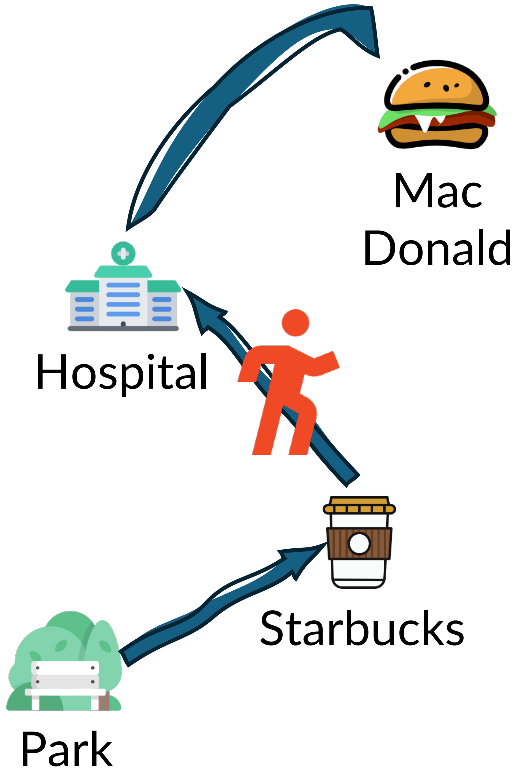
TraCS: Trajectory Collection in Continuous Space under Local Differential Privacy

Authors: Ye Zheng, Yidan Hu

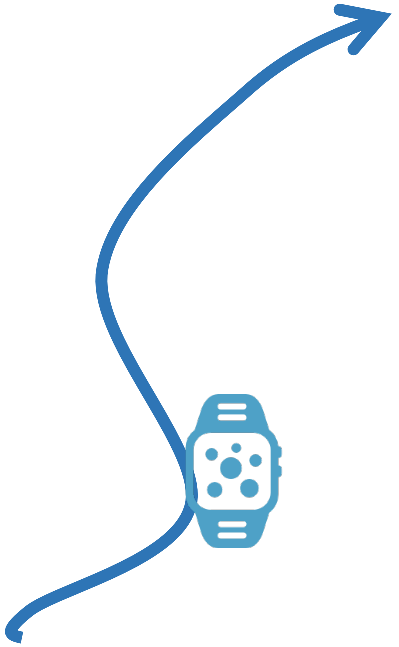
Rochester Institute of Technology

Trajectory Collection

- Sensitive trajectories: daily routine, wearable-sensor trajectories



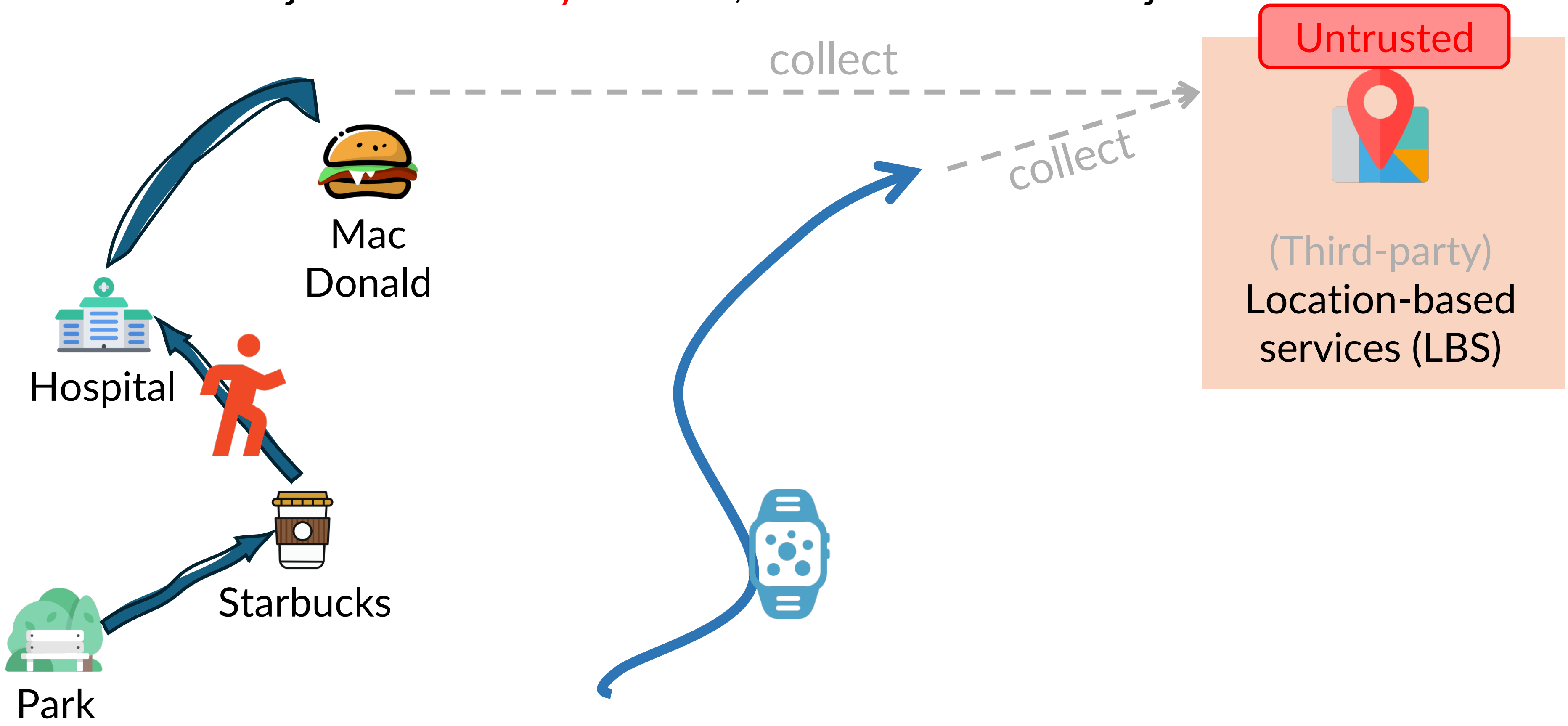
Discrete locations
w/ labels



Continuous locations
w/o labels

Trajectory Collection

- Sensitive trajectories: daily routine, wearable-sensor trajectories



Discrete locations w/ labels

Continuous locations w/o labels

LDP-fy a Trajectory

- LDP-fy: perturb a trajectory with LDP guarantee (provable privacy)
  - cannot distinguish location τ_1 from τ_2 with confidence quantified by e^{ε}

$$\forall \tau_1, \tau_2, \tau_* \in \mathcal{S}: \quad \frac{\Pr[\mathcal{M}(\tau_1) = \tau_*]}{\Pr[\mathcal{M}(\tau_2) = \tau_*]} \leq e^{\varepsilon}$$

location space

distinguishability of τ_1 from τ_2 when observing τ_*

LDP-fy a Trajectory

- LDP-fy: perturb a trajectory with LDP guarantee (provable privacy)
  - cannot distinguish location τ_1 from τ_2 with confidence quantified by e^{ε}

$$\forall \tau_1, \tau_2, \tau_* \in \mathcal{S}: \quad \frac{\Pr[\mathcal{M}(\tau_1) = \tau_*]}{\Pr[\mathcal{M}(\tau_2) = \tau_*]} \leq e^{\varepsilon}$$

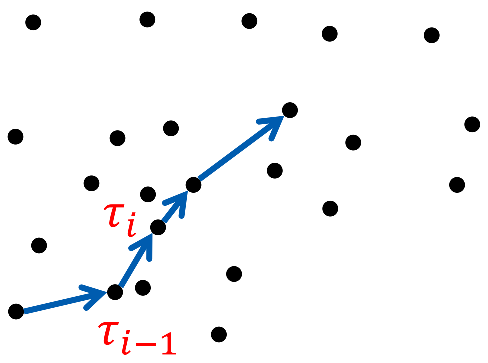
location space

distinguishability of τ_1 from τ_2 when observing τ_*

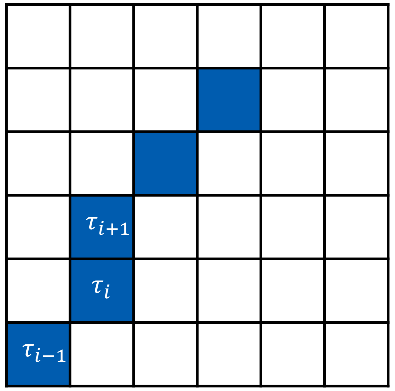
- Target: LDP mechanism $\mathcal{M}$ (for the location space)
  - provable privacy for users' trajectories, in continuous spaces or discrete spaces
  - as high trajectory utility as possible

