1 Introduction

Trajectory data serves as a cornerstone for numerous real-world applications, ranging from smart transportation systems to urban logistics. The rich sources of trajectory data offer profound insights into mining movement patterns and enable intelligent decision-making for large-scale users. Yet, trajectories also carry inherent privacy risks [1]. The spatiotemporal nature of trajectory data can inadvertently expose sensitive personal information, such as the data owner's home or work locations and travel routines, thereby raising significant concerns about personal privacy [2].

Local Differential Privacy (LDP) [3–6] has emerged as the de facto standard for protecting personal privacy during trajectory data collection. The first work, NGram [4], leverages an n-gram model and external public knowledge (*e.g.*, POI information) to perturb the raw trajectories of each user. Subsequently, LDPTrace [5] and PrivTC [6] have further enhanced the time efficiency of NGram by capturing global motion patterns to generate trajectories. By adopting LDP to perturb trajectories, platforms can safely collect users' data without compromising personal privacy [7].

Motivation. Although recent methods [5, 6] demonstrate better efficiency than NGram [4], we observe that *the data quality of these methods is not satisfactory* when compared to raw data. Specifically, prior studies [4–6] usually use errors of statistical queries to evaluate the quality of perturbed trajectories, but this metric may overlook structural integrity. Motivated by this, we adopt trajectory similarity, which is widely used to quantify the structural correlations between trajectories [8, 9], to assess the quality. Surprisingly, we find that the dissimilarity between perturbed trajectories by prior solutions and original ones often yields high scores, indicating significant degradation in data quality. The *main reasons* are twofold: (1) they fail to retain the trajectory backbones due to randomness of the location perturbation, and (2) most solutions still overlook the reachability between adjacent points, which may produce unrealistic trajectories. Thus, it is still challenging to *simultaneously ensure strong privacy protections and maintain high-quality perturbed trajectories*

Contribution. To address the challenge, we propose a novel framework called DQ-LDP for trajectory data collection. In DQ-LDP, we first identify the key points that form the backbone of a trajectory. Then, we perturb each key point using a distance-aware exponential mechanism while adhering to the reachability constraint. Finally, we re-construct the entire trajectory by interpolating locations between adjacent key points under LDP. To validate the performance of DQ-LDP, we conduct extensive experiments on two real-world datasets, comparing our solution against 4 state-of-the-art methods. The results demonstrate that DQ-LDP consistently generates higher-quality trajectories than these baselines under the same privacy requirement.

Roadmap. The rest of this paper is organized as follows. We present the problem definition in Section 2, and introduce our algorithm framework in Section 3. Then, we evaluate our solution and state-of-the-art methods in Section 4. We provide a comprehensive literature review in Section 5. Finally, we conclude the paper in Section 6.

2 Problem Statement

In this section, we first introduce the basic concept and then formally define the studied problem.

Formulation of Trajectory Quality. In this paper, we use trajectory similarity to measure the quality of a perturbed trajectory with respect to the original trajectory. Among the trajectory similarity metrics, we adopt Dynamic Time Warping (DTW), which has been widely used in existing work [8, 9]. Based on the definition of DTW, we propose a new concept called *similarity-aware trajectory quality*, as follows.

Definition 1 (Similarity-aware Trajectory Quality)**.** Given an original trajectory τ and the corresponding perturbed trajectory τ', the quality of τ' compared with τ is evaluated through trajectory similarity by DTW:

$$\begin{aligned}\text{Similarity}(\tau,\tau') &= \text{DTW}(\tau,\tau') \\ &= \begin{cases} d(p,q), \text{ if } \tau=\{p\} \text{ and } \tau'=\{q\} \\ \infty, \text{ if } \tau=\emptyset \text{ or } \tau'=\emptyset \\ d(p,q)+\min\begin{cases} DTW\left(\tau_h,\tau'_h\right) \\ DTW\left(\tau,\tau'_h\right) \\ DTW\left(\tau_h,\tau'\right) \end{cases} & \text{otherwise} \end{cases}\end{aligned} \tag{1}$$

where p, q represent the head points in the original trajectory τ and the perturbed trajectory τ' respectively, and τ_h, τ'_h represent the sub-trajectory excluding the head point p and q in the original trajectory τ and the perturbed trajectory τ' respectively.

Intuitively, a lower DTW value implies higher similarity (*i.e.*, lower dissimilarity) between τ and τ', indicating better quality for the perturbed trajectory τ' compared with the original trajectory τ. Now, we present our problem definition as follows.

