Differentially Private Map Matching for Mobility Trajectories

Ammar Haydari, Chen-Nee Chuah, Michael Zhang, Jane Macfarlane,
Sean Peisert

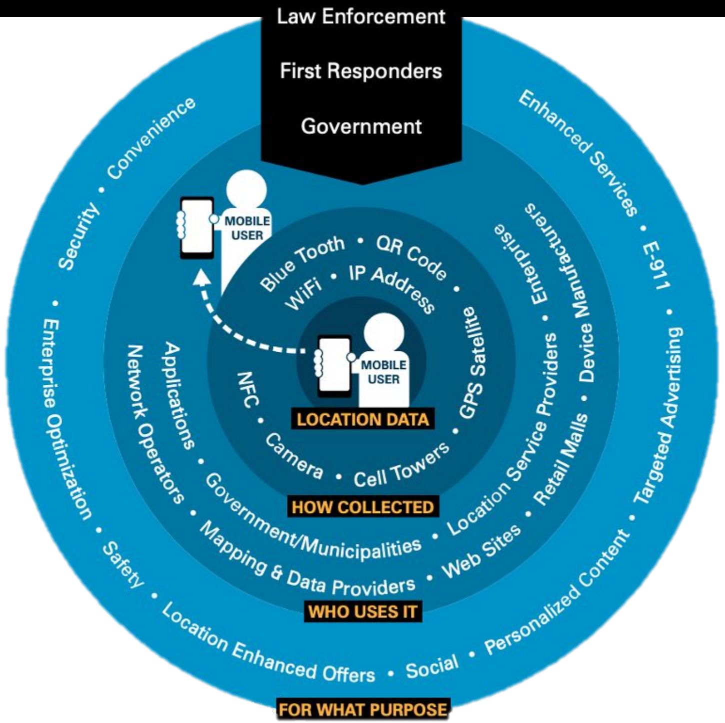
ACSAC 2022

UCDAVIS
UNIVERSITY OF CALIFORNIA

Berkeley
UNIVERSITY OF CALIFORNIA

BERKELEY LAB

Location Ecosystem



Law Enforcement

First Responders

Government

MOBILE USER

MOBILE USER

LOCATION DATA

Blue Tooth • QR Code • WiFi • IP Address • GPS Satellite • Cell Towers • Camera • NFC •

HOW COLLECTED

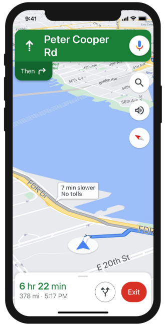
Enterprise • Device Manufacturers • Retail Malls • Web Sites • Location Service Providers • Mapping & Data Providers • Government/Municipalities • Applications • Network Operators

WHO USES IT

Security • Convenience • Enhanced Services • E-911 • Targeted Advertising • Personalized Content • Social • Location Enhanced Offers • Safety • Enterprise Optimization •

FOR WHAT PURPOSE

Source: Skyhook



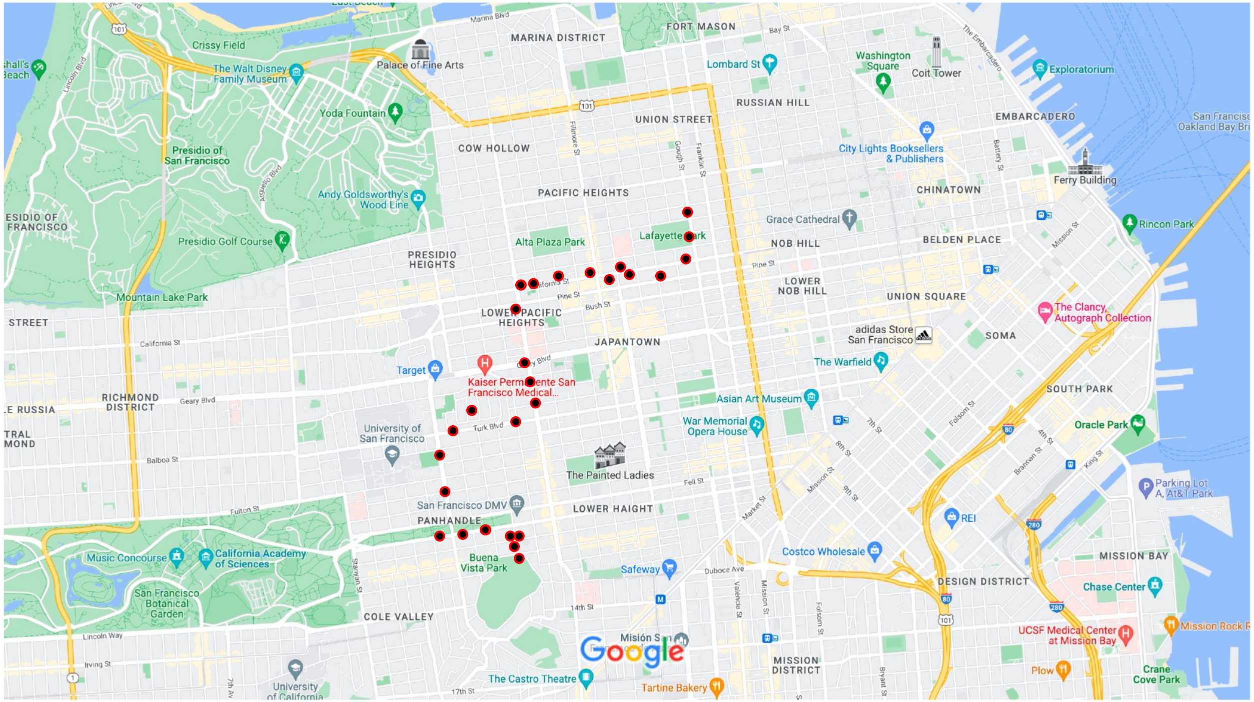
GPS Trajectory

- Sequence of GPS coordinates
- Continuous motion of an object
- Inherits rich information for mobility analysis
- However, user trajectory data is **sensitive**

