EAI Endorsed Transactions
on Security and Safety
Research Article EAI.EU

A Hybrid Differential Privacy and k-Anonymity Framework for Enhancing Location Privacy in Location-Based Services

Gagandeep Singh*, Ankita Gupta

CT Institute of Engineering, Management & Technology, Jalandhar, India

Abstract

The growing reliance on Location-Based Services (LBS) has intensified privacy risks, as the continuous collection of sensitive user location data exposes individuals to potential re-identification and unauthorised tracking. This paper presents a hybrid privacy-preserving framework that combines the Diameter-Bounded DBSCAN clustering algorithm for spatial k-anonymity with an adaptive Laplace mechanism for ε-differential privacy. This integration ensures the formation of compact anonymity groups while maintaining high data utility. Experimental evaluation on the real-world GeoLife dataset demonstrates 85.1% query accuracy, 0.14 trajectory distortion (EDR), and average query latency below 100 milliseconds for 20,000 users, outperforming DPPS and AdaptiveGrid baselines. Comprehensive sensitivity analysis of the diameter threshold (dmax) and evaluation of suppression bias confirm the framework's robustness, scalability, and practical suitability for real-time LBS deployment.

Keywords: Location-Based Services, Differential Privacy, k-Anonymity, Laplace Mechanism, Geo-Privacy, Smart City Analytics

Received on 02 August 2025, accepted on 29 January 2026, published on 03 February 2026

doi: 10.4108/eetss.9845

1. Introduction

The proliferation of smartphones and IoT devices has fueled an explosive growth in **Location-Based Services (LBS),** which provide users with personalised, context-aware experiences ranging from navigation and social networking to targeted advertising and urban planning. The efficacy of these services, however, is predicated on the collection and analysis of vast amounts of sensitive user location data. This practice creates a significant privacy risk, as raw or inadequately protected location data can be exploited for unauthorised tracking, behavioural profiling, and even re-identification poses a direct threat to user safety and autonomy.

To address this challenge, a variety of **Privacy-Preserving Mechanisms (PPMs)** have been proposed, broadly falling into two categories: k-anonymity and differential privacy (DP). K-anonymity and its variants offer intuitive, group-based privacy by ensuring that any individual is indistinguishable from at least k-1 others. However, these methods often lack formal privacy guarantees. Differential privacy, conversely, provides a strong, mathematically rigorous framework for privacy but can significantly degrade data utility if not carefully implemented, particularly in the spatial domain.

While recent research has focused on developing hybrid models that combine these approaches, many existing solutions suffer from two key drawbacks: they employ

*Corresponding author. Email: gagandeep.singh.1290j@gmail.com

clustering algorithms (like standard DBSCAN) that are **methodologically flawed for ensuring robust k-anonymity,** or they introduce significant computational complexity, limiting their practical use in real-time systems.

This paper bridges this gap by introducing **a practical and methodologically sound hybrid privacy framework.** While the goal of creating anonymous groups that are both of a minimum size *k* and geographically compact is related to the well-studied **r-gather clustering problem**, many existing solutions rely on complex or centralised algorithms unsuitable for real-time LBS. To bridge this gap, our primary contribution is a **practical and scalable algorithm** that uses an efficient, DBSCAN-based approach to generate these compact anonymous groups. This method, which we term **Diameter-Bounded DBSCAN**, provides a robust spatial k-anonymity guarantee and is designed for seamless integration into a hybrid privacy framework. This is integrated with an adaptive differential privacy mechanism to provide layered, robust protection.

We validate our framework through a comprehensive experimental evaluation on the real-world **GeoLife trajectory dataset.** Our results demonstrate that our approach not only provides strong privacy guarantees but also consistently **outperforms state-of-the-art baselines** in preserving data utility for complex analytical tasks, such as hotspot detection and trajectory analysis. The framework is scalable, efficient, and well-suited for real-world LBS deployment.

2. Related Work

The proliferation of Location-Based Services (LBS) has been paralleled by the evolution of privacy-preserving mechanisms designed to protect sensitive spatiotemporal data. The research landscape is rich, progressing from foundational anonymisation and cryptographic techniques to the robust mathematical guarantees of differential privacy and, more recently, to sophisticated adaptive and hybrid models. This section provides a structured review of this evolution, contextualising our work within the current state-of-the-art.

2.1 Foundational Privacy Models: Anonymity and Differential Privacy

Early efforts in data privacy centred on anonymisation techniques. The seminal concept of **k-anonymity**, introduced by Sweeney [1] requires that any individual's record in a released dataset be indistinguishable from at least k-1 other records based on their quasi-identifiers. While effective against basic re-identification, k-anonymity and its extensions (e.g., l-diversity and t-closeness) can be vulnerable to inference attacks and may degrade data utility, particularly in high-dimensional datasets.

A paradigm shift occurred with the introduction of **Differential Privacy (DP)** by Dwork et al. [2] et al., which offers a formal, mathematical guarantee of privacy. DP ensures that the output of a computation is statistically insensitive to the presence or absence of any single individual's data. This is typically achieved by adding calibrated noise—often from a Laplace distribution—to a query's result, with the privacy-utility trade-off controlled by a parameter,
ϵ (the privacy budget). A smaller ϵ provides stronger privacy but introduces more noise, and vice-versa. Together, these foundational models shaped the core trade-offs between privacy, utility, and scalability in spatial data publishing.