Problem Statement. A geospatial platform aims to collect user trajectories, but users are often unwilling to share raw data with the platform due to privacy concerns. To address this, the platform designs a privacy protection mechanism $\mathcal{M}$, which perturbs each user's original trajectory τ into a new trajectory τ', meeting the following requirements:

(1) **Reachability constraint** [4]: any two adjacent points in τ' are physically reachable within in their time window. Otherwise, perturbed trajectories τ' will downgrade the downstream applications like speed estimation and map editing.

(2) **Privacy constraint** [4–6]: $\mathcal{M}$ satisfies ϵ–LDP.

(3) **Optimization goal**: to collect trajectories with good data quality, the differences (*i.e.*, dissimilarity) between τ' and τ should be as small as possible. In other words, we aim to minimize DTW(τ, τ') defined in Equation 1 under the previous constraints.

Discussion. Although perturbed trajectories are inherently expected to differ from the raw ones to preserve privacy, trajectory similarity is suitable quantify the trajectory quality due to following reasons:

(1) **It is necessary to establish a rational metric to quantify the differences between perturbed trajectory and original trajectory while under the privacy requirement.** We choose trajectory similarity as the metric, since trajectory similarity has been widely used to quantify the spatiotemporal correlations between trajectories in real-world applications and prior research. When trajectory similarity is high under the privacy requirement, it implies that the perturbed trajectory is closer to the original trajectory. This is what real-world platform usually expects for collecting trajectories from users under ϵ-LDP.

(2) **Compared to existing metric for the quality of perturbed trajectories, trajectory similarity better captures structural integrity.** Existing work [5, 6] usually uses query errors (*e.g.*, range counting query) to evaluate the quality of the perturbed trajectories. However, this existing metric mainly focuses on aggregated statistical utility but may overlook the structural integrity of trajectories. Fig. 1 shows such a simplified counterexample of trajectory query error. For both the original trajectory τ and perturbed trajectory τ', the result of the range counting query is identical, so the trajectory query error is 0. However, as indicated in the illustrations, these trajectories are inherently different due to their opposite moving directions. By contrast, the DTW between τ and τ' is 11, which reflects the low quality of the perturbed trajectory. Thus, in these scenarios, trajectory dissimilarity is a better choice than currently used query errors to quantify the quality of the perturbed trajectory.

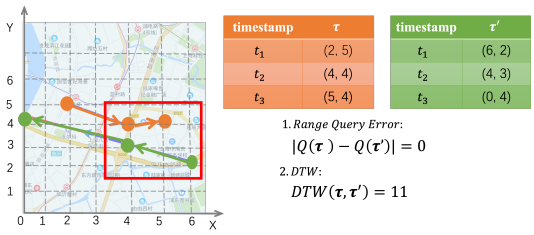


Fig. 1 Example of trajectory quality measured by query error (*i.e.*, range query error) and trajectory dissimilarity (*e.g.*, DTW)

3 Methodology

In this section, we first present our proposed framework, and then prove the privacy guarantee and error bound.

3.1 Our Framework DQ-LDP.

As shown in Fig. 2, the main workflows of our proposed framework DQ-LDP are as follows:

(1) Key Point Selection. Intuitively, inflections, origin and destination contain richer structural information than the other points in a trajectory. To capture these "key points", we define the deviation of a trajectory movement as follows:

$$\theta_i = |arctan(\overrightarrow{p_{i-1}, p_i}) - arctan(\overrightarrow{p_{i-1}, p_{i+1}})| \tag{2}$$

Here, a trajectory τ is defined as a sequence of spatiotemporal points, *i.e.*, $\tau = \{(p_1, t_1), \cdots, (p_{|\tau|}, t_{|\tau|})\}$, where $p_i \in \mathcal{P}$ denotes the location at the timestamp t_i, and $\mathcal{P}$ denotes the points set. Accordingly, we use $\sin\theta_i$ to measure the importance of each point and perform a Top-k selection to identify sufficient key points, where k is a user-defined parameter ($k = 0.6|\tau|$ in our real-world datasets). Besides, the Top-k selection procedure is protected via exponential mechanism [3] with an allocated privacy budget 0.5ϵ.

(2) Key Point Perturbation. To perturb key points while maintaining reachability, we need to address two main technical issues: *(i) the total privacy budget allocation* and *(ii) the obfuscation mechanism under LDP*.

For the *first issue*, prior work divides the privacy budget uniformly across all spatial points. By contrast, since key points form the backbone of trajectories, we allocate the remaining privacy budget to key points and leverage post-processing property of LDP [3, 4] to generate non-key points.