It may reveal

- Home - office address
- Route taken by the user
- Interaction of user with communities

Google Maps



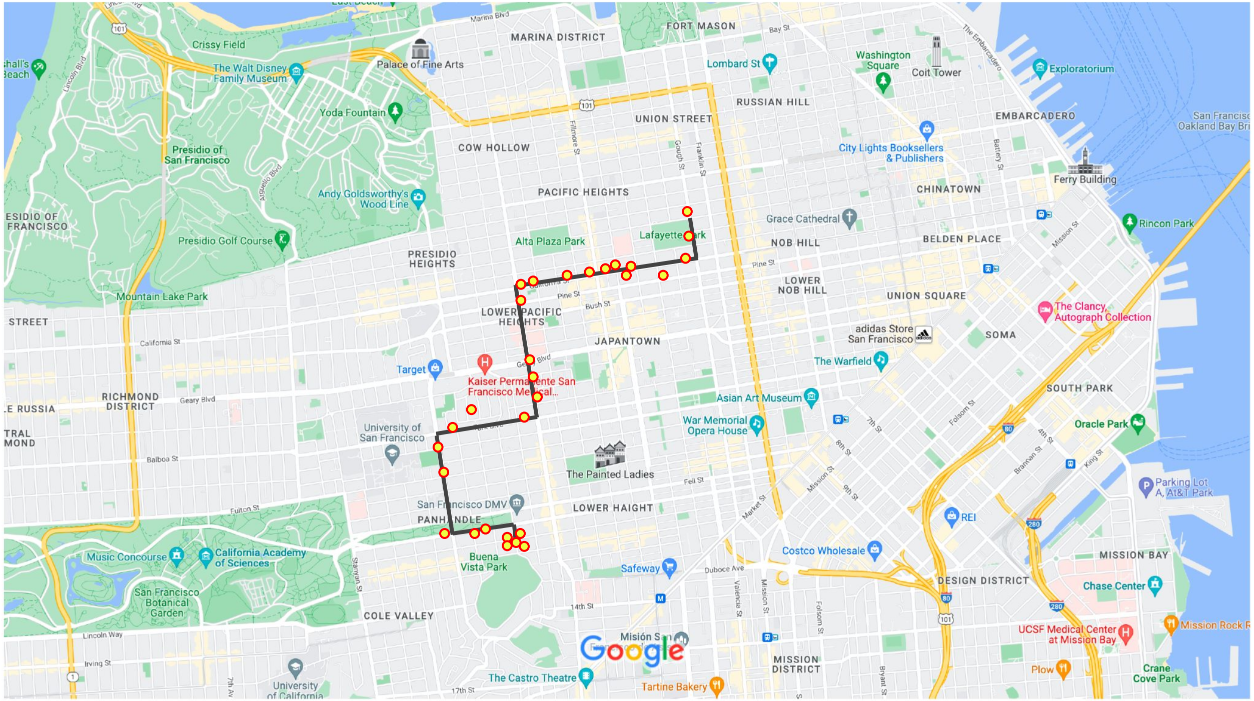
Map data ©2022 Google 2000 ft

Map-Matched Trajectory

Map-matching

- Minimizes trajectory error
- Ordered set of road links
- Reflects exact travel path

Google Maps



Map data ©2022 Google 2000 ft

Prior Works on Trajectory Privacy

Prior works attempt to provide private trajectories through generative models

- Human mobility modeling
- Prefix-tree
- Generative adversarial networks

Other methods are

- Noise injection to locations
- Location swapping
- Dumppy location injection
- Trajectory reconstruction

Our Goal

This paper attempts to protect the privacy of every individual trajectory regardless of the rest of the data

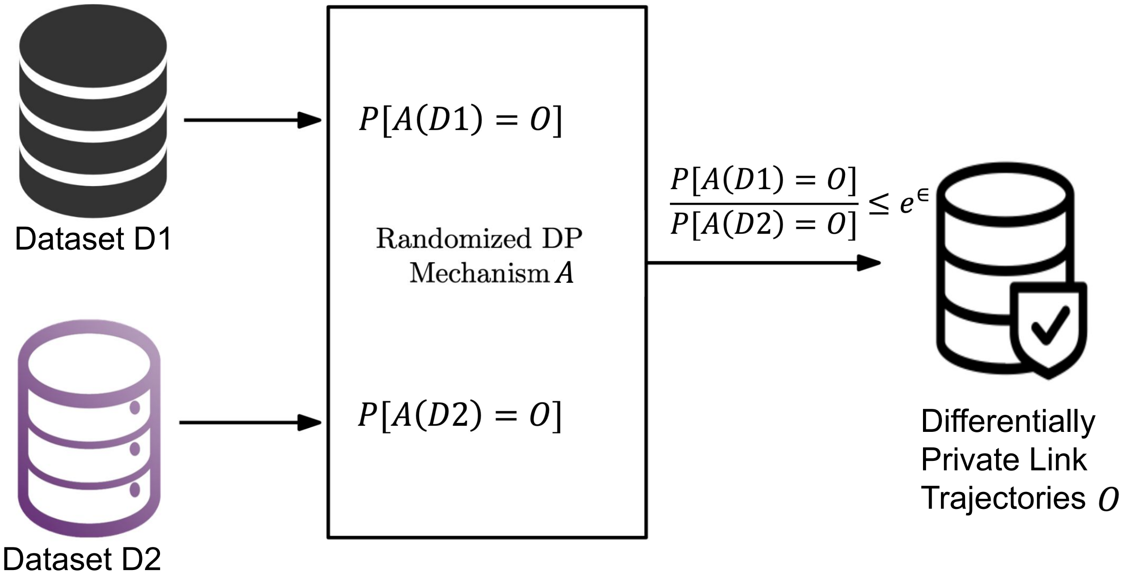
- masking origin and destinations (OD) with adaptive noise injection
- randomizing travel paths with exponential selection.