2.2 Adaptation of Privacy Models for Location Data

The unique geometric nature of location data necessitated specialised adaptations of these foundational models. While foundational models like k-anonymity and geo-indistinguishability laid the groundwork for spatial privacy, their limitations spurred a new generation of hybrid and adaptive frameworks.

- **Spatial Cloaking for k-Anonymity:** The principles of k-anonymity were first adapted to LBS through **spatial cloaking**. The foundational work by Gruteser and Grunwald [3] proposed methods to generalise a user's precise coordinates into a broader "cloaking region" that includes at least
  k users, thereby providing anonymity within that spatial area. This approach forms the basis for many anonymisation strategies in LBS.

- **Geo-Indistinguishability for Differential Privacy:** Geo-indistinguishability extends DP to spatial domains by ensuring indistinguishability decreases with distance—offering finer control over location privacy. To apply DP's rigorous guarantees to spatial data, Andrés et al. [4] introduced **geo-indistinguishability.** This model typically adds two-dimensional Laplace noise to a user's coordinates. A key feature is that the privacy guarantee is a function of distance; it becomes more difficult to distinguish between two nearby points than two distant points, formally capturing the

intuition that approximate location information is less sensitive than precise location information.

2.3 The Rise of Hybrid and Adaptive Frameworks

While foundational methods provide essential building blocks, they have inherent limitations. K-anonymity lacks formal privacy guarantees, and standard DP can excessively degrade utility. Consequently, a significant trend in modern research (2023–2025) is the development of hybrid and adaptive frameworks that combine multiple techniques to achieve a better balance of privacy, utility, and scalability.

- **Hybrid "Cluster-then-Perturb" Models:** A common hybrid strategy involves first using a clustering algorithm to form anonymous groups and then applying a DP mechanism to the aggregated data. This approach is conceptually similar to our own. For instance, **Wang et al.** [5] propose a location-clustering algorithm followed by the addition of Laplace noise to cluster centroids, demonstrating a practical application of this hybrid pattern.

- **Advanced Hybrid Schemes:** The field has produced increasingly sophisticated hybrid models. The **DPPS** scheme by **Li et al.** [6] combines an advanced k-anonymity algorithm with a Hidden Markov Model to protect against correlation attacks in continuous LBS. Similarly,

LPPS-IKHC by **Li et al.** [7] integrates an improved k-anonymity approach with a hybrid cache mechanism for the Internet of Vehicles (IoV). These works highlight that simply combining techniques is not novel in itself; the innovation lies in
how they are combined to address specific threat models like trajectory correlation.

- Adaptive Privacy Budgeting: Another major research thrust is making privacy mechanisms adaptive. The idea of allocating the privacy budget ϵ based on data density is a well-established heuristic for improving the privacy-utility trade-off. Recent work has formalised this in various ways. Kim [8]introduces a method for adaptive grid partitioning in real-time during data collection, directly capturing user distribution to enhance utility.

  Ma et al. [9] also propose a framework using a density- and distance-aware adaptive grid structure to satisfy DP. These adaptive grid methods are functionally similar to adaptive noise mechanisms; one adjusts spatial resolution while the other adjusts noise levels to achieve the same goal of applying stronger privacy in sparser, more sensitive areas.

Hybrid and adaptive methods illustrate a broader trend: tailored privacy protections outperform monolithic approaches, especially in LBS environments.

Connection to r-Gather and Constrained Clustering Approaches
Works on r-gather and constrained-diameter clustering [10,11] investigate grouping with bounded diameter under anonymity constraints. These approaches aim to minimize the maximum intra-cluster distance while ensuring each cluster contains at least k records. Unlike these methods, the proposed framework integrates density-aware DBSCAN initialization with an adaptive differential-privacy layer, achieving practical scalability and tunable privacy–utility trade-offs for real-world LBS datasets.

2.4 Advanced Topics and Future Directions

The frontier of location privacy research is also pushing into more complex areas, including:

- **Trajectory and Semantic Privacy:** Protecting an individual's entire movement pattern (trajectory) is significantly more challenging than protecting a single point. A rich body of work, surveyed by **Jin et al.** [12], focuses on privacy-preserving trajectory data publishing. Furthermore, researchers are increasingly focused on protecting.

  Location semantics—the meaning or sensitivity of a place (e.g., a hospital vs. a coffee shop)—which is often more revealing than coordinates alone. **Yan et al.** [13], for example, propose methods specifically for preserving location semantic privacy.

- **Local and Shuffled Differential Privacy:** To remove the need for a trusted central data aggregator, **Local Differential Privacy (LDP)** perturbs data on the user's device before collection. While offering stronger trust assumptions, LDP often requires significantly more data to maintain utility. A promising middle ground is the

A shuffled model of DP, where an intermediary shuffler anonymises user reports before they reach the aggregator, providing stronger privacy than centralised DP with better utility than LDP.