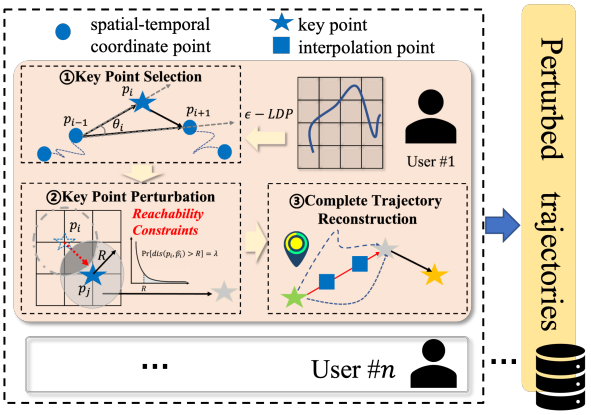


Fig. 2 Overview of our framework DQ-LDP

Regarding the *second issue*, we introduce a distance-aware exponential mechanism. Specifically, we use the distance d to measure the utility q (*i.e.*, similarity) between the original point p_k and the perturbed point p'_k, *i.e.*, $q = -d(p_k, p'_k)$. To optimize overall efficiency, we only consider candidate points $\mathcal{W}(p_k)$ in Eq. (3) that are located within the reachable regions of p_k. Here, Δ_q in Eq. (4) is the sensitivity of the utility. We provide the parameter λ to allow users to adjust the candidate set, if λ not be set, $\mathcal{W}(p_k)$ is $\mathcal{P}$.

$$\mathcal{W}(p_k) = \left\{ p'_k \mid d(p_k, p'_k) \leq \frac{2\Delta_q}{\epsilon} \cdot \ln\left(\frac{1}{\lambda}\right) \right\} \tag{3}$$

$$\Delta_q = \max_{p,p'} \left| d(p_k, p) - d(p_k, p') \right| \tag{4}$$

Eq. (3) guarantees that the cumulative probability of a perturbed point being farther than $\frac{2\Delta_q}{\epsilon} \ln \frac{1}{\lambda}$ from the original point is bounded by λ. Moreover, the reachability constraint between the previous key point p_{k-1} and the current key point p_k further refines the search space as follows:

$$\mathcal{W}^*(p_k) = \left\{ p'_k \in \mathcal{W}(p_k) \mid d(p_{k-1}, p'_k) \leq \text{speed} \cdot (t_k - t_{k-1}) \right\}$$

Finally, we calculate the appearance probability of each perturbed point using the exponential mechanism:

$$\Pr\left(p'_k = p'\right) = \frac{\exp\left(-\epsilon d\left(p_k, p'\right)/2\Delta_q\right)}{\sum_{p \in \mathcal{W}^*(p_k)} \exp\left(-\epsilon d\left(p_k, p\right)/2\Delta_q\right)} \tag{5}$$

(3) Complete Trajectory Reconstruction. We reconstruct the entire trajectory by interpolating additional points between perturbed key points, guided by the GPS sampling rates.

3.2 Theoretical Analysis.

Theorem 1 presents the main result of privacy guarantees.

Theorem 1. Our algorithm DQ-LDP satisfies ϵ–LDP.

Proof. The process of Top-k selecting in i−th iteration can be regarded as a LDP mechanism $\mathcal{M}_i$. We allocate privacy budget ϵ_1/k for each mechanism $\mathcal{M}_i$. Because each key point comes from the same user and the composition theorem of local differential privacy, the mechanism $\mathcal{M} = \{\mathcal{M}_i | i = 1, ..., k\}$ after k iteration satisfies $(\sum_{i=1}^{k} \epsilon_1/k)$−LDP. Similarly, when $\mathcal{W}(p_k)$ is $\mathcal{P}$, the mechanism $\mathcal{E} = \{\mathcal{E}_i | i = 1, ..., k\}$ used in key point perturbation can be regarded as one mechanism which excludes all spatiotemporal points that do not satisfy the reachability constraint with privacy budget ϵ_2. The mechanism $\mathcal{E}$ satisfies $(\sum_{i=1}^{k} \epsilon_2/k)$−LDP. *Complete Trajectory Reconstruction* is only a post-processing step and does not affect privacy. Since $\epsilon_1 = \epsilon_2 = \epsilon/2$, the total budget is $\sum_{i=1}^{k}(\epsilon_1/k + \epsilon_2/k) = \epsilon$. Thus, the DQ-LDP method satisfies ϵ–LDP. □

Error Bound. Theorem 2 derives the error bound.

Theorem 2. Given parameter λ, our solution DQ-LDP ensures that the distance between the perturbed point and the actual one doesn't exceed $\frac{2\Delta_q}{\epsilon} \ln \frac{1}{\lambda}$.