Differentially **P**rivate **M**ap-**M**atching (DPMM)

- Preserves high utility in protected trajectories
- Prevents geospatial mismatches
- Keeps the trajectories in reasonable path

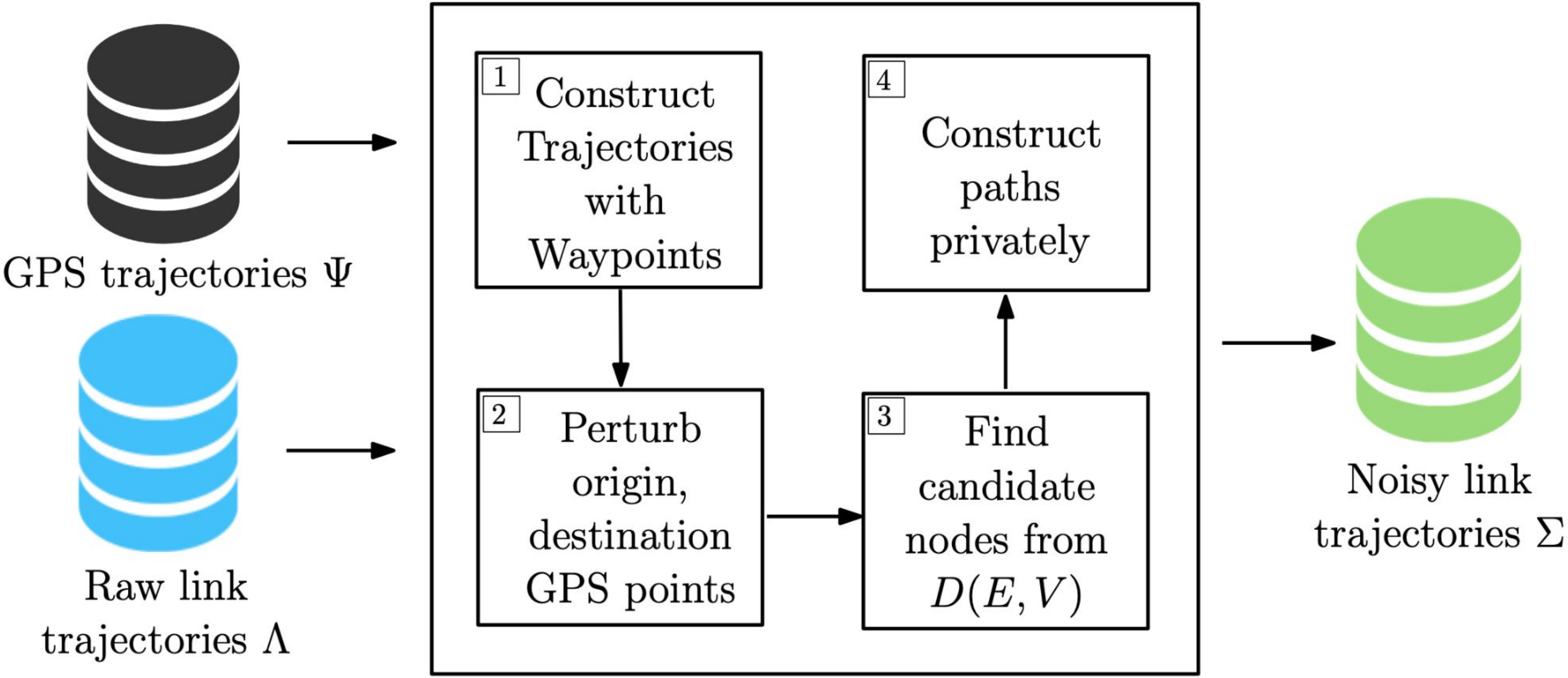
Differential Privacy

Differential Privacy statistically guarantees the privacy of individual user trajectories independent of the background knowledge and other samples.