LDP-fy a Trajectory in A **Discrete Space**

- Discrete location spaces: Point of interests or discretized cells



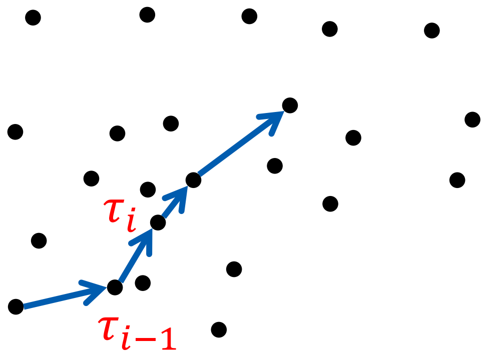
$\mathcal{S} = \{p_1, \dots, p_n\}$



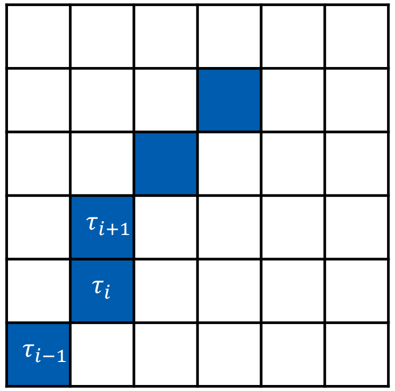
$\mathcal{S} = \{c_1, \dots, c_n\}$

LDP-fy a Trajectory in A **Discrete Space**

- Discrete location spaces: Point of interests or discretized cells



$\mathcal{S} = \{p_1, \ldots, p_n\}$



$\mathcal{S} = \{c_1, \ldots, c_n\}$

- Exponential mechanism:

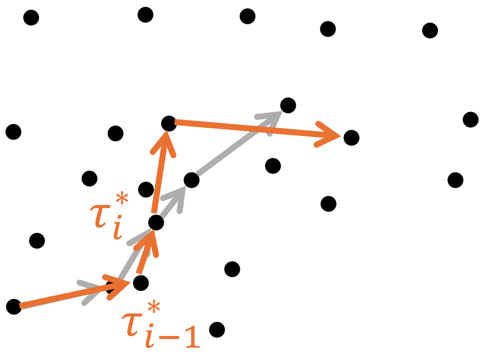
$$\Pr[\mathcal{M}_{\exp}(\tau) = \tau^*] = \frac{\exp\left(\varepsilon d(\tau, \tau^*)\right)}{\sum_{\tau' \in S} \exp\left(\varepsilon d(\tau, \tau')\right)}$$

Pairwise distance (numerator)

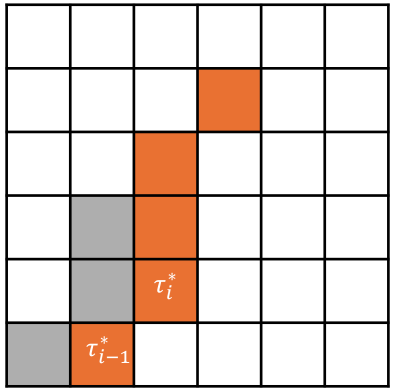
Sum of distance (denominator)

LDP-fy a Trajectory in A **Discrete Space**

- Discrete location spaces: Point of interests or discretized cells



$\mathcal{S} = \{p_1, \dots, p_n\}$



$\mathcal{S} = \{c_1, \dots, c_n\}$

- Exponential mechanism:

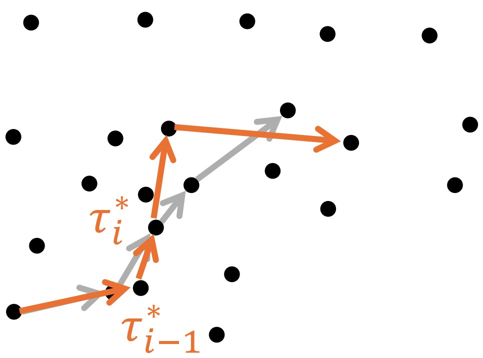
$$\Pr[\mathcal{M}_{\exp}(\tau) = \tau^*] = \frac{\exp\left(\varepsilon d(\tau, \tau^*)\right)}{\sum_{\tau' \in S} \exp\left(\varepsilon d(\tau, \tau')\right)}$$

← Pairwise distance

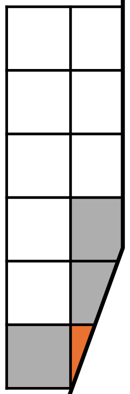
← Sum of distance

LDP-fy a Trajectory in A **Discrete Space**

- Discrete location spaces: Point of interests or discretized c



$\mathcal{S} = \{p_1, \dots, p_n\}$



Limitations

1. Efficiency

- each sample costs $\mathcal{O}(n)$

2. Trajectory Utility

- decreases as n increases

3. Applicability

- cannot apply to continuous $\mathcal{S}$

- Exponential mechanism:

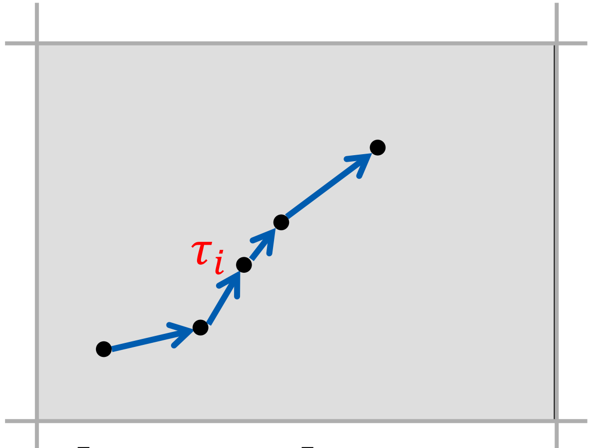
$$\Pr[\mathcal{M}_{\exp}(\tau) = \tau^*] = \frac{\exp\left(\varepsilon d(\tau, \tau^*)\right)}{\sum_{\tau' \in S} \exp\left(\varepsilon d(\tau, \tau')\right)}$$

← Pairwise distance

← Sum of distance

Continuous Spaces: Better & Universal

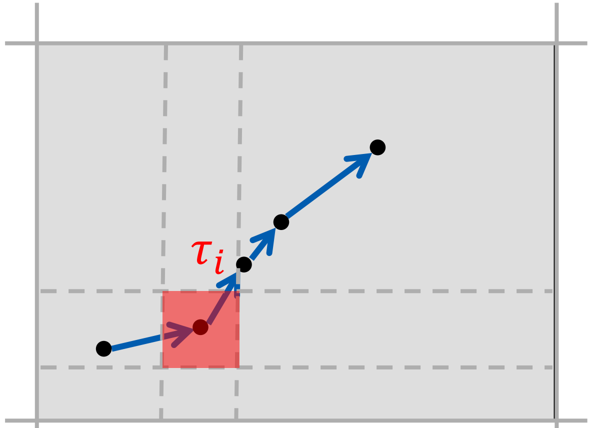
- Operate on a continuous space (covering the discrete space)



$\mathcal{S} = [a_{\text{sta}}, a_{\text{end}}] \times [b_{\text{sta}}, b_{\text{end}}]$

Continuous Spaces: Better & Universal

- Operate on a continuous space (covering the discrete space)



$\mathcal{S} = [a_{\mathrm{sta}}, a_{\mathrm{end}}] \times [b_{\mathrm{sta}}, b_{\mathrm{end}}]$