Proof. The exponential mechanism assigns probabilities to candidate coordinates, and the distance r between the probability density function and the sensitive value has a negative exponential relationship:

$$\Pr(r) \propto \exp(-\frac{\epsilon r}{2\Delta_q}) \tag{6}$$

Based on the normalization conditions, the normalization constant C is determined by solving:

$$\Pr(r \geq 0) = \int_0^{\infty} C \cdot \exp\left(-\frac{\epsilon \cdot r}{2\Delta q}\right) dr = 1 \tag{7}$$

Since $C = \frac{\epsilon}{2\Delta_q}$, the probability that r is greater than d is as follows:

$$\Pr(r{>}d) = \int_d^{\infty} \frac{\epsilon}{2\Delta_q} \cdot \exp\left(-\frac{\epsilon \cdot r}{2\Delta q}\right) dr = \lambda \tag{8}$$

After solving the integral, the maximum error between the perturbed coordinate and the actual one is $d = \frac{2\Delta_q}{\epsilon} \ln \frac{1}{\lambda}$. □

Discussion. Our DQ-LDP selects k key points based on their importance to the trajectory backbone. The Top-k selection procedure is allocated a privacy budget of 0.5ϵ to ensure that data access to all points is differentially private. The key point perturbation procedure allocates the remaining privacy budget 0.5ϵ to all key points. Thus, each key point is perturbed with a privacy budget of $\frac{0.5\epsilon}{k}$, indicating a stronger

level of privacy preservation for each point compared to that for the entire trajectory.

As for the non-key points, there are $|\tau| - k$ in total, and our DQ-LDP considers them to be much less significant (than key points) for the trajectory backbone. We leverage the post-processing property of LDP [10] to protect their privacy. Specifically, we generate non-key points by interpolating certain points between perturbed key points, guided by the GPS sampling rates. As a result, we can reconstruct the entire trajectory without allocating additional privacy budget to the non-key points.

4 Experimental Evaluation

This section presents our experimental studies. Specifically, we first introduce the experimental setup. Then, we present the overall performance and the evaluation on privacy protection level. Finally, we conduct scalability tests and abalation studies.

4.1 Experimental Setup

Dataset. Our evaluation uses two real-world datasets for conventional tests: *Dazhong* and *Geolife*. *Dazhong* contains moving trajectories of 13,012 vehicles and *Geolife* comprises trajectories from 182 users. *Geolife* is a public trajectory dataset. The *Dazhong* dataset, which is a private dataset, contains moving trajectories of 13,012 Dazhong-brand personal cars. This real-world dataset was collected from 2015 to 2017. Overall, this dataset contains more than 195,516,555 spatial locations and forms 115,349 trajectories, where the average length of these trajectories is 1,695. All the spatial data fall within the geographic spatial range of $29.0°N \sim 32.9°N$ and $119.0°E \sim 122.4°E$. In contrast to the *Geolife* dataset, which primarily consists of human trajectories, the *Dazhong* dataset focuses on vehicle trajectories.

We have also prepared two more datasets for scalability tests, named *Chengdu* and *Xian*. They are open-sourced by Didi Chuxing [11]. Both datasets contain massive trajectories from taxis in Didi Chuxing during the ride-hailing service.

Overall, the detailed statistics of all four datasets are summarized in Table 1. These datasets collectively cover diverse application scopes and parameter settings.

Baseline. We compared our DQ-LDP with 4 existing methods: EM [3], NGram [4], LDPTrace [5], and PrivTC [6]). In the experiments, all the algorithms adopt the same privacy budget ϵ, and ϵ is varied from 0.01 to 5 to assess the impact of the privacy preservation level for each compared algorithm. As for other parameters, we use their recommended configurations:

Table 1 The statistics of four real datasets in our evaluation

Dataset	#(trajectory)	Spatial Region	Average Length
Dazhong	115,349	$29.0°N \sim 32.9°N$ $119.0°E \sim 122.4°E$	1,695
Geolife	13,595	$39.9°N \sim 40.1°N$ $116.1°E \sim 116.5°E$	713
Chengdu	207,107	$30.6°N \sim 30.7°N$ $104.0°E \sim 104.2°E$	201
Xian	119,018	$34.2°N \sim 34.3°N$ $108.9°E \sim 109.0°E$	238

1. EM [3]: we set the utility function of the exponential mechanism perturbation is consistent with the **Key Point Perturbation** of DQ-LDP in the manuscript.
2. NGram [4]: we set $n = 2$ for n-gram model and choose the speed in the *Geolife* the dataset is 8 km/h, while the speed in other datasets is 50 km/h.
3. LDPTrace [5]: we set $\alpha = 0.3, \beta = 0.2$ for the reweighting function used in LDPTrace.
4. PrivTC [6]: we set $\sigma = 0.2, \alpha = 0.02$ for grid construction in this PrivTC.