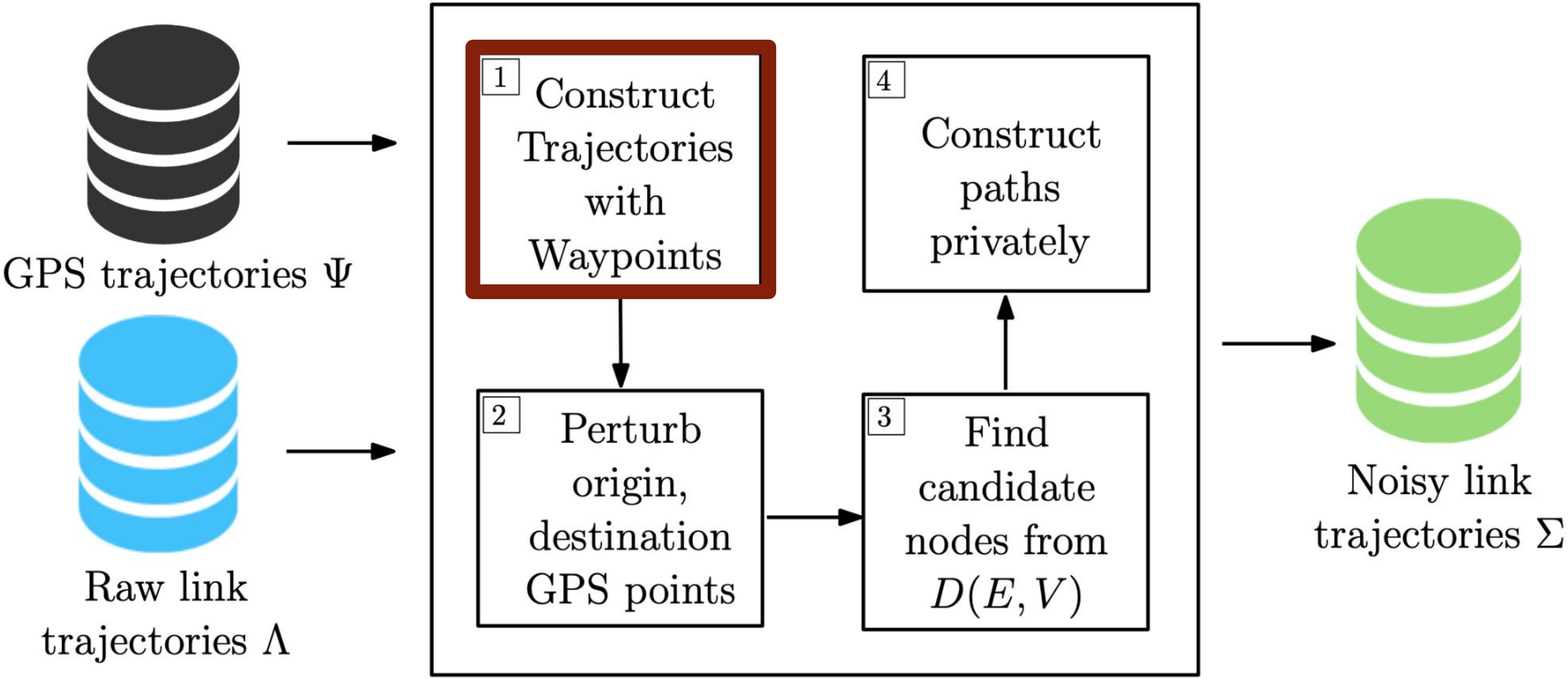


Dataset D1
Dataset D2
$P[A(D1) = O]$
Randomized DP Mechanism A
$P[A(D2) = O]$
$\frac{P[A(D1) = O]}{P[A(D2) = O]} \leq e^{\epsilon}$
Differentially Private Link Trajectories O

Differentially Private Map-Matching


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

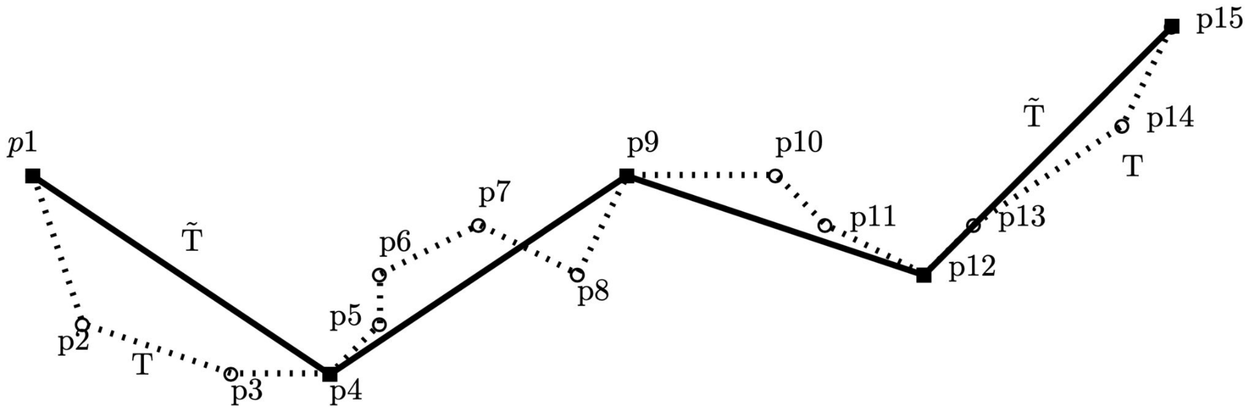
Trajectory Simplification


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

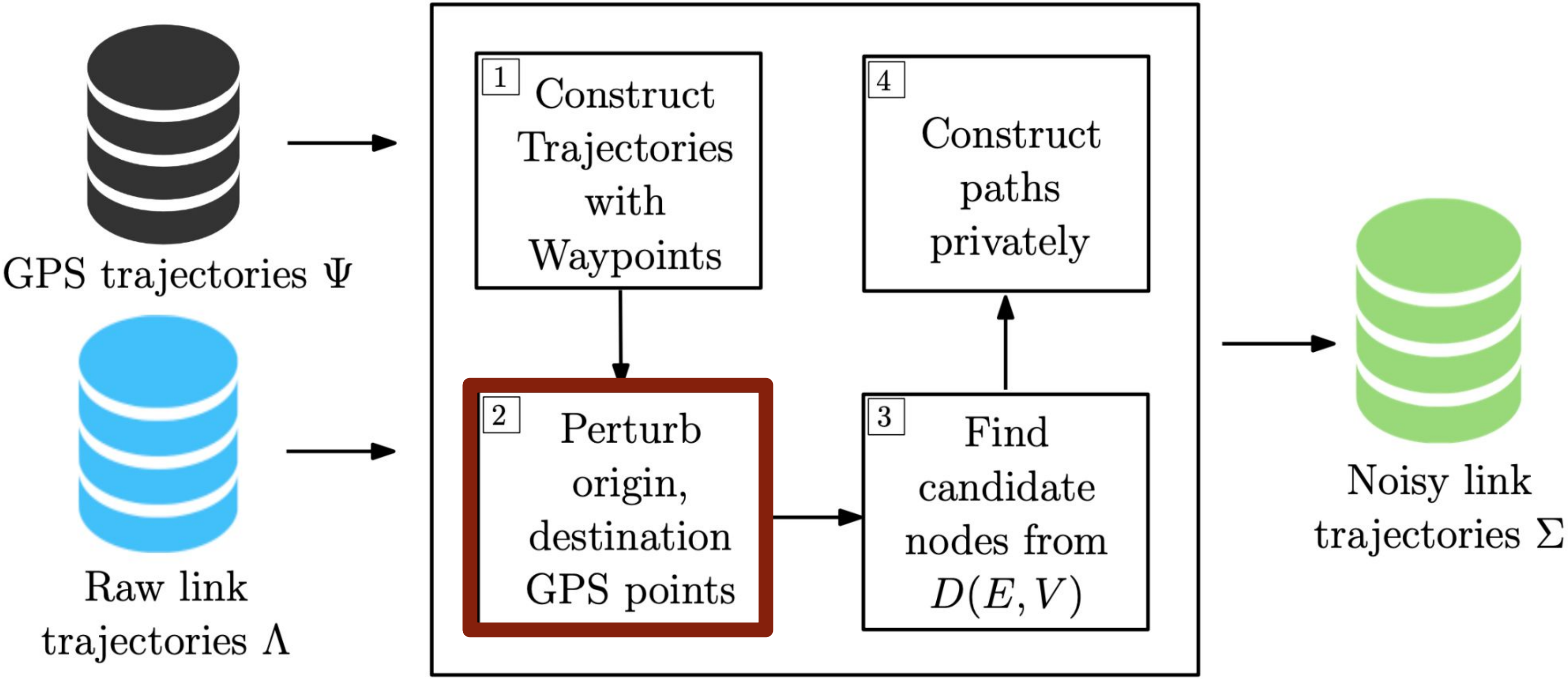
Ramer–Douglas–Peucker (RDP)

