ALOG: Adaptive Longitudinal Grids for Geospatial Data using Local Differential Privacy

Eduardo R. Duarte Neto
Universidade Federal do Ceará, Brazil
Fortaleza, Ceará, Brazil
eduardo.rodrigues@lsbd.ufc.br

José S. Costa Filho
Universidade Federal do Ceará, Brazil
Fortaleza, Ceará, Brazil
serafim.costa@lsbd.ufc.br

Antonio A. Marreiras Neto
Universidade Federal do Ceará, Brazil
Fortaleza, Ceará, Brazil
antonio.marreiras@lsbd.ufc.br

Javam C. Machado
Universidade Federal do Ceará, Brazil
Fortaleza, Ceará, Brazil
javam.machado@lsbd.ufc.br

Abstract

This study introduces ALOG (Adaptive Longitudinal Grids for Geospatial Data using Local Differential Privacy), a novel framework designed to optimize geospatial data collection and frequency estimation while ensuring robust user privacy. ALOG leverages adaptive grids to dynamically adjust spatial granularity based on data density, eliminating the need for prior knowledge about data distribution. We evaluate ALOG and its variations using both synthetic and real-world datasets, comparing their effectiveness against state-of-the-art protocols. Experimental results demonstrate ALOG's superior performance in balancing privacy and utility, particularly under varying grid sizes and privacy budgets. The findings highlight the effectiveness of adaptive grid refinement in achieving precise frequency estimates in privacy-sensitive applications without relying on prior knowledge of data density.

Keywords

Local Differential Privacy, Adaptive Grids, Geospatial Data, Frequency Estimation

1 Introduction

The growing popularity of mobile devices and location-based services has led to an unprecedented influx of geospatial data, capturing detailed information on individual movements, location preferences, and popular routes [7, 12]. This surge in location data opens up opportunities for a wide range of applications across sectors, from optimizing public transportation systems and enhancing emergency response times to supporting intelligent urban planning and resource allocation [25]. One key approach in analyzing this data is through frequency analysis, which identifies areas with high concentrations of people over time [17]. By quantifying how often certain locations are visited, frequency analysis provides crucial insights that enable decision-makers to allocate resources effectively and design services that meet population needs.

Frequency analysis, for instance, is widely used for generating heat maps [5, 18, 32], which visually represent high-frequency areas or routes based on user interactions. By highlighting patterns in pedestrian traffic, vehicle concentration, or other location-based trends [30], these maps offer actionable insights into critical areas that could benefit from additional services or infrastructural improvements [24]. Furthermore, tracking frequency changes over time enhances understanding of dynamic trends, allowing analysis of complex, time-dependent phenomena. For instance, observing shifts in traffic flow patterns helps city planners design better road systems and optimize signal timings to alleviate congestion [19]. In public health, evolving frequency patterns can reveal the spread of infectious diseases [4], guiding officials to proactively allocate resources where they are most needed. Additionally, analyzing changes in movement patterns informs a range of personalized services and applications, from targeted advertising and location-based recommendations to social behavior research [33].

However, while these geospatial analyses offer substantial value across fields, they also raise serious privacy concerns. The data required to generate heat maps over time and track patterns often includes sensitive information about individuals' locations and movements, leading to potential risks related to user identification, behavioral profiling, and location-based discrimination [24, 34, 36]. As a result, preserving the privacy of users while leveraging the benefits of geospatial data has become a crucial issue.

Differential Privacy (DP) has emerged as a prominent framework for enabling data analysis while preserving individual privacy [9, 14]. However, its traditional model relies on a trusted curator, introducing potential vulnerabilities due to the risk of data breaches. To address this limitation, Local Differential Privacy (LDP) has been proposed as a robust alternative. LDP allows individuals to anonymize their data locally before sharing it for analysis, eliminating the need for a centralized curator [15, 20]. This decentralized approach has gained significant traction in recent years, with major technology companies such as Apple [38], Google [16], and Microsoft [13] incorporating LDP into their consumer products.

LDP is particularly well-suited for scenarios involving longitudinal data—data collected from the same individual over time. For instance, in systems where users repeatedly send location reports or other personal information at multiple time intervals, the resulting data forms a longitudinal dataset. LDP ensures privacy protection in such settings by anonymizing each data point before transmission.

Despite the advantages of LDP, current techniques applied to geospatial data often rely on uniform grid structures to partition space for simplicity [1, 42], which can lead to an inadequate representation of the data. These approaches apply the same level of granularity to all regions, regardless of the density of points of interest, potentially compromising both the accuracy of the analyses and the utility of the aggregated data. This issue, known as non-uniformity error, arises when the assumption of

EDBT '26, Tampere (Finland)

uniform data distribution within grid cells does not align with the actual data distribution [27]. In dense regions, this mismatch can lead to underrepresentation of critical areas, introducing bias into the results, while in sparse regions, excessive granularity can amplify noise, reducing the utility of the data. Because these methods don't consider how data is distributed within a region, they can produce large errors. For example, in dense regions, uniform grids may be too coarse, leading to significant bias when answering queries. On the other hand, regions with sparse data may be mapped by a fine-grained grid, resulting in weak utility and high noise during query answering [27].