Metric. To assess the quality of perturbed trajectories, we use Dynamic Time Warping (DTW) in Eq. (1) to measure trajectory similarity relative to the raw data. Besides, we evaluate the average processing time required for each trajectory.

4.2 Main Result

We investigate the impact of different trajectory lengths $|\tau|$ and privacy budgets ϵ on both datasets. We ignore the results of NGram that take over 24 hours.

- **Impact of Trajectory Length** $|\tau|$. As shown in Fig. 3, our DQ-LDP always achieves the lowest DTW, indicating the minimal dis-similarity between perturbed trajectories and the raw data. For example, on the *Dazhong* dataset, the DTW of DQ-LDP is up to 13.7× lower than that of compared baselines. Meanwhile, DQ-LDP is faster than most baselines.
- **Impact of Privacy Budget** ϵ. we can observe that the DTW of our DQ-LDP outperforms the compared baselines regardless of the privacy budget, as shown in Fig. 4. The gap of DTW remains over 20.4× on the *Geolife* dataset. It implies that the data quality of DQ-

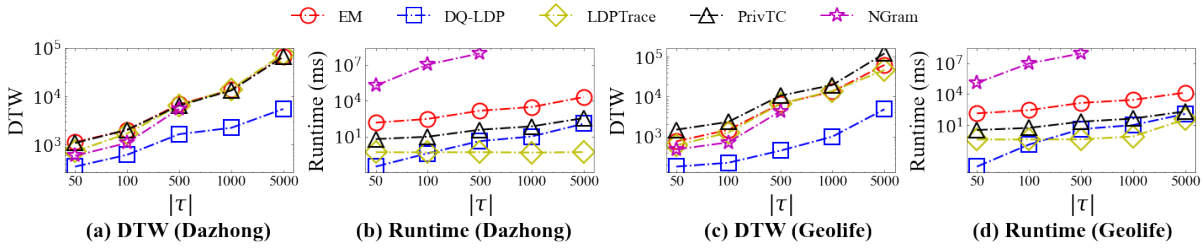


Fig. 3 Impact of trajectory lengths $|\tau|$ on the *Dazhong* and *Geolife* datasets

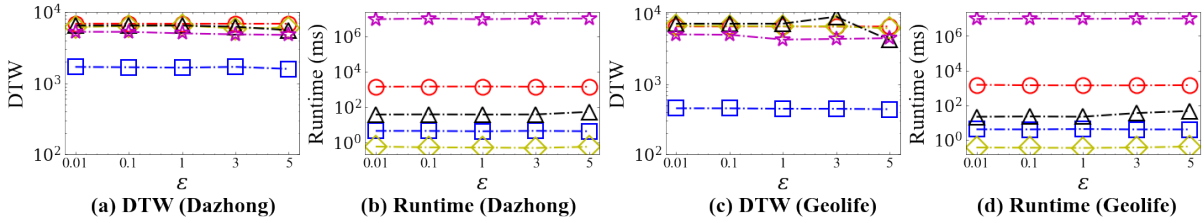


Fig. 4 Impact of total privacy budget ϵ on the *Dazhong* and *Geolife* datasets

LDP is superior to that of the other methods. DQ-LDP is also more efficient than EM and PrivTC.

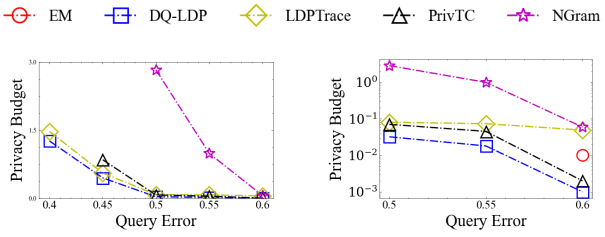
4.3 Evaluation on Privacy Protection Level

Since all existing baselines and our solution adopt the privacy budget ϵ of local differential privacy to reflect the privacy protection level, this new experiment fixes the query error over perturbed trajectories and assesses the values of privacy budgets ϵ. Under the same query error, if an algorithm obtains a smaller budget ϵ than others, it implies that this algorithm has the most stringent privacy preservation level.