Represent the sequence of location with waypoints

- retain the movement characteristics
- enhances the path quality
- decreases the computational complexity

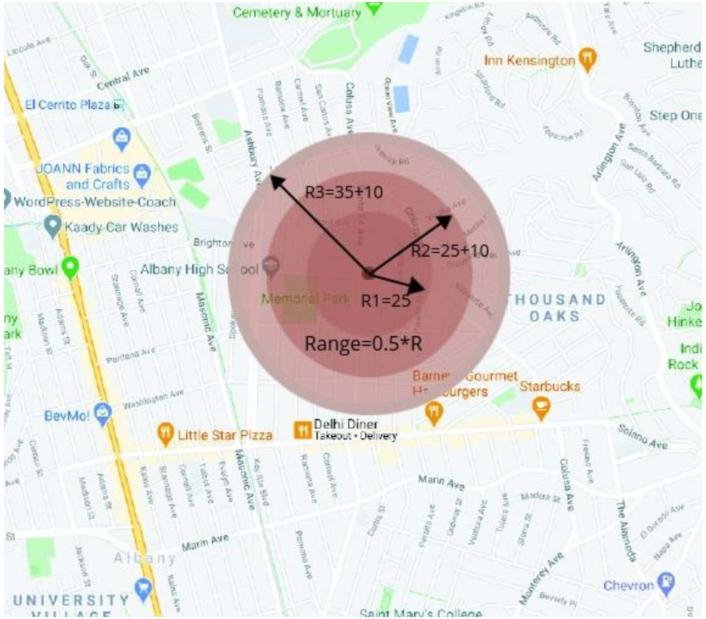


Private Origin-Destinations

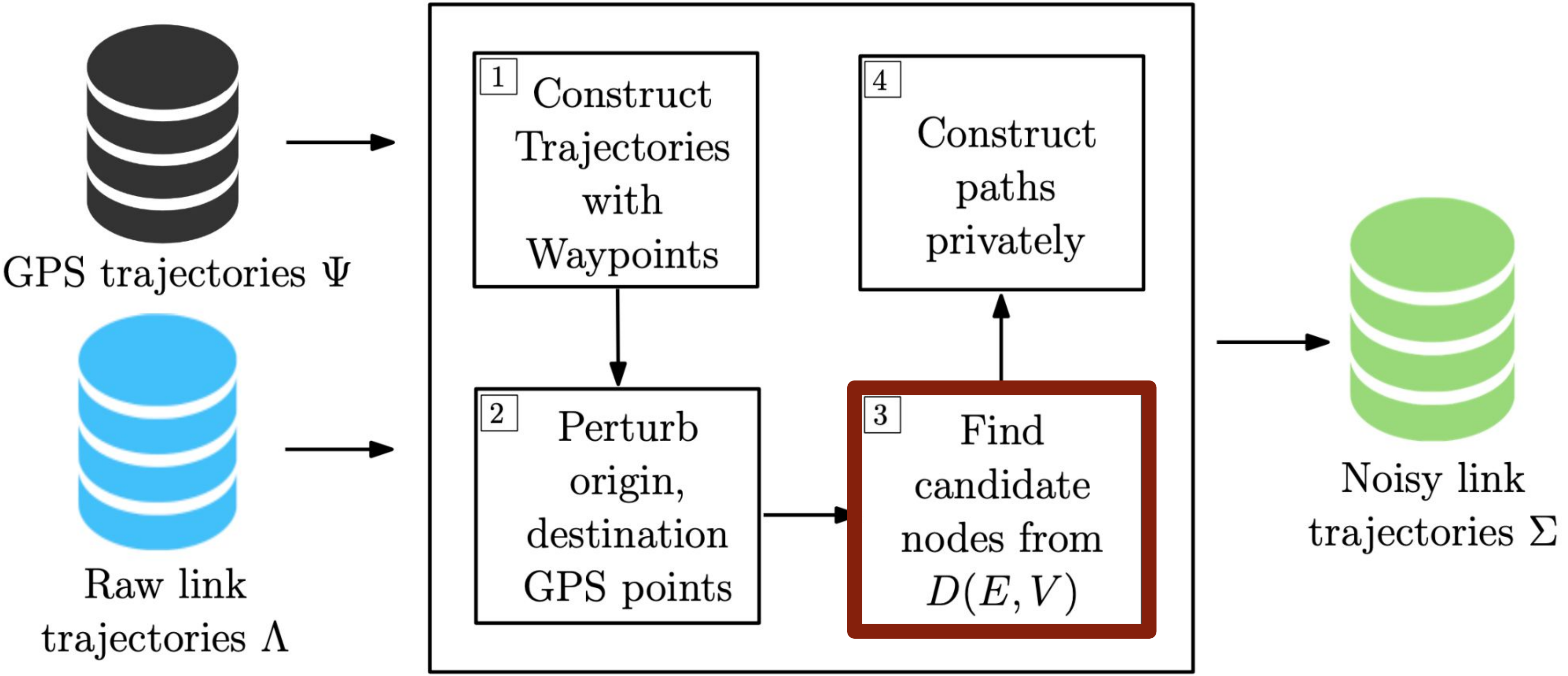

GPS trajectories Ψ
Raw link trajectories Λ
1 Construct Trajectories with Waypoints
2 Perturb origin, destination GPS points
3 Find candidate nodes from $D(E, V)$
4 Construct paths privately
Noisy link trajectories Σ

