $\epsilon = 1.0$. The proposed DPMM performs better for all metrics due to its ability to handle each trajectory separately. On the other hand, AdaTrace and DPT achieve varying performance on different metrics. While DPT outperforms the AdaTrace in terms of the most visited places (*Query Error*) and origin-destination densities (*OD Similarity*), AdaTrace produces more similar trajectories to proposed DPMM in terms of the trip length, as shown with the *VMT Change* statistics. In summary, the DPMM succeeds in keeping the trajectory patterns in the same region while hiding true OD locations and travel paths. Therefore, the results in Table 1 reflect the superiority of the proposed algorithm.

Next, we evaluate the proposed DPMM and benchmarks with original link trajectories in terms of link count distribution to understand how link counts changes as a function of privacy. Figure 7 shows the link count distribution with respect to the original trajectories for different privacy mechanism with $\epsilon = 1$, the ideal privacy level. The results illustrate that DPMM preserves the link densities compared to baseline models AdaTrace and DPT.

Finally, we compare the spatial densities of the benchmark models with the original trajectories using the same number of samples. Figure 8 shows the visual representation of spatial densities for the raw GPS points, AdaTrace [20] and DPT [24]. The population densities and major routes are clearly observed in raw GPS distributions. However, AdaTrace and DPT has some sort of density awareness while missing the major routes. Note that since AdaTrace and DPT do not consider geospatial constraints, resulting trajectories are sampled in traffic-free areas such as city-parks and national-preserve areas. Compared to these baselines, the proposed DPMM model provides privacy-protected trajectories at the road network level that prevents to have such unrealistic trajectories.



(a) AdaTrace

(b) DPT

(c) Original Trajectories

Figure 8: Visual representation of the original trajectories vs privacy preserved trajectory densities for benchmark models. Proposed DPMM does not produce GPS trajectories, hence, it does not have visual comparison with benchmarks.

7 CONCLUSION

In this paper, we present a differentially-private map-matching algorithm for the privacy of mobility trajectories. Proposed mechanism protects individual OD locations with adaptive noise injection model and travel paths with exponential DP method. The DPMM injects planar Laplace noise to the individual OD GPS points by considering the density of the localized road network and the functional class of the links. The actual perturbation level for each GPS point is adjusted by considering the localized link density. Next, proposed DPMM uses a waypoint sampling method for constructing travel paths privately. We evaluate our DPMM method for a variety of noise levels by comparing it with several comparative privacy models at individual trajectory and aggregated statistics.

The advantage over the literature of DPMM does not rely on population density with respect to other samples in the database, rather it considers link density in the road network. Due to map-matching, DPMM prevent geographical mismatches with the road structures which is a common problem for other baseline models. While this project provides OD location privacy with travel path privacy for individual user trajectories, DPMM does not guarantee the generation of the repeated trajectories due to the randomized nature of the mechanism. This resulting distortion is a form of the utility trade-off. Future work will include extending this investigation to different types of mobility datasets while also addressing the aforementioned limitations.

ACKNOWLEDGMENTS

This work was supported by Contractor Supported Research (CSR) funds provided by Lawrence Berkeley National Laboratory, operated for the U.S. Department of Energy (DOE) under Contract No. DE-AC02-05CH11231, NSF HDR:TRIPODS grant CCF-1934568, the U.S. DOE Vehicle Technologies Office (VTO) under the Big Data Solutions for Mobility Program, an initiative of the Energy Efficient Mobility Systems (EEMS) Program, and Department of Transportation - University Transportation Centers, "Center for Transportation, Environment, and Community Health (CTECH)". Prasad Gupte, of the DOE Office of Energy Efficiency and Renewable Energy (EERE), played an important role in establishing the project concept, advancing implementation, and providing ongoing guidance.

REFERENCES

[1] Miguel E Andrés, Nicolás E Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2013. Geo-indistinguishability: Differential Privacy for Location-based Systems. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*. 901–914.

[2] Mustafa Berber, Aydin Ustun, and Mevlut Yetkin. 2012. Comparison of accuracy of GPS techniques. *Measurement* 45, 7 (2012), 1742–1746.

[3] Vincent Bindschaedler and Reza Shokri. 2016. Synthesizing plausible privacy-preserving location traces. In *2016 IEEE Symposium on Security and Privacy (SP)*. IEEE, 546–563.