Fig. 5 shows the results of such an experiment on *Dazhong* dataset. We can observe that DQ-LDP consistently achieves lower ϵ values than other baseline methods under the same query error, indicating stronger privacy protection. For instance, when the (average) query error is 0.4, DQ-LDP requires $\epsilon = 1.259$, while the best baseline, LDPTrace, requires $\epsilon = 1.468$. In other words, the privacy protection level has been improved by 14.2% compared to LDPTrace. Moreover, other baseline methods fail to achieve a query error of 0.4 within a reasonable setting for ϵ (*e.g.*, ≤ 10). When the query error varies between 0.5 and 0.6, Fig. 5(b) illustrates the detailed comparison results between our DQ-LDP and existing baselines. The result indicates that DQ-LDP still obtains the best privacy protection under the same query error.

Beyond the evaluations in Fig. 5, we have also tried to compare the privacy budget under the same trajectory quality (*i.e.*, the same DTW). However, even when the privacy budget ϵ increases to 20, the DTW of our DQ-LDP is still better than that of existing baselines. In other words, we cannot fix the DTW value under a rational setting for ϵ and plot a meaningful curve between DTW and ϵ. Thus, we ignore the result of this experiment in the revision.

In summary, at the same level of trajectory query errors, our DQ-LDP always consumes a smaller privacy budget, indicating its capability to preserve trajectory privacy more rigorously.



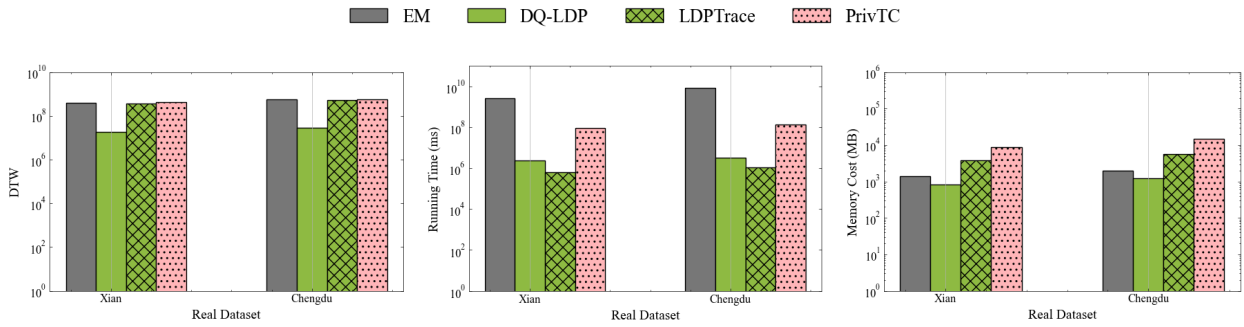
(a) Result of privacy budget ϵ when query error increases from 0.4 to 0.6

(b) Result of privacy budget ϵ when query error increases from 0.5 to 0.6

Fig. 5 Experiment on the privacy preservation level (*i.e.*, privacy budget ϵ) when fixing query errors

4.4 Scalability Test

The scalability tests are conducted on two large-scale datasets of taxi trajectories: *Xian* and *Chengdu*. We evaluate the algorithm performance from three perspectives: trajectory quality (which is measured with total DTW in Definition 1), running time, and memory overhead. The experimental results are presented in Fig. 6.



(a) Trajectory quality (total DTW) (b) Running time (c) Memory cost

Fig. 6 scalability test on real-world datasets

Result of trajectory quality. Regarding trajectory quality, our algorithm DQ-LDP achieves the best results (*i.e.*, lowest total DTW) across all datasets in Fig. 6(a). On both datasets, the total DTW is reduced by up to 95% compared to existing baselines. Among the baselines, PrivTC tends to obtain higher total DTW than others, indicating lower quality for the perturbed trajectories. The results of EM and LDPTrace are comparably close.

Result of running time. In terms of running time, the baseline LDPTrace is the most efficient, while our algorithm DQ-LDP is consistently the runner-up. The average gaps of each trajectory between LDPTrace and DQ-LDP are 14.08 and 10.6 milliseconds on the *Xian* and *Chengdu* datasets, respectively. Considering the application needs and improvements in trajectory quality, the gap is acceptable. Moreover, DQ-LDP is up to 40.13 × faster than PrivTC. Compared with the EM, the performance improvement is more significant.

Result of memory usage. As shown in Fig. 6(c), our algorithm DQ-LDP requires the least space consumption among the compared solutions. This is because DQ-LDP leverage key point selection to concentrate on relevant locations in the trajectory. By contrast, existing baselines [5, 6] require additional storage for Markov transition matrices, which introduces extra memory overhead. For example, on the *Xian* dataset, our DQ-LDP saves over an order of magnitude in memory overhead compared to the baseline PrivTC.