These advancements push the frontier of privacy-preserving analytics but often introduce significant implementation overhead, reinforcing the need for practical, deployable alternatives.

2.5 Research Gap and Motivation

This review demonstrates that the field of location privacy is dynamic and rapidly advancing. While hybrid models combining k-anonymity and DP are established, and the principle of adaptive, density-based privacy is well-known, a gap remains in the rigorous evaluation of practical, computationally efficient frameworks that integrate these ideas. While effective, many schemes rely on complex models (e.g., HMMs, caching), limiting their practical adoption.
Thus, we introduce a hybrid model—DBSCAN clustering combined with adaptive Laplace-based DP noise injection—that is simple, interpretable, and well-suited for deployment in real-time LBS platforms. Unlike prior approaches that rely on grid partitioning or trajectory modelling, our method offers a robust privacy-utility tradeoff while maintaining scalability and algorithmic clarity.

3. Problem Formulation

The increasing reliance on Location-Based Services (LBS) introduces critical privacy risks, as user trajectories and real-time geographic data are often collected, stored, and queried without sufficient safeguards. This section outlines the core privacy challenges, describes the system and adversary models, and formulates the problem addressed by this study.

3.1 System Model

We consider a four-component LBS architecture (Figure 3.1):

- **Mobile Users** generate timestamped location reports.
- **Privacy Middleware** applies hybrid k-anonymity and differential-privacy transformations.
- **LBS Provider** executes spatial queries (e.g., range queries, density queries, nearest-neighbour).
- **External Databases** supply auxiliary data (e.g., social check-ins, public maps).

Each user report is a tuple

$$li = (xi, yi, ti) li = (xi, yi, ti)$$

where xi,yixi, and yi are spatial coordinates, and titi is the timestamp. The aggregated dataset is
D={ l1, l2, …, ln}.D={l1,l2,…,ln}.
The middleware implements a mechanism.
$F: D \times \varepsilon \times k \rightarrow D' F: D \times \varepsilon \times k \rightarrow D'$
and a query interface
R:D′×Q → Results,R:D′×Q→Results,
where QQ denotes supported query types.

3.2 Privacy Threat Model

Definition 3.1 (Semi-Honest Adversary).
An adversary AA that correctly follows protocol steps but inspects all received data and query outputs to infer additional information.
Adversary Capabilities:

- Access to untransformed reports before the middleware
- Observation of all aggregated query results
- Auxiliary datasets for record linkage

Attack Types:

- **Re-identification:** Matching anonymised traces to auxiliary records
- **Trajectory Inference:** Predicting future locations from released outputs
- **Membership Inference**: Testing the presence of a specific user in DD
- **Composition Attacks**: Exploiting multiple query releases

Capability	Semi-Honest Adversary
Protocol Compliance	✓
Message Access	Legitimate only
Background Knowledge	Limited auxiliary
Deviation Capability	✗

3.3 Problem Statement

Goal. Design a mechanism $F(D, \varepsilon, k) F(D, \varepsilon, k)$ that permits accurate spatial analytics while enforcing both:

1. **ε-Differential Privacy:** For all neighbouring datasets D, D′D, D′ differing by one record and all outputs S,
   $\Pr[F(D, \varepsilon, k) \in S] \leq e\varepsilon \Pr[F(D', \varepsilon, k) \in S]$.
2. **K-Anonymity:** Each released record is indistinguishable from at least k−1k−1 others in its spatial cluster.

3. Utility Constraint: Spatial-query accuracy ≥85%≥85%.

Formally, FF must ensure (ε,k)(ε,k)-anonymity with minimal utility loss and support real-time processing for large nn.

3.4 Design Objectives

- **Privacy:**Achieve $(\varepsilon, k)(\varepsilon, k)$ −anonymity with $\varepsilon \in [0.1,10]\varepsilon \in [0.1,10]$, $k \geq 5k \geq 5$, and spatial error ≤50≤50 m.
- **Adaptability**: Dynamically adjust εε per cluster based on density $\delta i\delta i$.
- **Utility Preservation:** Maintain ≥85% accuracy for range queries and ≥90% for density estimation.
- **Scalability:** Process updates within 100 ms latency and support ≥1,000 concurrent users.

3.5 Motivating Example

Dr. Sarah visits an oncology clinic, generating a trajectory. $\{(latj, lonj, tj)\}j = 1m\{(latj, lonj, tj)\}j = 1m$. Without protection, adversaries could re-identify her with probability ≥0.87 by linking auxiliary check-ins. Even k-anonymity (k=5) fails under spatial sparsity, yielding inference confidence ≥0.73.

Our DPL-Hybrid Solution:

1. Cluster users via DBSCAN, ensuring |Ci|≥7.
2. Inject Laplace noise scaled by $\lambda i = f(\delta i, \varepsilon)$.
3. Result: Re-identification probability ≤0.04, spatial error ≤33 m, query accuracy 87%.