In this paper, we propose a new approach based on adaptive, non-uniform grids that dynamically adjust to data point densities, allowing for greater accuracy in regions with high individual concentration and coarser granularity in less dense areas. Our methodology combines the flexibility of adaptive grids with the privacy guarantees offered by Local Differential Privacy. By iteratively refining cell sizes as new data is reported, our approach incrementally adjusts to temporal variations in data distribution. This temporal refinement is essential to accurately capture the correlations present among consecutive reports submitted under Local Differential Privacy (LDP), enabling our framework to deliver more precise and context-aware analyses. Thus, these enhancements make our framework, called ALOG (Adaptive Longitudinal Grids), particularly well-suited for addressing the challenges of privacy-preserving geospatial data analysis.

Contributions. In summary, this paper makes the following contributions:

- We propose ALOG, an adaptive grid-based framework for LDP-compliant longitudinal data collection, which exploits temporal correlations and dynamically refines spatial partitions to balance privacy and utility.
- We analyze various grid refinement strategies to detect and adapt to changing patterns in data density throughout the grid adaptation windows with a detailed investigation of the factors that influence the utility-privacy trade-off.
- To assess the effectiveness of our adaptive longitudinal data collection framework, we conduct comprehensive evaluations using both synthetic and real-world datasets while comparing them with different prior state-of-the-art techniques.

Organization. Section 2 formalizes the problem tackled in this work, and presents the foundations and essential definitions used throughout the work. In Section 3, we present the related work and discuss how the main existing approaches work and their limitations. We present our approach in Section 4. Experimental results are shown in Section 5. Finally, Section 6 concludes this work.

2 Foundations and Key Concepts

This work addresses the challenge of estimating the frequency of user visits within a spatial region of interest, S (e.g., a city), while ensuring user privacy. The region S is partitioned into a uniform grid G, where each cell $g_i \in G$ represents a distinct sub-region. For a set of users $U = \{u_1, u_2, ..., u_n\}$ moving within S across discrete timestamps, the objective is to estimate the frequency $\hat{f}_i$ of visits to each cell g_i. Users generate true location reports r_i at each timestamp, which are then anonymized locally using a Local Differential Privacy (LDP) mechanism Ψ to produce private reports $\hat{r}_i = \Psi(r_i)$ [11]. An aggregator collects these private reports $\{\hat{r}_i\}_{i=1}^{|U|}$ and estimates the cell frequencies $\hat{f}_i$ using an estimator function $F(\cdot)$. Key concepts like LDP and Frequency Estimators, detailed below, form the basis of this approach. Table 1 summarizes commonly used notations.

Table 1: Commonly Used Notations

Notations	Meaning
S	Geo-space
g_i,G_i	Grid cell and Grid partition of S
u,U	User and the set of all users
t,T	Timestamp and set of all timestamps
f_i,$\hat{f}_i$	Frequency and frequency estimation for cell g_i
r_i,$\hat{r}_i$	User true report and user private report at each timestamp i
w, k	Window and grid size

2.1 Local Differential Privacy

Local Differential Privacy (LDP) differs significantly from the centralized model of Differential Privacy (DP) [14] in where data perturbation occurs. In LDP, each user applies a randomized algorithm Ψ locally to perturb their true data *before* it is transmitted. Only the perturbed output is sent to the server, protecting the raw data and making LDP suitable for settings without a fully trusted data curator. This model is well-suited for longitudinal data, where users repeatedly report information over time, ensuring privacy at each step [2].

Definition 2.1. (ϵ - LDP [16]) An algorithm $\Psi(\cdot)$ satisfies ϵ-local differential privacy, where $\epsilon \geq 0$, if and only if for any pair of inputs (v, v'), and any set R of possible outputs of Ψ, we have

$$Pr[\Psi(v) \in R] \leq e^{\epsilon} Pr[\Psi(v') \in R] \quad (1)$$

Intuitively, *ϵ-LDP* ensures that given the perturbed output in R, it is not possible for the server or for an adversary to distinguish whether the original true value was v or v' beyond the probability ratio e^{ϵ}. Here, ϵ controls the privacy level. Lower ϵ yields stronger privacy. Analogous to central DP, LDP also benefits from key foundational properties, including robustness to post-processing and composition [15].