[4] Shan Chang, Chao Li, Hongzi Zhu, Ting Lu, and Qiang Li. 2018. Revealing privacy vulnerabilities of anonymous trajectories. *IEEE Transactions on Vehicular Technology* 67, 12 (2018), 12061–12071.

[5] Pingfu Chao, Yehong Xu, Wen Hua, and Xiaofang Zhou. 2020. A survey on map-matching algorithms. In *Australasian Database Conference*. Springer, 121–133.

[6] Rui Chen, Benjamin Fung, and Bipin C Desai. 2011. Differentially private trajectory data publication. *arXiv preprint arXiv:1112.2020* (2011).

[7] Rui Chen, Benjamin CM Fung, Bipin C Desai, and Nériah M Sossou. 2012. Differentially private transit data publication: a case study on the montreal transportation system. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 213–221.

[8] Yan Dai, Jie Shao, Chengbo Wei, Dongxiang Zhang, and Heng Tao Shen. 2018. Personalized semantic trajectory privacy preservation through trajectory reconstruction. *World Wide Web* 21, 4 (2018), 875–914.

[9] Ekler P de Mattos, Augusto CSA Domingues, and Antonio AF Loureiro. 2019. Give me two points and i'll tell you who you are. In *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 1081–1087.

[10] Yves-Alexandre De Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. 2013. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports* 3, 1 (2013), 1–5.

[11] Edsger W Dijkstra et al. 1959. A note on two problems in connexion with graphs. *Numerische mathematik* 1, 1 (1959), 269–271.

[12] David H Douglas and Thomas K Peucker. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: the international journal for geographic information and geovisualization* 10, 2 (1973), 112–122.

[13] Cynthia Dwork. 2006. Differential Privacy. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming, part II (ICALP) (Lecture Notes in Computer Science)*, Vol. 4052. 1–12.

[14] Cynthia Dwork. 2008. Differential Privacy: A Survey of Results. In *International Conference on Theory and Applications of Models of Computation*. Springer, 1–19.

[15] Ehab ElSalamouny and Sébastien Gambs. 2016. Differential privacy models for location-based services. *Transactions on Data Privacy* 9, 1 (2016), 15–48.

[16] Jie Feng, Zeyu Yang, Fengli Xu, Haisu Yu, Mudan Wang, and Yong Li. 2020. Learning to simulate human mobility. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3426–3433.

[17] Marco Fiore, Panagiota Katsikouli, Elli Zavou, Mathieu Cunche, Françoise Fessant, Dominique Le Hello, Ulrich Aivodji, Baptiste Olivier, Tony Quertier, and Razvan Stanica. 2020. Privacy in trajectory micro-data publishing: a survey. *Transactions on Data Privacy* 13 (2020), 91–149.

[18] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[19] Marco Gramaglia and Marco Fiore. 2015. Hiding mobile traffic fingerprints with glove. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*. 1–13.

[20] Mehmet Emre Gursoy, Ling Liu, Stacey Truex, Lei Yu, and Wenqi Wei. 2018. Utility-aware synthesis of differentially private and attack-resilient location traces. In *Proceedings of the 2018 ACM SIGSAC conference on computer and communications security*. 196–211.

[21] Peter E Hart, Nils J Nilsson, and Bertram Raphael. 1968. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics* 4, 2 (1968), 100–107.

[22] Samiul Hasan, Christian M Schneider, Satish V Ukkusuri, and Marta C González. 2013. Spatiotemporal patterns of urban human mobility. *Journal of Statistical Physics* 151, 1 (2013), 304–318.

[23] Ammar Haydari, Michael Zhang, Chen-Nee Chuah, Jane Macfarlane, and Sean Peisert. 2021. Adaptive Differential Privacy Mechanism for Aggregated Mobility Dataset. *arXiv preprint arXiv:2112.08487* (2021).

[24] Xi He, Graham Cormode, Ashwin Machanavajjhala, Cecilia M Procopiuc, and Divesh Srivastava. 2015. DPT: differentially private trajectory synthesis using hierarchical reference systems. *Proceedings of the VLDB Endowment* 8, 11 (2015), 1154–1165.