Private Origin-Destinations

- Origin-destinations have the highest privacy concern
- 2D Planar Laplace noise added to the OD GPS points adaptively
- Planar Laplace Noise has two parameters
  - Epsilon (∈)
  - Range (r)
- Road link network density specifies the Radius
  - Number of road links in cloaking region
- **Output** is noisy a GPS point



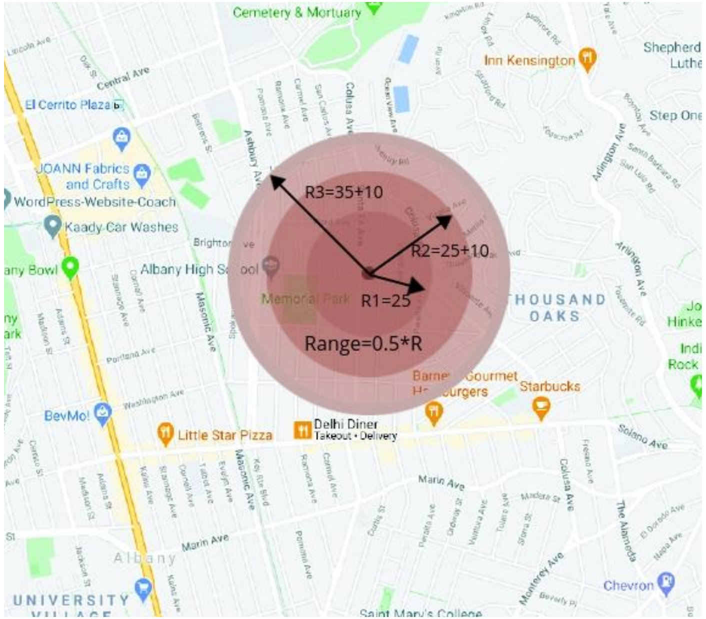
Candidate Nodes



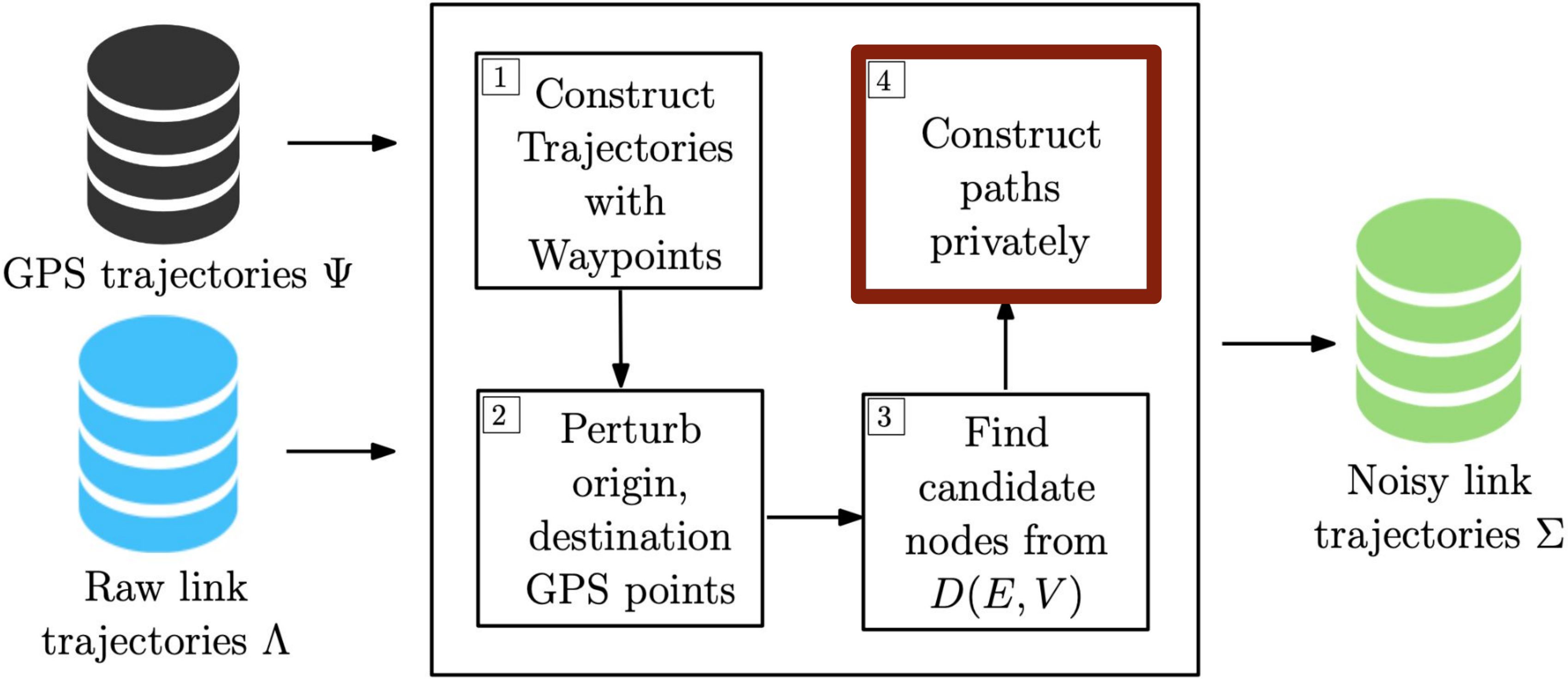
GPS trajectories Ψ
Raw link trajectories Λ
1
Construct Trajectories with Waypoints
2
Perturb origin, destination GPS points
3
Find candidate nodes from $D(E,V)$
4
Construct paths privately
Noisy link trajectories Σ