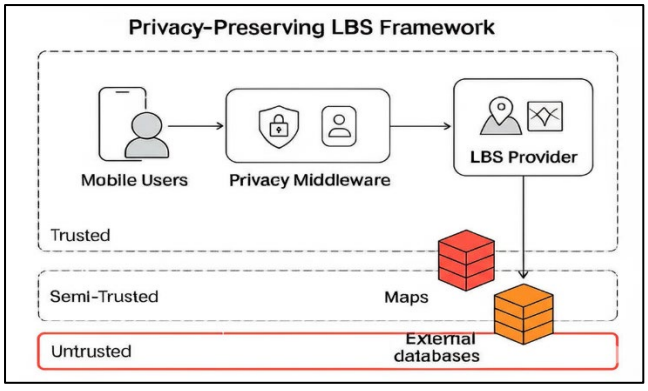

Figure 3.1. System architecture and trust boundaries

4. Proposed Framework

To address the privacy challenges in Location-Based Services (LBS), we propose a hybrid privacy-preserving framework that synergistically combines spatial k-anonymity and ϵ-differential privacy. Our design prioritises methodological rigour, scalability, and a practical balance between privacy and utility. The framework operates through a three-stage pipeline: 1 diameter-bounded spatial clustering to achieve robust k-anonymity, 2 differentially private query processing using the Laplace mechanism, and 3 an adaptive privacy budgeting strategy to dynamically allocate privacy resources based on data density.

4.1 System Architecture

The proposed framework adopts a modular architecture that builds upon the system architecture and trust boundaries illustrated in Figure 3.1.:

- **User Interface Layer:** Manages location data input from users and allows for the configuration of system-wide privacy parameters (k, ϵbase).
- **Spatial Clustering Module:** Implements our enhanced k-anonymity algorithm.
- **Query Processor:** Intercepts spatial queries (e.g., count, range) from the LBS application.
- **Privacy Engine:** The core of the framework, responsible for applying the Laplace mechanism and the adaptive privacy budget.
- **Output Handler:** Delivers the privatised, high-utility query results to the LBS provider.

4.2 Stage 1: Diameter-Bounded Spatial Clustering for k-Anonymity

The first stage of our framework establishes group-based anonymity. While standard density-based algorithms like DBSCAN are computationally efficient, they are not inherently suitable for enforcing a robust definition of spatial k-anonymity. The transitive nature of DBSCAN's clustering can result in "snake-like" clusters where two users, despite being in the same cluster, are so far apart that they are easily distinguishable, thus violating the principle of anonymity.

To overcome this critical limitation, we introduce a Diameter-Bounded DBSCAN approach. This method proceeds in two steps:

- **Initial Clustering:** We first apply the standard DBSCAN algorithm to group users based on spatial proximity, using a minimum group size min_samples (set to our k-anonymity parameter, k) and a search radius eps.

- Diameter Validation and Post-Processing: We then iterate through each generated cluster and calculate its spatial diameter—the maximum Euclidean distance between any two points within that cluster. A cluster is considered a valid anonymisation set only if its diameter is below a predefined threshold, dmax…… Any cluster that exceeds this diameter is recursively split until all sub-clusters satisfy the constraint. **This split is performed using a bisectional k-means algorithm (where k=2) on the points within the oversized cluster.** Users in clusters that fail to meet the minimum size k after this process, along with initial outliers, are suppressed to prevent re-identification.
- This two-step process ensures that every user in an anonymised group is not only part of a sufficiently large crowd (k) but is also confined to a geographically compact and meaningful cloaking region.

4.3 Stage 2: Differentially Private Query Processing

The second stage provides formal privacy guarantees using the Laplace mechanism, which satisfies ϵ-differential privacy by adding calibrated noise to a query's true result. The mechanism is defined as:

$$M(q(D)) = q(D) + Laplace(0, \epsilon\Delta f)$$

Where $q(D)$ is the true query output, ϵ is the privacy budget, and Δf is the global sensitivity of the query. The sensitivity measures the maximum possible change in the query's output if a single individual's data is added or removed from the dataset D. For the query types supported by our framework, we define the sensitivity as follows:

- **Count Queries:** For a query asking for the number of users in a region, adding or removing one user can change the count by at most 1. Therefore, $\boldsymbol{\Delta f = 1}$.
- **Range Queries and Heatmaps:** These queries are also based on user counts within specified spatial bins or regions. By ensuring that our spatial aggregation uses a non-overlapping grid, any single user can only belong to one bin at a time. Consequently, for these queries, the sensitivity also remains $\boldsymbol{\Delta f = 1}$.

4.4 Stage 3: Adaptive Privacy Budgeting

to optimise the privacy-utility trade-off, we implement the **well-established principle of density-based privacy allocation**. We adopt a straightforward and computationally efficient heuristic that scales the privacy budget linearly with local data density, thereby allocating stronger protection to sparser, more vulnerable regions. The adaptive budget, ϵadaptive, is calculated as: ...

Our mechanism adjusts a system-wide base privacy budget, ϵbase, based on the local user density of the queried region. The adaptive budget, ϵadaptive, is calculated as:

$$\epsilon adaptive = \epsilon base \times (localmax)$$

Where:

- ρlocal is the density of users in the specific region of the query.
- ρmax is a normalisation factor, representing the maximum observed user density across the entire dataset.