[25] Alejandro Henao and Wesley E Marshall. 2019. The impact of ride-hailing on vehicle miles traveled. *Transportation* 46, 6 (2019), 2173–2194.

[26] Baik Hoh and Marco Gruteser. 2005. Protecting location privacy through path confusion. In *First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SECURECOMM'05)*. IEEE, 194–205.

[27] Fengmei Jin, Wen Hua, Matteo Francia, Pingfu Chao, Maria Orlowska, and Xiaofang Zhou. 2021. A Survey and Experimental Study on Privacy-Preserving Trajectory Data Publishing. (2021).

[28] Meng Li, Liehuang Zhu, Zijian Zhang, and Rixin Xu. 2017. Achieving differential privacy of trajectory data publishing in participatory sensing. *Information Sciences* 400 (2017), 1–13.

[29] Jianhua Lin. 1991. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information theory* 37, 1 (1991), 145–151.

[30] Xiangyu Liu, Jinmei Chen, Xiufeng Xia, Chuanyu Zong, Rui Zhu, and Jiajia Li. 2019. Dummy-based trajectory privacy protection against exposure location attacks. In *International Conference on Web Information Systems and Applications*. Springer, 368–381.

[31] Frank McSherry and Kunal Talwar. 2007. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. IEEE, 94–103.

[32] Darakhshan J Mir, Sibren Isaacman, Ramón Cáceres, Margaret Martonosi, and Rebecca N Wright. 2013. Dp-Where: Differentially Private Modeling of Human Mobility. In *2013 IEEE International Conference on Big Data*. IEEE, 580–588.

[33] Jameson D Morgan. 2020. GeoAware-A Simulation-based Framework for Synthetic Trajectory Generation from Mobility Patterns. (2020).

[34] Jason W Powell, Yan Huang, Favyen Bastani, and Minhe Ji. 2011. Towards reducing taxicab cruising time using spatio-temporal profitability maps. In *International Symposium on spatial and temporal Databases*. Springer, 242–260.

[35] Wahbeh Qardaji, Weining Yang, and Ninghui Li. 2013. Differentially Private Grids for Geospatial Data. In *2013 IEEE 29th International Conference on Data Engineering (ICDE)*. IEEE, 757–768.

[36] Mohammed A Quddus, Washington Y Ochieng, and Robert B Noland. 2007. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation research part c: Emerging technologies* 15, 5 (2007), 312–328.

[37] Jinmeng Rao, Song Gao, Yuhao Kang, and Qunying Huang. 2020. LSTM-TrajGAN: A deep learning approach to trajectory privacy protection. *arXiv preprint arXiv:2006.10521* (2020).

[38] Caroline Rodier. 2009. Review of international modeling literature: Transit, land use, and auto pricing strategies to reduce vehicle miles traveled and greenhouse gas emissions. *Transportation Research Record* 2132, 1 (2009), 1–12.

[39] Meead Saberi, Hani S Mahmassani, Dirk Brockmann, and Amir Hosseini. 2017. A complex network perspective for characterizing urban travel demand patterns: graph theoretical analysis of large-scale origin–destination demand networks. *Transportation* 44, 6 (2017), 1383–1402.

[40] Julián Salas, David Megías, and Vicenç Torra. 2018. SwapMob: Swapping trajectories for mobility anonymization. In *International Conference on Privacy in Statistical Databases*. Springer, 331–346.

[41] Katrina Ward, Dan Lin, and Sanjay Madria. 2020. A Parallel Algorithm For Anonymizing Large-scale Trajectory Data. *ACM Transactions on Data Science* 1, 1 (2020), 1–26.

[42] Jianhao Wei, Yaping Lin, Xin Yao, and Jin Zhang. 2019. Differential privacy-based location protection in spatial crowdsourcing. *IEEE Transactions on Services Computing* (2019).

[43] Xiaokui Xiao, Gabriel Bender, Michael Hay, and Johannes Gehrke. 2011. iReduct: Differential privacy with reduced relative errors. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*. 229–240.

[44] Yonghui Xiao and Li Xiong. 2015. Protecting locations with differential privacy under temporal correlations. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 1298–1309.

[45] Xiaodong Zhao, Dechang Pi, and Junfu Chen. 2020. Novel trajectory privacy-preserving method based on clustering using differential privacy. *Expert Systems with Applications* 149 (2020), 113241.