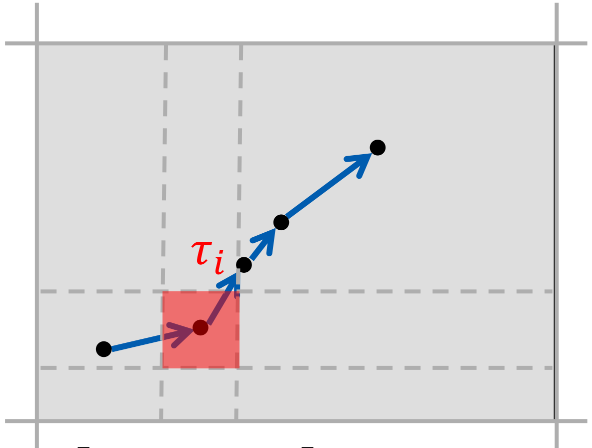
Simple sampling: (satisfying LDP for $\mathcal{S}$)

$\Pr[\mathcal{M}(x_i) = \tilde{x} \in \blacksquare] = p_\varepsilon$ (red square)

$\Pr[\mathcal{M}(x_i) = \tilde{x} \in \blacksquare] = p_\varepsilon / e^\varepsilon$ (grey square)

Continuous Spaces: Better & Universal

- Operate on a continuous space (covering the discrete space)



$\mathcal{S} = [a_{\text{sta}}, a_{\text{end}}] \times [b_{\text{sta}}, b_{\text{end}}]$

Simple sampling: (satisfying LDP for $\mathcal{S}$)

$$\Pr[\mathcal{M}(x_i) = \tilde{x} \in \blacksquare] = p_\varepsilon$$

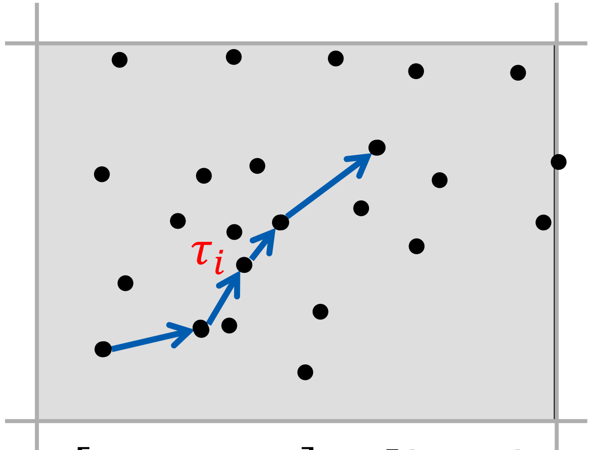
$$\Pr[\mathcal{M}(x_i) = \tilde{x} \in \blacksquare] = p_\varepsilon / e^\varepsilon$$

- Benefits:

1. **Efficiency:** $\mathcal{O}(1)$ sampling complexity
2. **Trajectory utility:** "n-independent"
3. **Applicability:** Both continuous spaces & discrete spaces

Continuous Spaces: Better & Universal

- Operate on a continuous space (covering the discrete space)



$\mathcal{S} = [a_{\mathrm{sta}}, a_{\mathrm{end}}] \times [b_{\mathrm{sta}}, b_{\mathrm{end}}]$

Simple sampling: (satisfying LDP for $\mathcal{S}$)

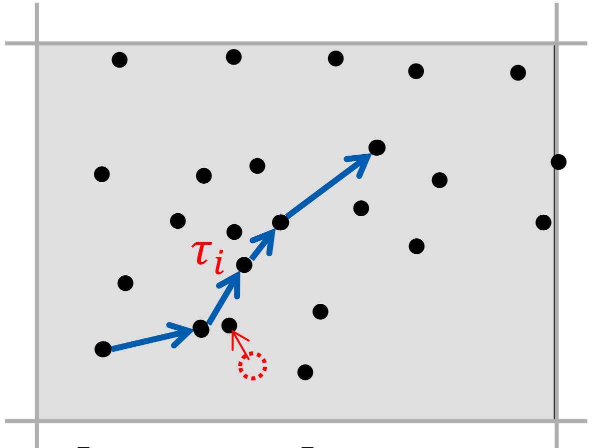
$$\Pr[\mathcal{M}(x_i) = \tilde{x} \in \text{■}] = p_\varepsilon$$

$$\Pr[\mathcal{M}(x_i) = \tilde{x} \in \text{■}] = p_\varepsilon / e^\varepsilon$$

- Benefits:
  1. **Efficiency:** $\mathcal{O}(1)$ sampling complexity
  2. **Trajectory utility:** "n-independent"
  3. **Applicability:** Both continuous spaces & discrete spaces

Continuous Spaces: Better & Universal

- Operate on a continuous space (covering the discrete space)



$\mathcal{S} = [a_{\mathrm{sta}}, a_{\mathrm{end}}] \times [b_{\mathrm{sta}}, b_{\mathrm{end}}]$

Simple sampling: (satisfying LDP for $\mathcal{S}$)

$$\Pr[\mathcal{M}(x_i) = \tilde{x} \in \blacksquare] = p_\varepsilon$$

$$\Pr[\mathcal{M}(x_i) = \tilde{x} \in \blacksquare] = p_\varepsilon / e^\varepsilon$$

- Benefits:

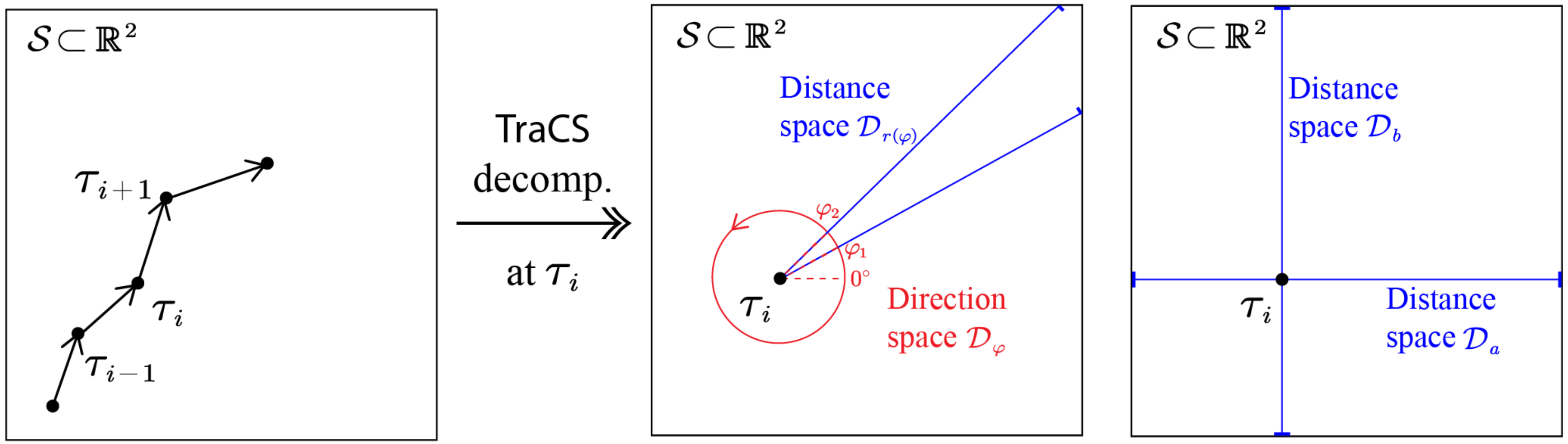
1. **Efficiency:** $\mathcal{O}(1)$ sampling complexity
2. **Trajectory utility:** "n-independent"
3. **Applicability:** Both continuous spaces & discrete spaces

Rounding-to-the-nearest (post-processing)