Candidate Nodes

- Retrieve candidate nodes from the road network for the noisy OD GPS points and waypoints
  - with the same cloaking region approach
- For OD GPS points
  - Consider functional class information from clean link trajectory
- Link functional class defines the type of the road that trajectory travels
  - Increase the similarity of the noisy and original link trajectories



Private Paths



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

Private Paths and Trajectory Construction

1. A* path selection algorithm
   - Find candidate paths between the nodes
   - Combines Dijkstra shortest path algorithm with greedy search
2. Exponential DP mechanism
   - Randomly selects a path from candidate paths
3. Connect sequence of candidate paths
   - Form the full trajectory
4. Trajectory Postprocessing
   - Remove the travel loops
   - Retains the DP guarantee

Attack Resilience of DPMM

Outlier Leakage on Trajectory ODs:

- ODs are sensitive that reveal user identity
- Rural areas are more unique than central areas
- Adaptively noise injection protects against outlier leakage

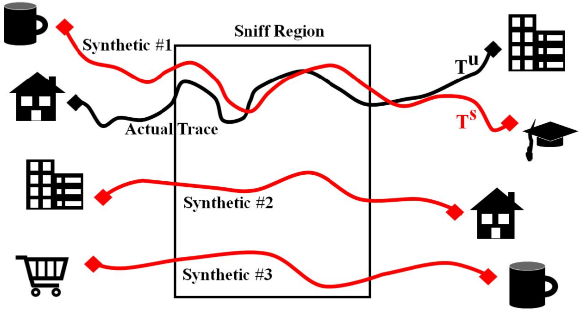
Attack Resilience of DPMM

Outlier Leakage on Trajectory ODs:

- ODs are sensitive that reveal user identity
- Rural areas are more unique than central areas
- Adaptively noise injection protects against outlier leakage

Partial Sniffing on Travel Paths:

- Adversary learns partial trajectory
- Match the rest of the trajectory with trajectory dataset
- Private path construction creates randomized paths
- Adversary cannot make correct inferences about user



Source: Gursoy et al., 2018

Experiment Setup

833 real fleet and consumer trajectories in San Francisco collected from different location-sharing apps and GPS devices.

Compared with different variants of DPMM and 2 external studies

- DPMM-No-WP
- DPMM-A*-WP
- DPMM-D-WP
- **DPT** (Ha et al., VLDB 2015)
- **AdaTrace** (Gursoy et al., CCS 2018)

Utility Metrics

- Privatized OD link ratio
- Trip length change
- Query error of spatial density
- OD similarity
- Vehicle miles traveled
- Road link count densities

Experimental Results

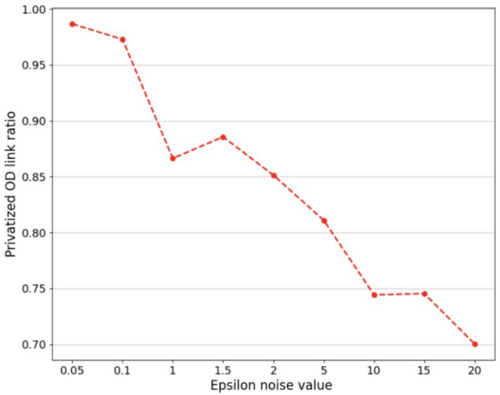
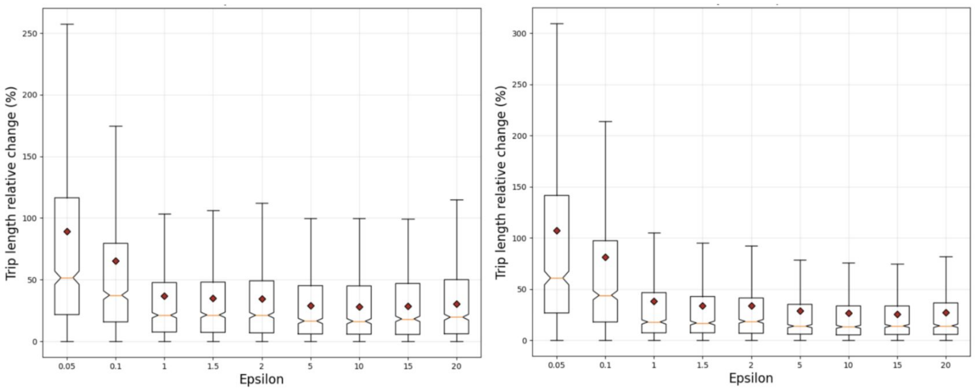


Figure 4: Comparison of different ϵ values and the change of OD-links for different for 1 hour period of trajectories between 1pm and 2pm.