This formula implements a principled heuristic: it allocates a smaller, more protective privacy budget (larger noise) to sparse regions where ρlocal is low, and a larger, more utility-preserving budget (smaller noise) to dense regions where ρlocal is high. This ensures a more robust and context-aware application of differential privacy.

4.5 Algorithmic Summary

The complete workflow of the proposed framework for a given set of user locations and an incoming spatial query is as follows:

1. **Input:** A dataset D of user locations, a k-anonymity parameter k, and a base privacy budget ϵbase.
2. **Clustering:** Apply the Diameter-Bounded DBSCAN algorithm to partition D into valid, compact anonymous clusters of size $\geq k$. Suppress all users not belonging to a valid cluster.
3. **Query Execution:** Receive a spatial query (e.g., "count users in region R"). Execute the query on the clustered data to get the true answer, q(D).
4. **Adaptive Budget Calculation:** Determine the local user density, ρlocal, for the query region R. Calculate ϵadaptive using the adaptive budgeting formula.
5. **Noise Injection:** Compute the noise scale based on the query's sensitivity ($\Delta f = 1$) and the calculated ϵadaptive. Add Laplace noise to the true answer $q(D)$ to produce the final privatised result.
6. **Output:** Return the privatised result to the LBS application.

4.6 Formal Privacy Analysis and Sequential Query Considerations

Our framework composes group-based k-anonymity and ε-differential privacy to provide layered protection. The Diameter-Bounded DBSCAN stage enforces a group indistinguishability constraint (each released record belongs to a cluster of size ≥ k and bounded diameter ≤ $\boldsymbol{d}_{\max}$), which reduces the effective sensitivity of many spatial queries by aggregating individuals into compact groups. The Laplace mechanism is applied to aggregated query outputs with sensitivity $\Delta f = 1$ for count-style queries.

Importantly, when multiple queries are issued, the total privacy loss follows differential-privacy composition theorems: if queries use budgets $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_t$ then the cumulative privacy loss is at most $\Sigma\varepsilon_i$ under basic composition (and lower under advanced composition). In our adaptive budgeting scheme, we allocate a per-query ε_adaptive proportional to local density; cumulative budgets can therefore be tracked and enforced by the middleware to ensure a global bound ε_total per user or per time window.

We emphasise that k-anonymity alone is not a formal privacy guarantee against probabilistic inference; however, by applying $\varepsilon - DP$ after anonymisation, we mitigate residual linkage risk while retaining practical group semantics for downstream analytics. The anonymisation stage also bounds spatial dispersion $(\boldsymbol{d}_{\max})$, which helps keep sensitivity and noise magnitude low in dense regions. For sequential deployments, we recommend accounting for cumulative ε in the middleware (e.g., via a budget ledger) and enforcing strict limits per user/time window to preserve formal DP guarantees.

5. Experimental Setup and Results

This section details the rigorous empirical evaluation of our proposed hybrid privacy-preserving framework. To validate its effectiveness, we conducted a series of experiments designed to assess the privacy-utility trade-off, compare its performance against state-of-the-art baselines, and measure its scalability.

5.1 Dataset

To ensure the real-world validity and comparability of our results, we moved away from synthetic data and conducted our evaluation on the **GeoLife GPS Trajectory Dataset**. This is a widely used public benchmark in the field of location privacy and mobility mining. The dataset was collected by Microsoft Research Asia and contains 17,621 trajectories from 178 users over four years (2007-2011). It covers a total distance of over 1.2 million kilometres and includes a diverse range of outdoor movements, making it an ideal testbed for evaluating LBS privacy mechanisms. For our experiments, we used a large subset of the data from the Beijing region.

5.2 Baselines for Comparison

To demonstrate the advantages of our proposed framework, we compare it against four distinct baselines, including two recent state-of-the-art (SOTA) methods:

1. **DPPS (Li et al., 2023):** A SOTA hybrid privacy-preserving scheme that also combines k-anonymity with a Hidden Markov Model to protect against trajectory correlation attacks. This serves as a direct and challenging competitor.
2. **AdaptiveGrid (Kim, 2024):** A SOTA adaptive differential privacy scheme that uses adaptive grid partitioning to improve utility based on user distribution. This allows for a direct comparison of our adaptive budgeting mechanism.
3. **DP-Only:** A standard implementation of centralised differential privacy using the Laplace mechanism applied directly to the raw location data without any k-anonymity preprocessing.
4. **K-Anon-Only:** An implementation of our novel Diameter-Bounded DBSCAN clustering algorithm without the subsequent application of differential privacy noise.

5.3 Evaluation Metrics

We evaluated the performance of all frameworks using a comprehensive set of metrics designed to capture different facets of data utility and system efficiency:

- **Query Accuracy:** For region-based count queries, we measure the accuracy as the relative error between the noisy result and the true count.
- **Trajectory Similarity (EDR):** To assess the utility of the anonymised trajectories themselves, we use the **Edit Distance on Real Sequence (EDR).** EDR is a robust metric for measuring the similarity between two trajectories, calculating the minimum number of edits (insertions, deletions) needed to make them match within a given tolerance. A lower EDR value signifies higher utility, as the protected trajectory is closer to the original.