TraCS: Trajectory Collection in Continuous Spaces

- TraCS-D: direction-distance perturbation
- TraCS-C: coordinates perturbation

- **Key idea:** decomposes $\mathcal{S}$ into two subspaces



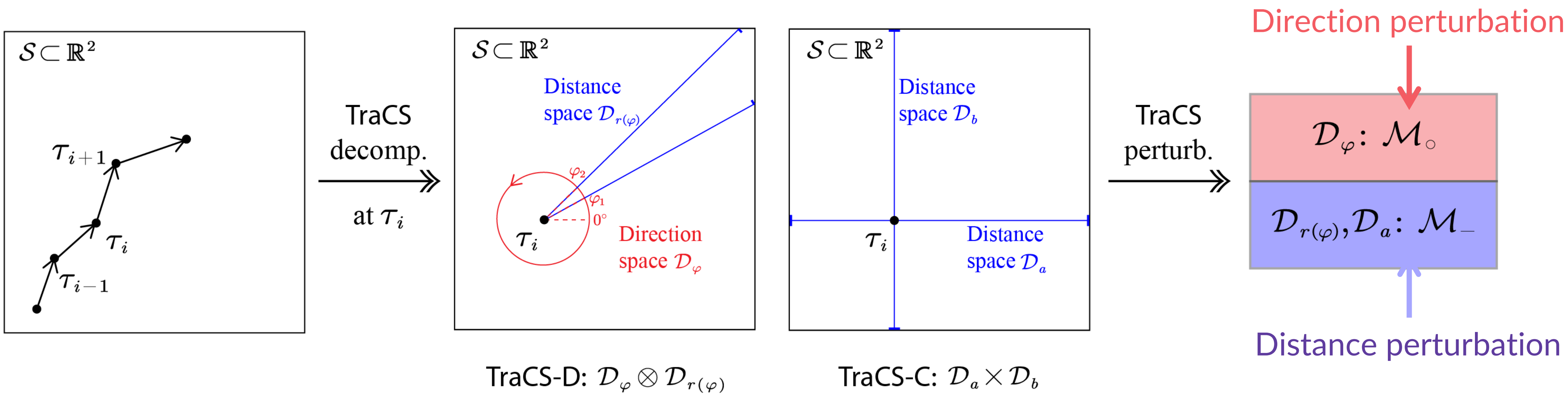
TraCS-D: $\mathcal{D}_\varphi \otimes \mathcal{D}_{r(\varphi)}$

TraCS-C: $\mathcal{D}_a \times \mathcal{D}_b$

TraCS: Trajectory Collection in Continuous Spaces

- TraCS-D: direction-distance perturbation
- TraCS-C: coordinates perturbation

- **Key idea:** decomposes $\mathcal{S}$ into two subspaces $\rightarrow$ design $\mathcal{M}$ for each subspace

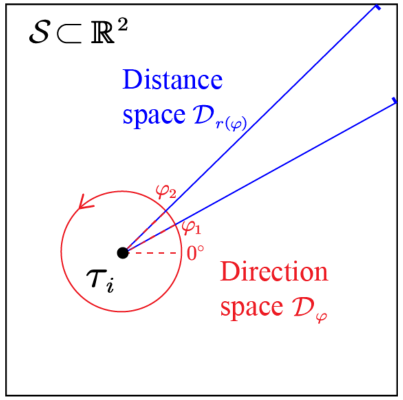


Decomposition of Continuous Spaces

- $\mathcal{S} = \mathcal{D}_{\varphi} \otimes \mathcal{D}_{r(\varphi)}$ ⟵ 1D subspace: linear

  2D space ↑   1D subspace: circular ↖

- Each location $\tau_{i+1} \in \mathcal{S}$ has a **unique** representation $(\varphi, r(\varphi))$
- Perturb φ and $r(\varphi)$ using 1D mechanisms



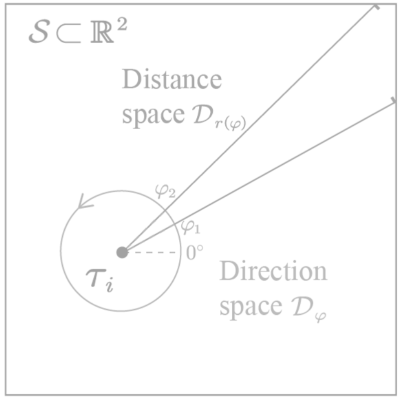
TraCS-D: $\mathcal{D}_{\varphi} \otimes \mathcal{D}_{r(\varphi)}$

Decomposition of Continuous Spaces

- $\mathcal{S} = \mathcal{D}_{\varphi} \otimes \mathcal{D}_{r(\varphi)}$ ← 1D subspace: linear

2D space 1D subspace: circular

- Each location $\tau_{i+1} \in \mathcal{S}$ has a **unique** representation $(\varphi, r(\varphi))$
- Perturb φ and $r(\varphi)$ using 1D mechanisms

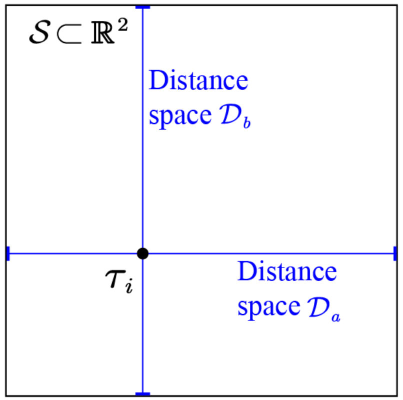


TraCS-D: $\mathcal{D}_{\varphi} \otimes \mathcal{D}_{r(\varphi)}$

- $\mathcal{S} = \mathcal{D}_a \times \mathcal{D}_b$

2D space 1D subspace: linear

- Each location $\tau_{i+1} \in \mathcal{S}$ has a **unique** representation (a, b)
- Perturb a and b using 1D mechanisms



TraCS-C: $\mathcal{D}_a \times \mathcal{D}_b$

LDP Mechanisms $\mathcal{M}_\circ$ and $\mathcal{M}_-$

- Leverage piecewise-based mechanisms $\mathcal{M}_\circ$ and $\mathcal{M}_-$

  $[0,2\pi) \rightarrow [0,2\pi)$ $[0,1) \rightarrow [0,1)$

  - other LDP mechanisms for bounded numerical domain also applicable

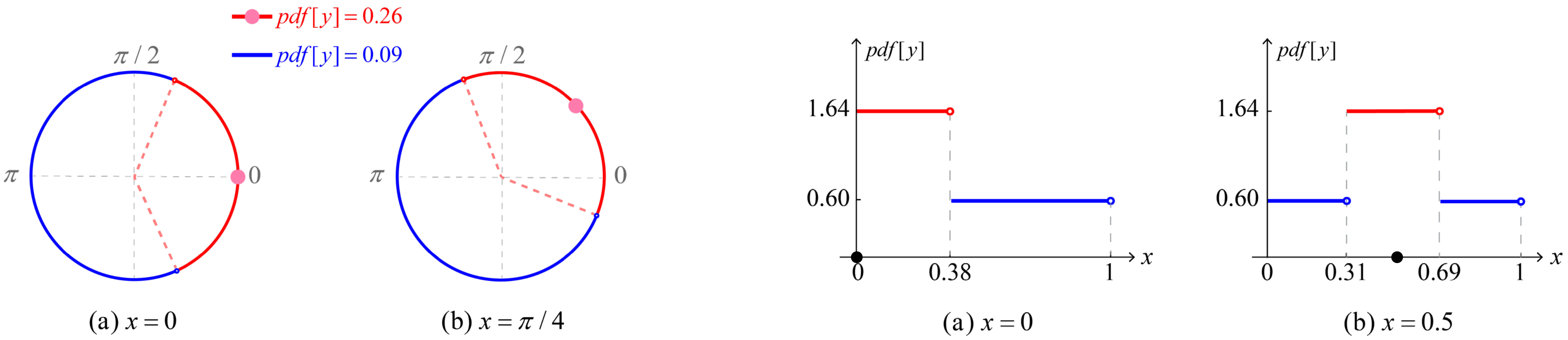
LDP Mechanisms $\mathcal{M}_\circ$ and $\mathcal{M}_-$

- Leverage piecewise-based mechanisms $\mathcal{M}_\circ$ and $\mathcal{M}_-$

  $[0,2\pi) \to [0,2\pi)$ $\quad$ $[0,1) \to [0,1)$

  - other LDP mechanisms for bounded numerical domain also applicable

- Examples of $\mathcal{M}_\circ$ and $\mathcal{M}_-$*



$\mathcal{M}_\circ(x; \varepsilon = 1)$ $\qquad$ $\mathcal{M}_-(x; \varepsilon = 1)$

* Optimal Piecewise-based Mechanism for Collecting Bounded Numerical Data under Local Differential Privacy, PETS'25

Evaluations

- Claims
  - continuous spaces: better utility than naïve baselines, e.g. planar Laplace + truncation
  - discrete spaces: better efficiency and utility than existing methods, ATP*, NGram**, L-SRR***