Overall, the above demonstrates that our solution can handle large-scale datasets more effectively than all existing baselines, with lower running time and memory usage than most of them.

4.5 Ablation Study

The ablation study aims to demonstrate that our Top-k key selection strategy is effective in enhancing the quality of perturbed trajectories without sacrificing time efficiency. Fig. 7 shows the experimental results of this ablation study on the *Geolife* dataset, when varying the trajectory length $|\tau|$ and privacy budget ϵ.

Impact on Effectiveness. DTW is used to measure the perturbed trajectory quality, which reflects the effectiveness of the compared algorithms. As shown in Fig. 7, our DQ-LDP consistently outperforms DQ-LDP-Random in terms of DTW across different trajectory lengths and privacy budgets. For example, when $|\tau| = 5000$, the DTW for DQ-LDP is 60.8% lower than that of DQ-LDP-Random. When $\epsilon = 0.01$, DQ-LDP reduces DTW by 62.8% compared to DQ-LDP-Random. The comparisons indicate that our Top-k selection can effectively improve the perturbed trajectory quality compared to random selection.

Impact on Efficiency. Meanwhile, we can also observe that both DQ-LDP and DQ-LDP-Random consume almost the same running time when varying the trajectory length and privacy budget. This is because the difference in time cost between Top-k selection and random selection is minor. The comparison here indicates that the improvement in terms of effectiveness does not degrade the overall time efficiency.

5 Related Work

Differential Privacy (DP) has become the de facto standard for ensuring trajectory privacy. Early approaches [12–17] assumed the presence of trusted central servers to collectively aggregate trajectories before applying perturbation with DP noise. While these methods demonstrated superior performance, they rely on a critical assumption: that centralized

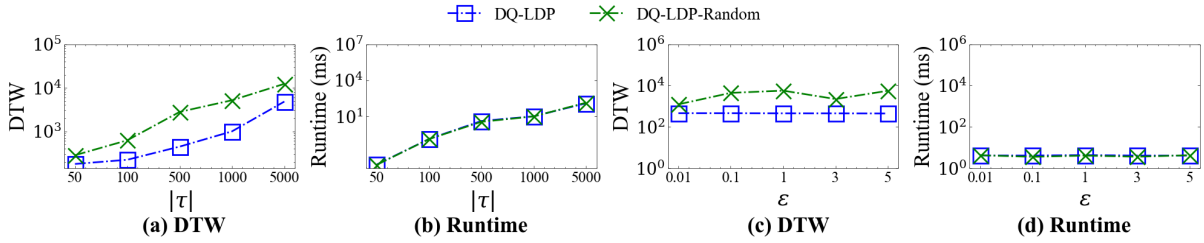


Fig. 7 Results of the ablation study of our Top-k key selection strategy

data curators are fully trusted. This reliance on trusted central servers poses significant challenges in real-world applications, where trust in central authorities may be limited or absent.

Consequently, in recent years, there has been a growing need for designing Local Differential Privacy (LDP) based mechanisms for collecting trajectories without this strong trust assumption. In contrast to central DP, LDP enables privacy-preserving trajectory collection directly on the user side. Existing location perturbation methods [18–20] could be used to protect the privacy of a location sequence while satisfying ϵ-Geo-indistinguishability. However, these methods may overlook the correlations between adjacent locations.

Related studies have focused on protecting the privacy of the entire trajectory and improving its utility by considering the correlations. In general, these approaches can be broadly categorized into *point perturbation methods* [4, 21] and *synthesis-based methods* [5, 6].

- Point perturbation methods preserve privacy by injecting noise directly into individual spatial-temporal points. The seminal work, NGram [4], leverages an n-gram model and external public knowledge (*e.g.*, POI information) to perturb the raw trajectories of each user. ATP [21] leverages direction information with trajectory perturbation in LDP. We did not compare with ATP in the experiments, because it has a different problem formulation from our paper. For example, ATP does not support the reachability constraint, so it is hard to compare under a relatively fair condition.
- Synthesis-based methods generate synthetic trajectories that statistically mimic real data distributions. For example, PrivTC [6] uses the Hidden Markov Model (HMM) to improve the utility of perturbed trajectories. LDPTrace [5] further enhances this type of solution by capturing more spatial patterns (*e.g.*, start/end points, intra-trajectory transitions, and length distributions) through a Markov-based synthesizer.