Proposition 2.2. *(Post-Processing) If Ψ is ϵ-LDP, then $f(\Psi)$ is also ϵ-LDP for any function f.*

Proposition 2.3. *(Sequential Composition) Let each Ψ_i be an ϵ-LDP mechanism, and Ψ is the sequential composition Ψ_1,...,Ψ_m. Then, Ψ satisfies $\sum_{i=1}^{m} \epsilon$-LDP.*

2.2 Frequency Estimator

A frequency estimator is an LDP mechanism designed to estimate how often each value $r \in V$ appears in the dataset, where V is the set of all possible values of a given attribute [6, 8]. A frequency estimator consists of two components: A sanitizer and an aggregator. The sanitizer is employed by users to locally perturb their data. An aggregator receives the perturbed data and estimates the frequencies regarding the input. Frequency estimators have been employed for a variety of different tasks such as answering range queries [11], identifying heavy hitters [29, 44], and the problem we attack in this paper, namely private queries on geospatial data [17, 23, 35]. Figure 1 illustrates a typical frequency estimator. In

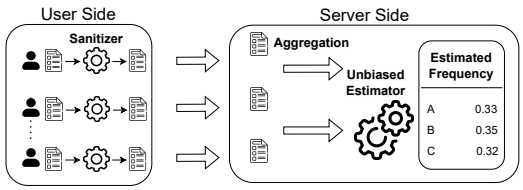


Figure 1: Frequency Estimator: users sending their sanitized report to the aggregator to estimate the frequency of values for a given attribute.

this example, the domain consists of three possible values: A, B, and C, which could represent items a user purchased, such as a doll, a car, or a dog. Each user applies a sanitization mechanism to their true report before submitting the sanitized version to the server. On the server side, the aggregator collects all sanitized reports and estimates the frequency of each value in the domain.

2.3 Space Partitioning

Spatial regions are commonly represented using map projections such as the Mercator projection, which converts latitude and longitude into a planar format [26]. While suitable for visualization, many analytical tasks require discretizing this continuous space into manageable units. Grid-based partitioning is a widely used approach due to its simplicity and effectiveness in dividing the region into uniform, contiguous cells [31].

Each cell represents a discrete sub-region of the original space, allowing any point (latitude and longitude) to be mapped to a specific cell. This structure supports efficient spatial queries, aggregations, and systematic data processing, particularly in high-density datasets [43].

To represent positions within the grid computationally, unary encoding can be applied. In a grid with k cells, each location is represented as a binary vector of length k, with a 1 indicating the active cell and 0s elsewhere. For example, if a user is in cell 20 of a 100-cell grid, their position is encoded as a 100-bit vector with a 1 at index 20. Unary encoding is well-suited for privacy-preserving applications, as it enables straightforward integration with differential privacy mechanisms [41].

By discretizing spatial data and encoding positions in this manner, it becomes possible to anonymize and aggregate location information effectively, enabling privacy-aware frequency estimation and spatial analysis in sensitive domains.

3 Related Work

Centralized models for estimating individual frequencies in geospatial areas typically use grids to partition space for location discretization [28, 39, 40]. Qardaji et al. [28] address the problem of constructing a differentially private synopsis for two-dimensional datasets, such as geospatial datasets. They highlight that a key challenge is balancing noise error and non-uniformity error in partition-based synopsis methods and study the uniform-grid approach. The authors propose a method for choosing the grid size. They also introduce an adaptive-grid method that refines cell granularity based on data density. Kim et al. [21] proposes a method that combines Geo-Indistinguishability [1] with adaptive grids, adjusting grid granularity based on real-time user density to improve location-based analysis accuracy. Unlike centralized approaches, several studies have explored Local Differential Privacy (LDP) to enhance location privacy. Erlingsson et al. [16] introduced RAPPOR, a protocol that uses LDP in longitudinal data collection. Arcolezi et al. [2] proposed L-OSUE, an enhanced LDP protocol that improves frequency estimate accuracy across time and multiple dimensions. Additionally, Arcolezi et al. [3] developed LOLOHA, an extension of OLH, for handling evolving data in trajectory-based applications, which refines hash functions dynamically to enhance location-based analysis accuracy. Cunningham et al. [10] propose an LDP mechanism based on perturbing hierarchically structured, overlapping n-grams of trajectory data. It uses a multi-dimensional hierarchy over real-world places of interest to improve the realism and utility of perturbed, shared trajectories.

While effective, applying LDP protocols to location privacy presents challenges. Specifically, discretizing continuous location data into a grid structure requires dynamic grid evolution to adapt to variations in user density, balancing privacy and data utility.

This work explores the use of adaptive grids with LDP protocols to improve privacy and maintain high data utility. It will be compared experimentally against state-of-the-art LDP methods, such as RAPPOR, LOSUE, and LOLOHA, to demonstrate the effectiveness of adaptive grid structures in handling evolving location-based datasets.