* Trajectory Data Collection with Local Differential Privacy, VLDB'23

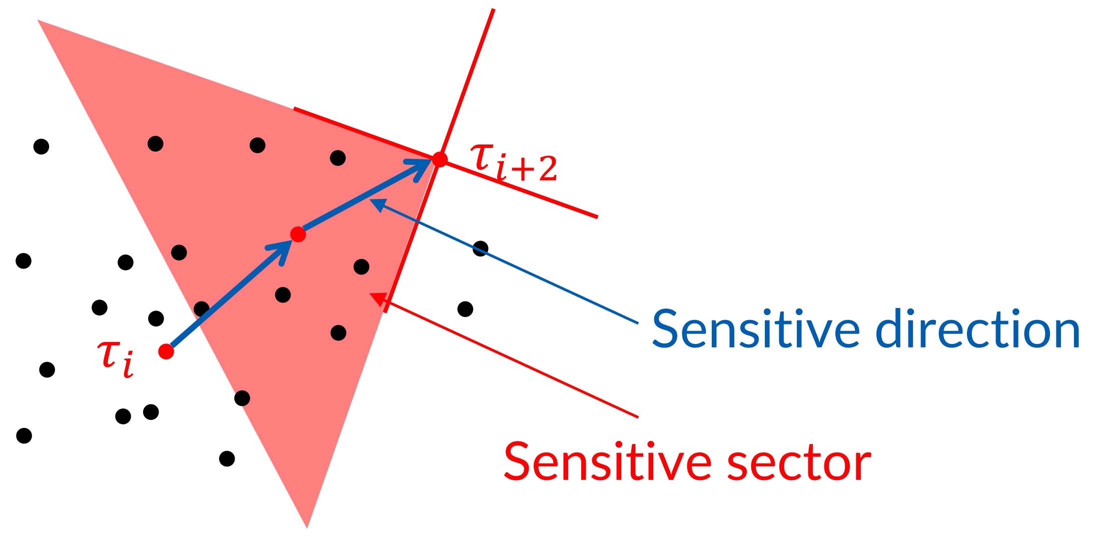
** Real-World Trajectory Sharing with Local Differential Privacy, VLDB'21

*** L-SRR: Local Differential Privacy for Location-Based Services with Staircase Randomized Response, CCS'22

Evaluations

- Claims
  - continuous spaces: better utility than naïve baselines, e.g. planar Laplace + truncation
  - discrete spaces: better efficiency and utility than existing methods, ATP*, NGram**, L-SRR***

- **ATP:** direction perturbation
  1. divide direction sectors, e.g. $k = 4$



* Trajectory Data Collection with Local Differential Privacy, VLDB'23

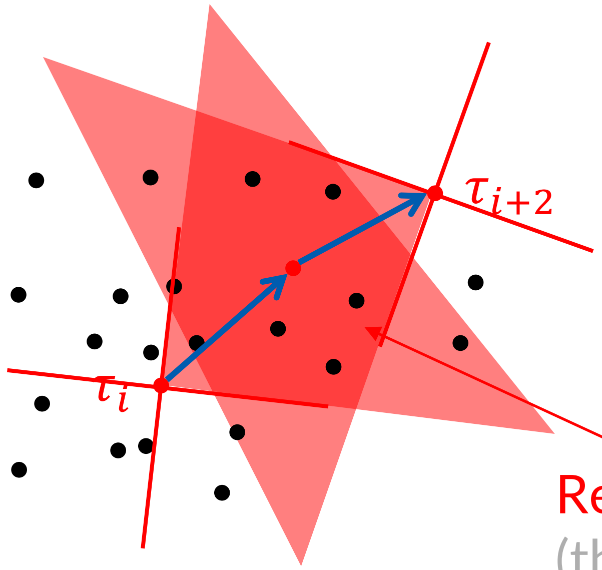
** Real-World Trajectory Sharing with Local Differential Privacy, VLDB'21

*** L-SRR: Local Differential Privacy for Location-Based Services with Staircase Randomized Response, CCS'22

Evaluations

- Claims
  - continuous spaces: better utility than naïve baselines, e.g. planar Laplace + truncation
  - discrete spaces: better efficiency and utility than existing methods, ATP*, NGram**, L-SRR***

- **ATP:** direction perturbation
  1. divide direction sectors, e.g. $k = 4$
  2. perturb sector
  3. $\mathcal{M}_{\exp}(\tau)$ on reduced $\mathcal{S}$



Reduced location space $\mathcal{S}$
(then use EM)

* Trajectory Data Collection with Local Differential Privacy, VLDB'23

** Real-World Trajectory Sharing with Local Differential Privacy, VLDB'21

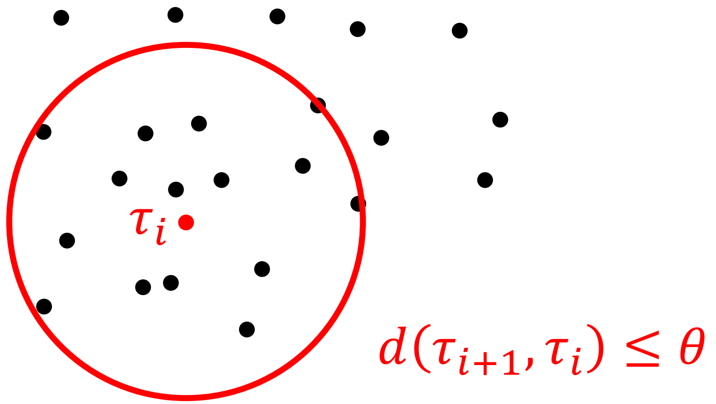
*** L-SRR: Local Differential Privacy for Location-Based Services with Staircase Randomized Response, CCS'22

Evaluations

- Claims
  - continuous spaces: better utility than naïve baselines, e.g. planar Laplace + truncation
  - discrete spaces: better efficiency and utility than existing methods, ATP*, NGram**, L-SRR***

- **NGram:** reachability constraint from public knowledge

1. distance reachability, i.e. the next location cannot be too far
2. $\mathcal{M}_{\exp}(\tau)$ on reduced $\mathcal{S}$



* Trajectory Data Collection with Local Differential Privacy, VLDB'23

** Real-World Trajectory Sharing with Local Differential Privacy, VLDB'21

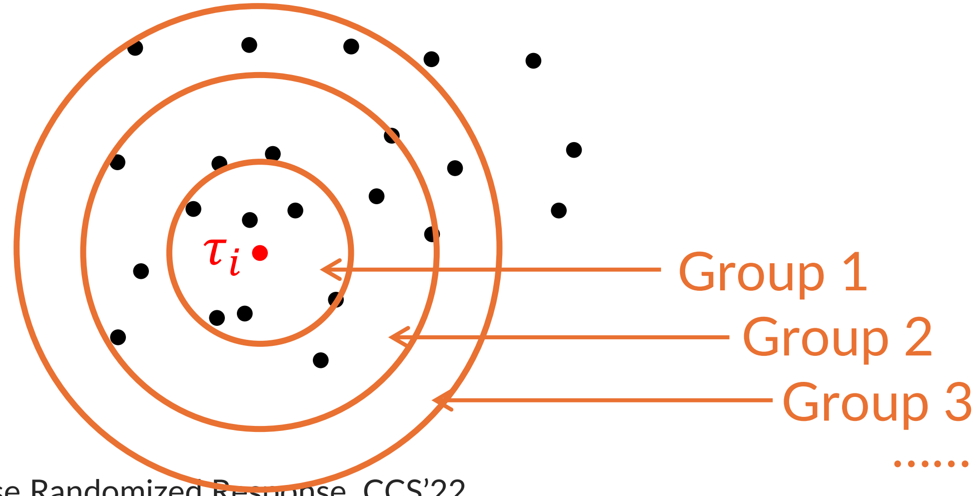
*** L-SRR: Local Differential Privacy for Location-Based Services with Staircase Randomized Response, CCS'22

Evaluations

- Claims
  - continuous spaces: better utility than naïve baselines, e.g. planar Laplace + truncation
  - discrete spaces: better efficiency and utility than existing methods, ATP*, NGram**, L-SRR***

- **L-SRR:** Staircase Randomized Response

  1. group the locations to k groups
  2. $\mathcal{M}_{\text{SRR}}(\tau)$ on k groups