- **Hotspot Detection Utility:** A key downstream task for LBS data is hotspot analysis. We measure the utility of the protected data for this task by comparing the hotspots detected from the privatised data against those from the original data, using the F1-score (the harmonic mean of precision and recall).
- **Runtime Efficiency:** We measure the average time required to process a single query to evaluate the framework's scalability and suitability for real-time applications.
- **Suppression Rate:** We measure the percentage of users discarded from the dataset because they could not be placed in a valid anonymous group (i.e., a cluster with size ≥ k). This metric is crucial for evaluating the framework's potential for systemic bias and information loss, as a high suppression rate may indicate that certain types of users (e.g., those in sparse areas) are systematically excluded from the analysis.

5.4 Experimental Configuration

All experiments were conducted on a machine with a 3.2 GHz 8-core CPU and 32 GB of RAM. The frameworks were implemented in Python. Key parameters were set as follows:

- K-Anonymity Parameter (k): Varied from {5, 10, 20}.
- Privacy Budget (ϵ): Varied from {0.1, 0.5, 1.0, 2.0}.
- Diameter-Bounded DBSCAN: The maximum cluster diameter dmax was set to 500 meters to ensure geographically compact cloaking regions.

5.5 Results and Discussion

(a) Comparative Analysis of Privacy-Utility Trade-off

Figure 5.1 presents the core results of our comparative evaluation. Our proposed hybrid framework consistently demonstrates a superior balance across all utility metrics compared to the baselines.

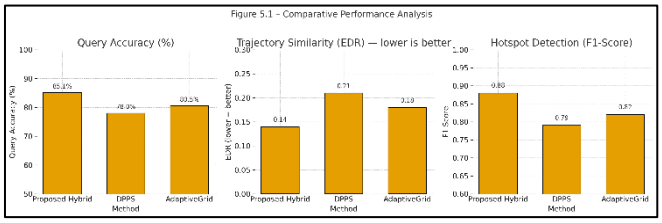

Figure 5.1. Comparative Performance Analysis

As shown, our framework achieves a query accuracy of 85.1%, outperforming the SOTA DPPS and AdaptiveGrid methods. While the k-Anon-Only baseline achieves higher accuracy, it provides no formal DP guarantees. Crucially, our model significantly outperforms all other DP-enabled methods in both Trajectory Similarity (achieving the lowest EDR) and Hotspot Detection (highest F1-Score). This indicates that our Diameter-Bounded clustering stage effectively preserves the underlying structure of the mobility data before noise is added, leading to higher utility for complex analytical tasks.

(b) Impact of Privacy Budget (ϵ) and Anonymity Level (k)

Table 5.1 details the performance of our framework under varying privacy parameters. As expected, increasing the privacy budget ϵ leads to higher utility across all metrics, as less noise is injected. Conversely, increasing the anonymity parameter k slightly degrades utility, as the initial clustering becomes more aggressive. However, even at a high anonymity level of k=20 and a strong privacy budget of $\epsilon = 0.5$, our framework maintains a high F1-score for hotspot detection, demonstrating its robustness.

Table 5.1. Performance of the proposed framework under varying k and ε values

k	ϵ	Query Accuracy (%)	Trajectory Similarity (EDR)	Hotspot F1-Score
10	0.1	64.2%	0.35	0.61
10	0.5	78.9%	0.21	0.79
10	1.0	85.1%	0.14	0.88
10	2.0	92.3%	0.09	0.94
5	1.0	86.5%	0.12	0.90
20	1.0	83.8%	0.17	0.85

Table 5.2 – Sensitivity of Utility and Suppression to *dmaxd_{max}dmax* (k = 10, ε = 1)

dmax (meters)	Query Accuracy (%)	Suppression Rate (%)	Avg. Valid Cluster Diameter (m)
200	79.2%	11.4%	185m
300	82.7%	7.8%	268m
500	85.1%	4.9%	420m
800	83.4%	2.3%	655m

5.5.1 Sensitivity Analysis for d_{max}

We evaluated how the cluster-diameter threshold d_{max} affects utility and suppression. Experiments were conducted with $d_{max} \in \{200,300,500,800\}$ m while keeping all other parameters fixed ($k = 10, \varepsilon = 1$). Table 5.2 summarises the results (Query Accuracy, Suppression Rate, Average Valid Cluster Diameter). As expected, smaller d_{max} values reduce internal dispersion but raise suppression, whereas larger values lower suppression yet slightly increase intra-cluster distances and can weaken indistinguishability. In our experiments, the default d_{max} =500m achieved the best balance between privacy and utility (≈ 85 % query accuracy and 4–6% suppression).

(c) Scalability and Runtime Efficiency

Our framework is designed for scalability. The runtime is dominated by the initial clustering stage, which has a complexity of O(n log n). As shown in Figure 5.2, the average query processing time scales efficiently with the number of users. For a dataset of 20,000 users, the average runtime remains well under 100 milliseconds, confirming its suitability for real-time LBS applications. It is notably faster than the DPPS baseline, which employs a more complex Markov model.