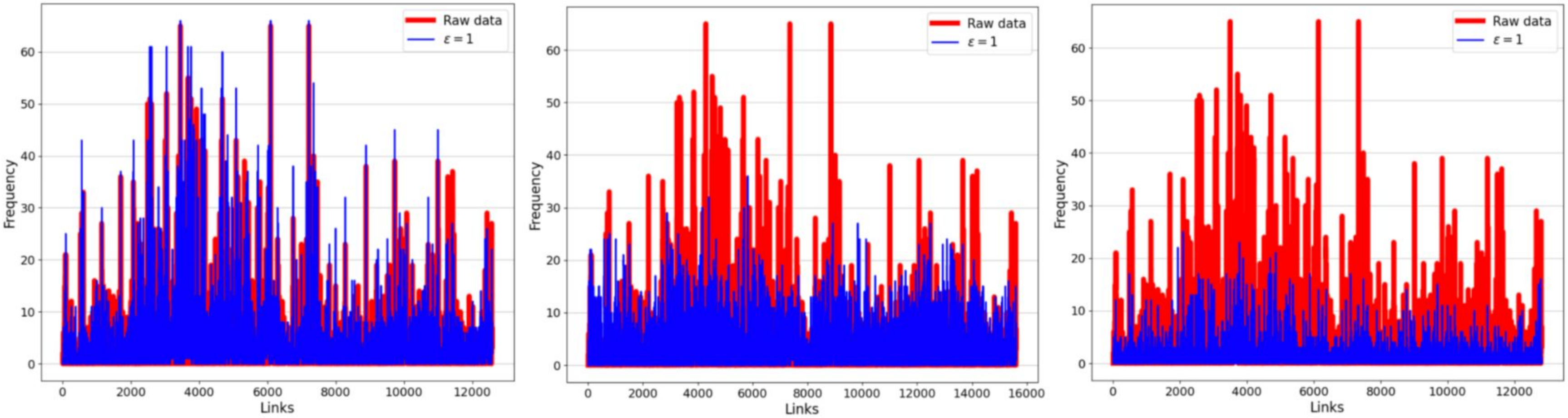


(a) DPMM vs original link trajectories

(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.

Experimental Results



(a) DPMM **(b)** AdaTrace **(c)** DPT

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

Experimental Results

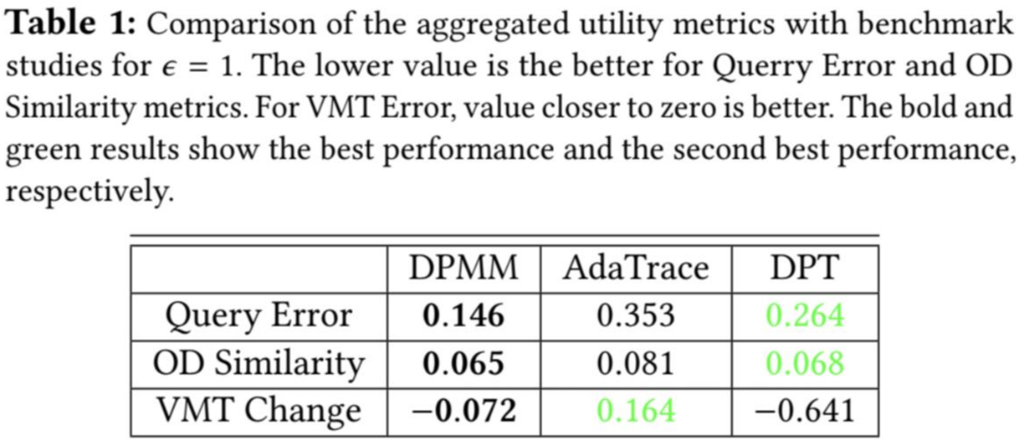
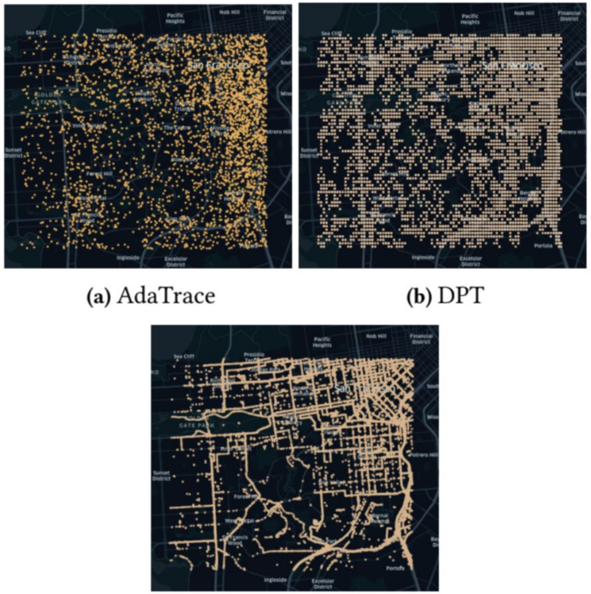


Table 1: Comparison of the aggregated utility metrics with benchmark studies for $\epsilon = 1$. The lower value is the better for Querry Error and OD Similarity metrics. For VMT Error, value closer to zero is better. The bold and green results show the best performance and the second best performance,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

Conclusion

- DPMM is a new trajectory privacy method with higher utility
  - OD privacy with noise injection DP mechanism
  - Path privacy with exponential DP mechanism
- Do not rely on the other samples in the dataset
- Minimizes the trajectory mismatches with geospatial reality
- Quantified at individual and aggregated utility metrics
- Superior against the all compared baselines

Thank You

Planer Laplace noise model

- Sample a GPS point from planer laplace distribution with two parameters
  - Epsilon (ε) adjusts the noise level
  - Range (r) determines the center of the laplace noise

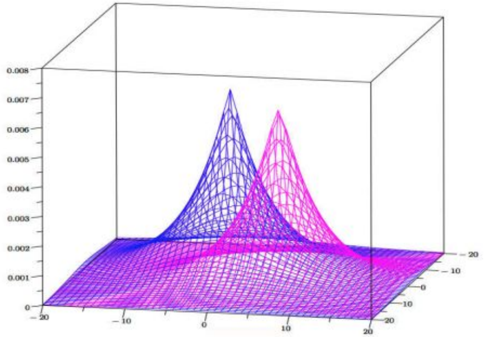


Figure 2: The pdf of two planar Laplacians, centered at $(-2,-4)$ and at $(5,3)$ respectively, with $\epsilon = 1/5$.

- Moving direction is selected randomly in angular form between [0 and 2**π**]
- While (ε) is the input parameter, the range value (r) is adaptivity parameter selected by the model

- (ε) defines the tightness of Laplace distribution, smaller epsilon moves more links