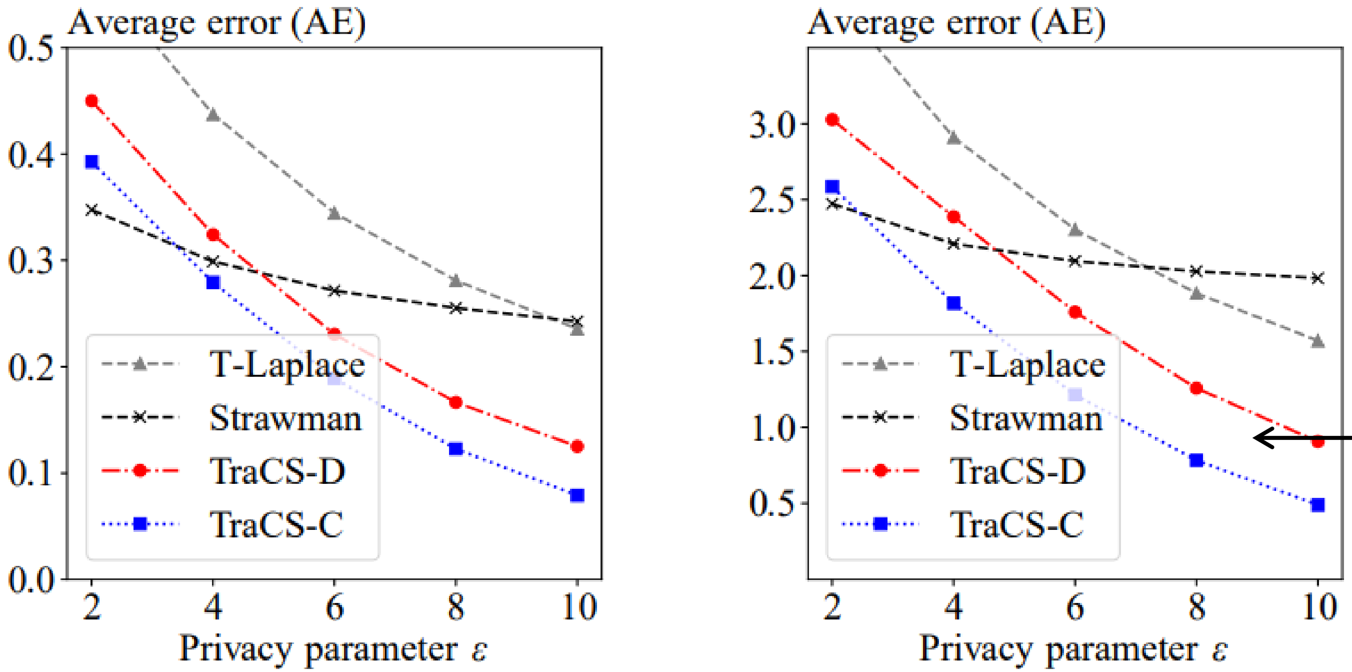
* Trajectory Data Collection with Local Differential Privacy, VLDB'23

** Real-World Trajectory Sharing with Local Differential Privacy, VLDB'21

*** L-SRR: Local Differential Privacy for Location-Based Services with Staircase Randomized Response, CCS'22

Evaluations – Continuous Spaces

- Trajectory utility metric: Average error (AE)
  - T-Laplace: Truncated Laplace; Strawman: k-RR + uniform sampling for direction perturbation

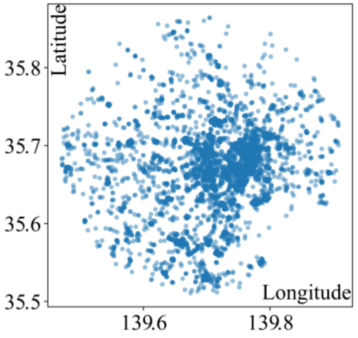


(a) $\mathcal{S} = [0, 1) \times [0, 1)$ **(b)** $\mathcal{S} = [0, 2) \times [0, 10)$

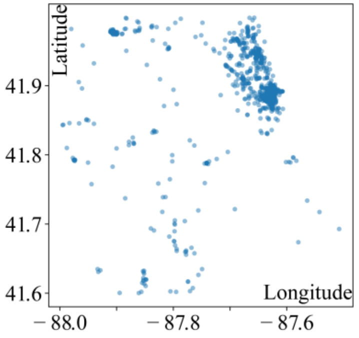
Lower error when ε becomes larger

Evaluations – Continuous Spaces

- TKY and CHI datasets
  - perturb the trajectories in continuous spaces formed by long-lati:



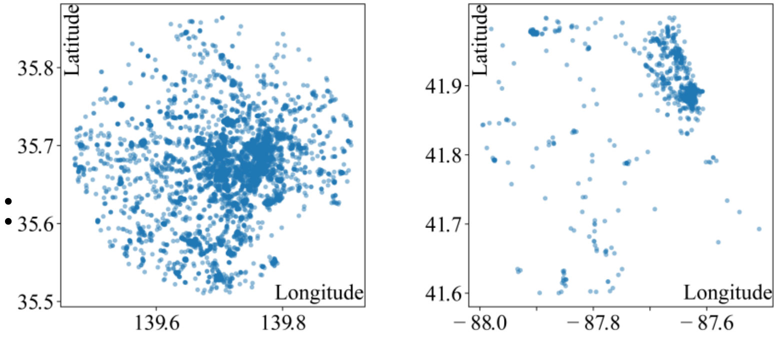
(a) TKY location space



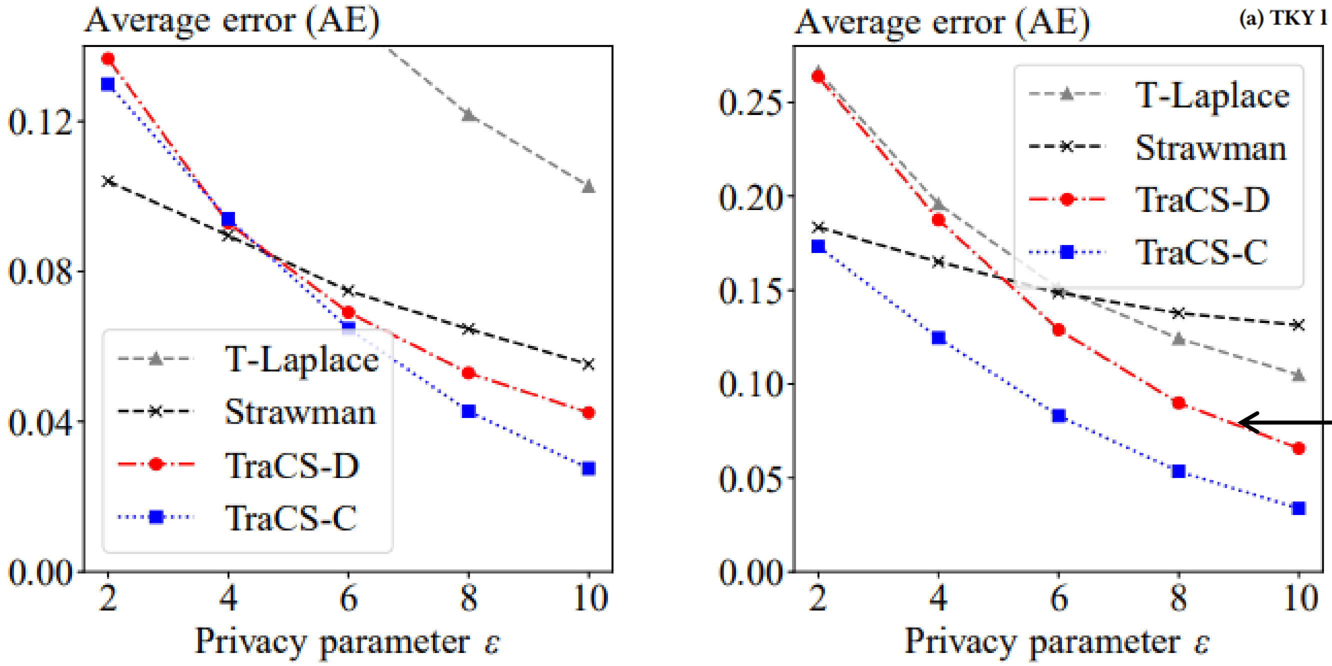
(b) CHI location space

Evaluations – Continuous Spaces

- TKY and CHI datasets
  - perturb the trajectories in continuous spaces formed by long-lati:



(a) TKY l cation space (b) CHI location space

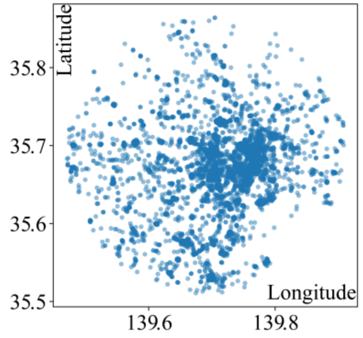


(a) TKY dataset (b) CHI dataset

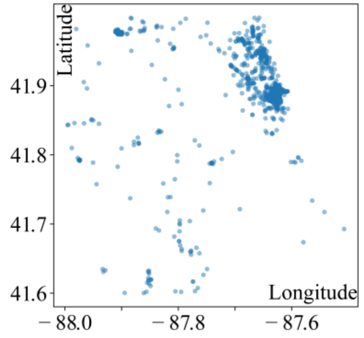
Lower error when ε becomes larger

Evaluations – Discrete Spaces

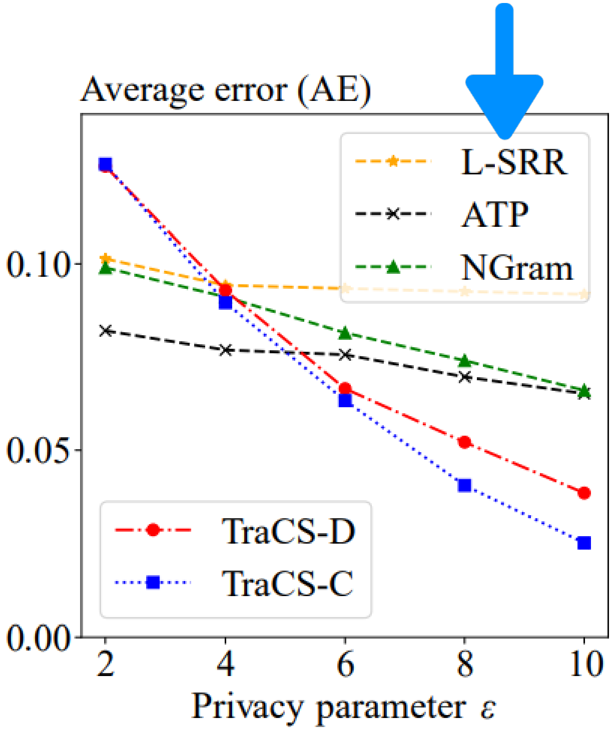
- TKY and CHI datasets
  - TraCS + rounding-to-the-nearest



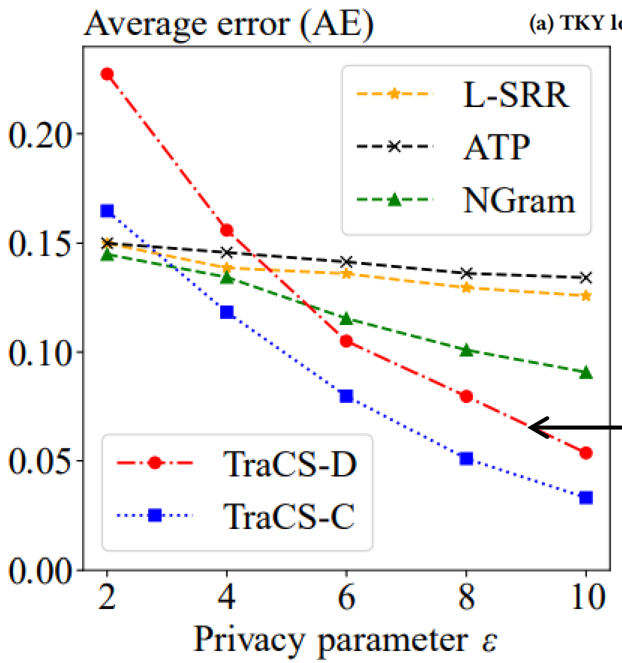
(a) TKY location space



(b) CHI location space



(a) TKY dataset



(b) CHI dataset

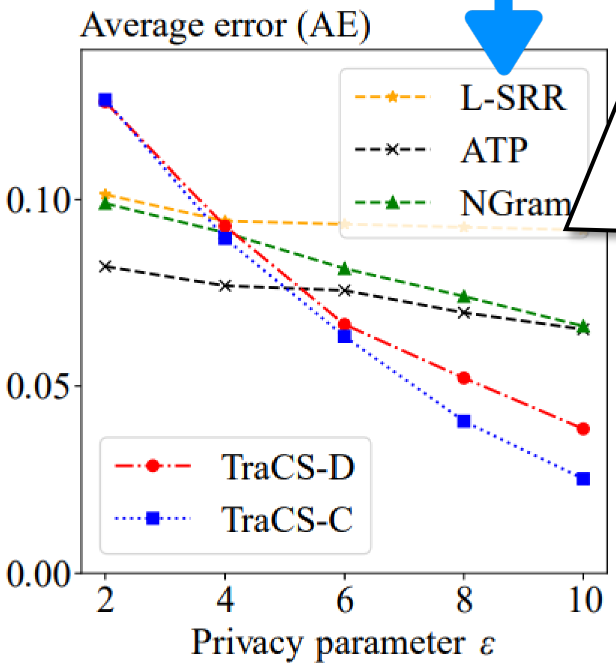
Lower error when ε becomes larger

Evaluations – Discrete Spaces

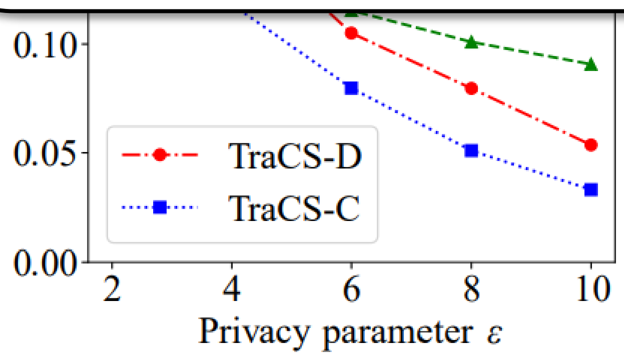
- TKY and CHI datasets
  - TraCS + rounding-to-the-nearest

Table 3: Time cost comparison (in milliseconds).
averaged

	ATP	NGram	L-SRR	TraCS-D	TraCS-C
Total	145.7	100.9	6.2	0.06	0.05
Perturb	125.8	92.8	0.003	0.018	0.003