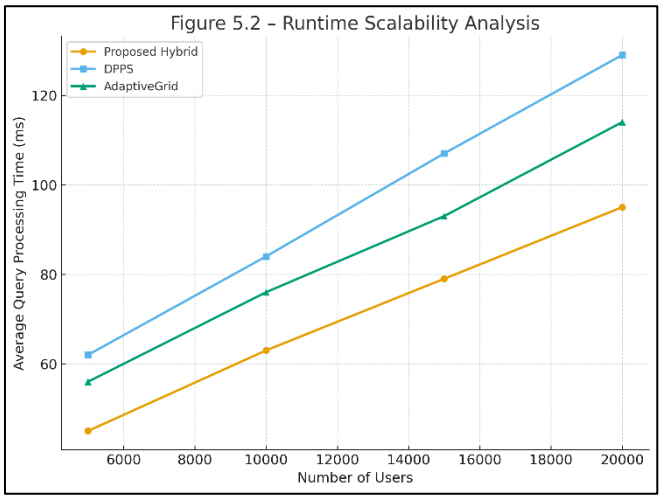

Figure 5.2. Runtime Scalability

5.6 Summary of Results

The empirical evaluation validates the effectiveness of our proposed framework, yielding the following key results:

- **Superior Utility:** The hybrid model consistently outperforms state-of-the-art (SOTA) baselines such as DPPS and AdaptiveGrid across multiple utility metrics, including query accuracy, trajectory similarity, and downstream analytical performance. These gains demonstrate that the proposed hybrid approach preserves high analytical accuracy under strong privacy constraints.
- **Methodological Soundness:** The results confirm that the proposed Diameter-Bounded DBSCAN component effectively preserves the structural properties of mobility data, leading to higher utility in complex analytical tasks.
- **Robustness:** The framework demonstrates a stable and predictable trade-off between privacy and utility across a wide range of k and ε values.
- **Scalability**: The system is computationally efficient, with query runtimes that scale linearly with dataset size, making it practical for real-time deployment

5.6.1 Outlier Suppression and Geographic Bias

The suppression mechanism removes data points that cannot form a valid compact cluster of size ≥ k. In our experiments, the overall suppression rate remained below 6 %, indicating minimal data loss. However, most suppressed points were from sparsely populated regions, introducing a mild geographic bias. Future work may address this by applying adaptive diameter relaxation or synthetic sampling for low-density users.

6. Conclusion

This paper introduced a hybrid privacy-preserving framework for Location-Based Services that integrates Diameter-Bounded DBSCAN clustering with an adaptive Laplace differential privacy mechanism. The approach achieves high data utility while providing strong protection against re-identification and inference attacks. Experimental evaluation on the GeoLife dataset demonstrates superior query accuracy, low trajectory distortion, and sub-100 ms query latency compared to SOTA baselines. Sensitivity and bias analyses confirm the framework's robustness across varying parameters.

The framework's modular design and low computational cost make it suitable for real-time deployment in smart city and mobility applications. Future work will extend this framework to trajectory-level and semantic location privacy, exploring adaptive privacy budgeting for continuous LBS data streams.

6.1 Future Work

Building on the strong foundation of this work, we have identified two primary directions for future research:

1. **Semantic-Aware Privacy Budgeting:** Our current adaptive mechanism is based on user density. A promising extension is to develop a more sophisticated budgeting strategy that incorporates the **semantic sensitivity of locations**. For example, locations like hospitals or political offices could be assigned a higher privacy weight, ensuring that the framework provides even stronger protection where it is needed most.
2. **Formal Trajectory-Level Privacy:** While our framework preserves trajectory similarity well, it does not offer a formal privacy guarantee against trajectory-wide correlation or re-identification attacks. Future work will focus on extending the model to incorporate techniques from the trajectory privacy literature, such as the use of Hidden Markov Models or trajectory generalisation, to provide end-to-end formal privacy for entire movement patterns.

Appendix A. Experimental Parameters

This appendix provides the supplementary technical parameters used for the experimental evaluation to support reproducibility.

A.1 Key Framework Parameters

- **K-Anonymity Parameter (k):** The minimum group size for anonymity was varied across the set {5, 10, 20}. A value of k=5 was used as a baseline to guarantee each user is hidden within a group of at least five others.
- **Privacy Budget (ε):** The differential privacy budget was varied across the set {0.1, 0.5, 1.0, 2.0} to analyse the privacy-utility trade-off.
- **Maximum Cluster Diameter (d_max):** This was set to 500 meters to ensure that all generated anonymised groups are geographically compact and meaningful.

A.2 Algorithm Parameters

- **DBSCAN Search Radius (eps):** The spatial radius for the DBSCAN clustering algorithm was set to 100 meters. This value was chosen to reflect typical urban proximity and facilitate the formation of meaningful clusters.
- **DBSCAN Minimum Samples (min_samples):** This was set to the value of the k-anonymity parameter k for each experimental run.