6 Conclusion

This paper focuses on privacy protection when collecting trajectories from multiple data sources under Local Differential Privacy (LDP). While existing solutions are efficient, they often result in unsatisfactory trajectory data quality. To address this, we propose a novel framework called DQ-LDP that captures key points in trajectories for perturbation. We also prove that DQ-LDP satisfies ϵ-LDP and introduces only moderate discrepancies compared to raw data. Moreover, extensive experiments demonstrate that our DQ-LDP is more effective than existing methods with a competitive efficiency.

References

1. Sun N, Wang W, Tong Y, Liu K. Blockchain based federated learning for intrusion detection for internet of things. Frontiers Comput. Sci., 2024, 18(5): 185328
2. Wang N, Zheng W, Wang Z, Wei Z, Gu Y, Tang P, Yu G. Collecting and analyzing key-value data under shuffled differential privacy. Frontiers Comput. Sci., 2023, 17(2): 172606
3. Li N, Lyu M, Su D, Yang W. Differential Privacy: From Theory to Practice. Morgan & Claypool Publishers, 2016
4. Cunningham T, Cormode G, Ferhatosmanoglu H, Srivastava D. Real-World Trajectory Sharing with Local Differential Privacy. PVLDB, 2021, 14(11): 2283–2295
5. Du Y, Hu Y, Zhang Z, Fang Z, Chen L, Zheng B, Gao Y. LDPTrace: Locally Differentially Private Trajectory Synthesis. PVLDB, 2023, 16(8): 1897–1909
6. Yang J, Cheng X, Su S, Sun H, Chen C. Collecting Individual Trajectories under Local Differential Privacy. In: MDM. 2022, 99–108
7. Liu F, Zheng Z, Shi Y, Tong Y, Zhang Y. A survey on federated learning: a perspective from multi-party computation. Frontiers Comput. Sci., 2024, 18(1): 181336
8. Su H, Liu S, Zheng B, Zhou X, Zheng K. A survey of trajectory distance measures and performance evaluation. VLDB J., 2020, 29(1): 3–32
9. Wang S, Bao Z, Culpepper J S, Cong G. A Survey on Trajectory Data Management, Analytics, and Learning. ACM Comput. Surv., 2022, 54(2): 39:1–39:36

10. Cormode G, Jha S, Kulkarni T, Li N, Wang T. Privacy at scale: Local differential privacy in practice. In: ACM SIGMOD International Conference on Management of Data. 2018, 1655–1658
11. Didi chuxing's open dataset. https://outreach.didichuxing.com/
12. Bonomi L, Xiong L. A two-phase algorithm for mining sequential patterns with differential privacy. In: 22nd ACM International Conference on Information and Knowledge Management, CIKM'13, San Francisco, CA, USA, October 27 - November 1, 2013. 2013, 269–278
13. Jin F, Hua W, Ruan B, Zhou X. Frequency-based randomization for guaranteeing differential privacy in spatial trajectories. In: 38th IEEE International Conference on Data Engineering, ICDE 2022, Kuala Lumpur, Malaysia, May 9-12, 2022. 2022, 1727–1739
14. He X, Cormode G, Machanavajjhala A, Procopiuc C M, Srivastava D. DPT: differentially private trajectory synthesis using hierarchical reference systems. Proc. VLDB Endow., 2015, 8(11): 1154–1165
15. Bindschaedler V, Shokri R. Synthesizing plausible privacy-preserving location traces. In: 2016 IEEE Symposium on Security and Privacy (SP). 2016, 546–563
16. Gursoy M E, Liu L, Truex S, Yu L. Differentially private and utility preserving publication of trajectory data. IEEE Trans. Mob. Comput., 2019, 18(10): 2315–2329
17. Gursoy M E, Liu L, Truex S, Yu L, Wei W. Utility-aware synthesis of differentially private and attack-resilient location traces. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 2018, Toronto, ON, Canada, October 15-19, 2018. 2018, 196–211
18. Andrés M E, Bordenabe N E, Chatzikokolakis K, Palamidessi C. Geo-indistinguishability: differential privacy for location-based systems. In: Proceedings of 2013 ACM SIGSAC Conference on Computer and Communications Security (CCS). 2013, 901–914
19. Weggenmann B, Kerschbaum F. Differential privacy for directional data. In: CCS '21: 2021 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, Republic of Korea, November 15 - 19, 2021. 2021, 1205–1222
20. Zhao Y, Yuan D, Du J T, Chen J. Geo-ellipse-indistinguishability: Community-aware location privacy protection for directional distribution. IEEE Trans. Knowl. Data Eng., 2023, 35(7): 6957–6967
21. Zhang Y, Ye Q, Chen R, Hu H, Han Q. Trajectory data collection with local differential privacy. Proc. VLDB Endow., 2023, 16(10): 2591–2604