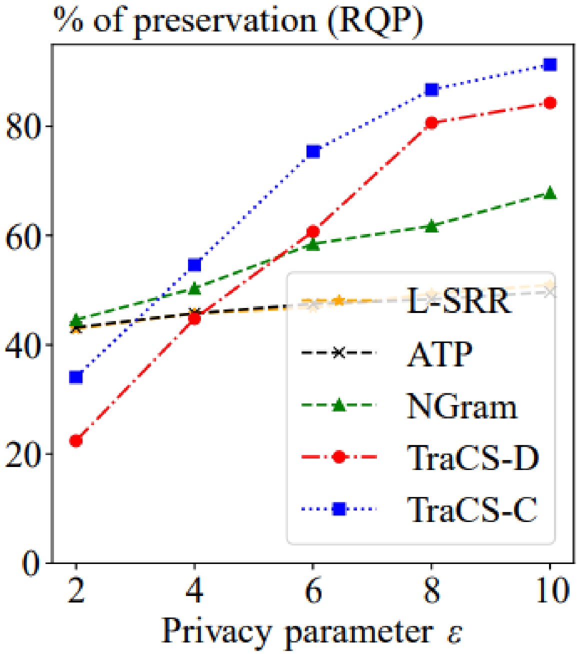
(a) TKY dataset



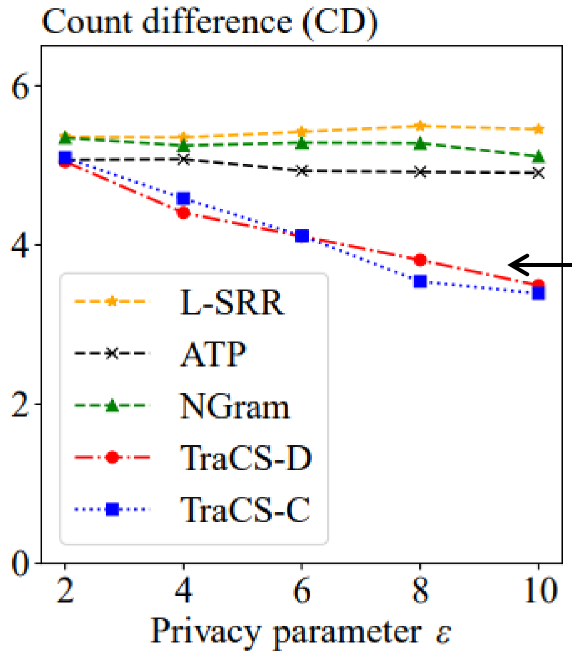
(b) CHI dataset

Evaluations – Discrete Spaces

- Other metrics: **Range query preservation, Hotspot preservation**
  - preservation of the perturbed location in a range δ ($\delta = 0.1$)
  - preservation of the most popular locations (top 20%)



(a) Range query preservation



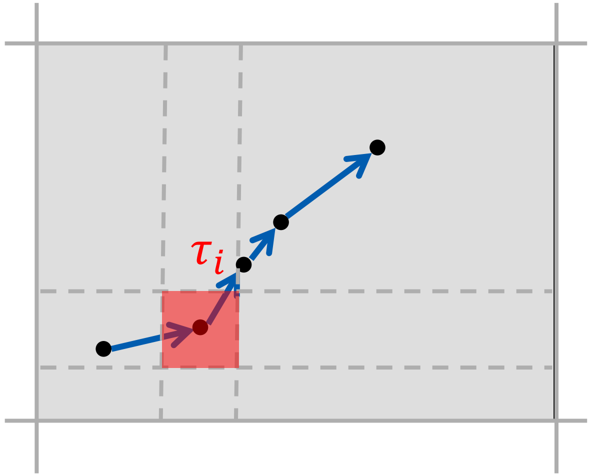
(b) Hotspot preservation

Better trajectory utility when ε becomes larger

Summary & Takeaway

- **RQ: Trajectory collection under LDP**

- Our results:
  - **operating in continuous spaces can do better**
  - better efficiency: $\mathcal{O}(1)$ sampling complexity
  - better trajectory utility, especially for larger ε
  - better applicability, for both continuous and discrete $\mathcal{S}$



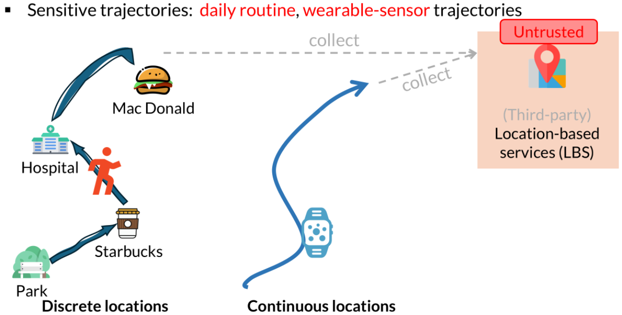
$\Pr[\mathcal{M}(x_i) = \tilde{x} \in \square] = p_\varepsilon$ (red square)

$\Pr[\mathcal{M}(x_i) = \tilde{x} \in \square] = p_\varepsilon / e^\varepsilon$ (grey square)

TraCS: Trajectory Collection in Continuous Space under LDP

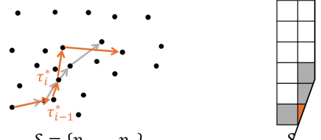
Trajectory Collection

- Sensitive trajectories: daily routine, wearable-sensor trajectories



LDP-fy a Trajectory in A **Discrete Space**

- Discrete location spaces: Point of interests or discretized ce



$\mathcal{S} = \{p_1, \ldots, p_n\}$

Limitations

1. **Efficiency**
   - each sample costs $O(n)$
2. **Trajectory Utility**
   - decreases as n increases
3. **Applicability**
   - cannot apply to continuous $\mathcal{S}$

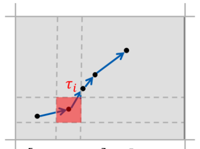
- Exponential mechanism:

$$\Pr[\mathcal{M}_{\exp}(\tau) = \tau^*] = \frac{\exp(\varepsilon d(\tau, \tau^*))}{\sum_{\tau' \in \mathcal{S}} \exp(\varepsilon d(\tau, \tau'))}$$

← Pairwise distance
← Sum of distance

Continuous Spaces: Better & Universal

- Operate on a continuous space (covering the discrete space)



$\mathcal{S} = [a_{\text{sta}}, a_{\text{end}}] \times [b_{\text{sta}}, b_{\text{end}}]$

Simple sampling: (satisfying LDP for $\mathcal{S}$)

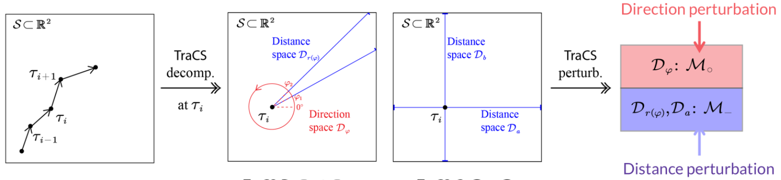
$\Pr[\mathcal{M}(x_i) = \tilde{x} \in \blacksquare] = p_\varepsilon$

$\Pr[\mathcal{M}(x_i) = \tilde{x} \in \blacksquare] = p_\varepsilon / e^\varepsilon$

- Benefits:
  1. **Efficiency:** $O(1)$ sampling complexity
  2. **Trajectory utility:** "n-independent"
  3. **Applicability:** Both continuous spaces & discrete spaces

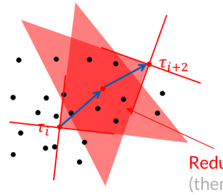
TraCS: Trajectory Collection in Continuous Spaces

- TraCS-D: direction-distance perturbation
- TraCS-C: coordinates perturbation
- **Key idea:** decomposes $\mathcal{S}$ into two subspaces → design $\mathcal{M}$ for each subspace



Evaluations

- Claims
  - continuous spaces: better utility than naïve baselines, e.g. planar Laplace + truncation
  - discrete spaces: better efficiency and utility than existing methods, ATP*, NGram**, L-SRR***
- **ATP:** direction perturbation
  1. divide direction sectors, e.g. $k = 4$
  2. perturb sector
  3. $\mathcal{M}_{\exp}(\tau)$ on reduced $\mathcal{S}$



* Trajectory Data Collection with Local Differential Privacy, VLDB'23
** Real-World Trajectory Sharing with Local Differential Privacy, VLDB'21
*** L-SRR: Local Differential Privacy for Location-Based Services with Staircase Randomized Response, CCS'22

Evaluations – Discrete Spaces

- TKY and CHI datasets
  - TraCS + rounding-to-the-nearest

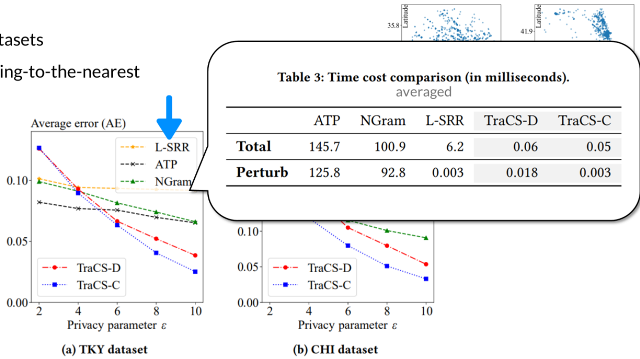


Table 3: Time cost comparison (in milliseconds).
averaged

	ATP	NGram	L-SRR	TraCS-D	TraCS-C
Total	145.7	100.9	6.2	0.06	0.05
Perturb	125.8	92.8	0.003	0.018	0.003

Thank you!