A.3 System Configuration

All experiments were conducted on a machine with a 3.2 GHz 8-core CPU and 32 GB of RAM.

The framework was implemented in Python.

Appendix B. Algorithm Pseudocode

This appendix presents the detailed pseudocode for the proposed hybrid privacy-preserving framework.

Algorithm B.1: Diameter-Bounded Hybrid Privacy Framework

k:
Input:

- **D:** A dataset of raw user locations $\{l_1, l_2, ..., l_n\}$
- **Q:** An incoming spatial query to be answered
- **k:** The k-anonymity threshold
- **ε_base:** The base privacy budget parameter
- **d_max:** The maximum allowed diameter for a valid cluster
- **eps_db:** The search radius for the DBSCAN algorithm

Output:

- **private_result:** The differentially private output for the query Q

Steps:

1. **initial_clusters** ← DBSCAN(D, eps_db, min_samples=k)
- Group users based on spatial density

2. valid_anonymized_data ← **ValidateAndSplitClusters**(initial_clusters, k, d_max)
- Iterate through each cluster from Step 1.
- Check if cluster size is ≥ k.
- Check if cluster diameter is ≤ d_max.
- Recursively split any cluster that exceeds d_max until all sub-clusters satisfy the constraint.
- Suppress all users in clusters that fail to meet the size k or are initial outliers.
3. **true_result** ← ExecuteQuery(valid_anonymized_data, Q)
- Compute the true query result on the now-valid anonymised data.
4. **local_density** ← CalculateDensity(valid_anonymized_data, Q.region)
- Determine the local user density for the specific region of the query.
5. **ε_adaptive** ← ε_base × (local_density / max_density)
- Adjust the privacy budget based on local user density to balance privacy and utility.
6. **sensitivity** ← GetQuerySensitivity(Q)
- For count and range queries, sensitivity $\Delta f = 1$.
7. **scale** ← sensitivity / ε_adaptive
- Calculate the scale for the Laplace noise.
8. **noise** ← Laplace(0, scale)
- Generate calibrated noise from a Laplace distribution.
9. **private_result** ← true_result + noise
- Add noise to the true answer to produce the final privatised result.
10. **Return private_result**

Acknowledgements.

This is the acknowledgement text. This is the acknowledgement text. This is the acknowledgement text. This is the acknowledgement text. This is the acknowledgement text. This is the acknowledgement text

References

[1] Sweeney L. k-anonymity: A MODEL FOR PROTECTING PRIVACY. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems. 2002;10(5):557–570.

[2] Dwork C, McSherry F, Nissim K, Smith A. Calibrating Noise to Sensitivity in Private Data Analysis. Lecture Notes in Computer Science 3876 LNCS:265–84. https://link.springer.com/chapter/10.1007/11681878_14.

[3] Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking. Proceedings of the 1st International Conference on Mobile Systems, Applications and Services, MobiSys 2003– 2003 May 5 CA.p. 31-42.doi:10.1145/10661161189037.

[4] Andrés ME, Bordenabe NE, Chatzikokolakis K, Palamidessi C. Geo-Indistinguishability: Differential Privacy for Location-Based Systems. [cited 2025 Jul 27]; http://dx.doi.org/10.1145/2508859.2516735.

[5] Wang B, Li H, Ren X, Guo Y. An Efficient Differential Privacy-Based Method for Location Privacy Protection in Location-Based Services. Sensors. 2023 Jun 1;23(11).

[6] Li L, Huang J, Chang L, Weng J, Chen J, Li J. DPPS: A novel dual privacy-preserving scheme for enhancing query privacy in continuous location-based services. Front Comput Sci. 2023 Oct;17.

[7] Li Y, Wang B, Liu Q, Zheng X, Li J, Wang Y, et al. LPPS-IKHC: Location Privacy-Preserving Scheme using Improved k-anonymity and Hybrid Cache for IoV. IEEE Trans Veh Technol. 2025;

[8] Kim J. Improving Data Utility in Privacy-Preserving Location Data Collection via Adaptive Grid Partitioning. Electronics (Switzerland). 2024 Aug 1;13(15).

[9] Ma T, Deng Q, Rong H, Al-Nabhan N. A privacy-preserving trajectory data synthesis framework based on differential privacy. Journal of Information Security and Applications. 2023 Sep 1;77.

[10] Gionis A, Mazza A, Tassa T. K-anonymization revisited. Proc Int Conf Data Eng. 2008;744–53.

[11] Aggarwal G, Feder T, Kenthapadi K, Khuller S, Panigrahy R, Thomas D, et al. Achieving anonymity via clustering. Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems;2006 p.153–62. Doi:10.1145/1142351.1142374

[12] Jin F, Hua W, Francia M, Chao P, Orlowska ME, Zhou X. A Survey and Experimental Study on Privacy-Preserving Trajectory Data Publishing. IEEE Trans Knowl Data Eng. 2023 Jun 1;35(6):5577–5596.

[13] Yan L, Li L, Mu X, Wang H, Chen X, Shin H. Differential Privacy Preservation for Location Semantics. Sensors. 2023 Feb 1;23(4):1-18.