3.1 RAPPOR

RAPPOR [16] is an influential work in the development of LDP and frequency estimators. It makes use of two rounds of sanitization and *memoization* to provide LDP guarantees for several queries over time. RAPPOR is especially relevant for longitudinal location data, as it allows for ongoing collection while protecting individual privacy by encoding and perturbing the data. In the context of location data, the domain size k refers to the number of possible location encodings within the grid. Each user's location at a given timestamp is encoded into this k-dimensional space, and RAPPOR's perturbation mechanisms are applied to this encoding. This process is repeated at each timestamp, allowing for frequency estimation of location visits over time while preserving privacy.

The first round of sanitization is referred to as the PRR step (permanent randomized response), and the second as the IRR step (instantaneous randomized response), both referencing the randomized response algorithm and their respective functions. In the *memoization* process, PRR adds noise to a value the first time it is processed, and then the sanitizer memoizes the noisy value, while IRR adds noise to every instance of a memoized value before sending a report to the aggregator. By memorizing a randomized version of a true value and consistently reusing it, or reusing it as input to a second round of sanitization, it becomes harder for an adversary to track changes in the data and reduces the impact of temporal correlations, making it more difficult to track changes in the data and infer the true values. As the domain size gets larger, it becomes a challenge to apply noise in a utility-preserving way, thus RAPPOR implements Unary Encoding as a means of guaranteeing LDP while reducing data utility loss. In Unary Encoding, an input value v from a domain D of size k is encoded as a binary vector B of length k : $B = [0, ..., 0, 1, 0, ..., 0]$, where only the v-th position is 1.

The perturbation protocol for RAPPOR and other UE-based LDP protocols is as follows: Given the probabilities p_1, p_2, q_1, q_2 to flip the i-th position of the binary vector B, for the first sanitization round, we have

$$Pr\left[\Psi_{UE_{(\epsilon)}} B'[i] = 1\right] = \begin{cases} p_1, \text{if } B[i] = 1 \\ q_1, \text{if } B[i] = 0 \end{cases},$$

and for the second

$$Pr\left[\Psi_{UE_{(\epsilon)}} B'[i] = 1\right] = \begin{cases} p_2, if B[i] = 1 \\ q_2, if B[i] = 0 \end{cases},$$

p_1, p_2, q_1, q_2 are computed as function of ϵ_∞ and ϵ_1 by symmetric unary encoding (SUE) to satisfy: $p_1 = \frac{e^{\epsilon_\infty/2}}{e^{\epsilon_\infty/2}+1}$, $q_1 = \frac{1}{e^{\epsilon_\infty/2}+1}$, $p_2 = 0.75$, and $q_2 = 1 - p_2$.

ϵ_∞ is an upper bound for the privacy budget as the number of reports sampled by the frequency estimator tends to infinity, and ϵ_1 is a lower bound for a single report. The value of ϵ_1 is computed using the following equation:

$$\epsilon_1 = ln\left(\frac{(p_1p_2 - q_2(p_1 - 1))(p_2q_1 - q_2(q_1 - 1) - 1)}{(p_2q_1 - q_2(q_1 - 1))(p_1p_2 - q_2(p_1 - 1) - 1)}\right) \quad (2)$$

All frequency estimators presented in this paper, including those based on the RAPPOR protocol discussed in this section and the L-OSUE and LOLOHA protocols covered in the following two sections, utilize the same unbiased function to estimate frequencies at the aggregator [2, 3, 16]:

$$\Phi_{f_L}(v) := \frac{C(v) - nq_1(p_2 - q_2) - nq_2}{n(p_1 - q_1)(p_2 - q_2)} = \frac{\frac{C(v)/n - q_2}{p_2 - q_2} - q_1}{p_1 - q_1}, \quad (3)$$

where n is the number of reports, $p1, p2, q1, q2$ are the probabilities for perturbing the data, and $C(v)$ is the count of reports with a value v, given $v \in D$.

3.2 L-OSUE

L-OSUE is an alternative to RAPPOR proposed in [2]. It uses the OUE [22] protocol for its first round of sanitization (PRR step) and SUE [37], which is equivalent to simple single-time RAPPOR, for the second (IRR step). The perturbation algorithm follows the same structure as RAPPOR, but with $p_1 = 0.5$, $q_1 = \frac{1}{e^{\epsilon}_{\infty}+1}$, $p2 = \frac{e^{\epsilon}\infty e^{\epsilon}_1 - 1}{e^{\epsilon_\infty} - e^{\epsilon}_1 + e^{\epsilon_\infty + \epsilon_1} - 1}$, $q_2 = 1 - p2$, and ϵ_1 calculated using equation 2. It provides better privacy guarantees than RAPPOR, while not introducing too much extra noise and preserving similar data utility. Similar to RAPPOR, L-OSUE can be applied to longitudinal location data by encoding locations into a k-dimensional space, where k is the number of cells in the grid. At each time step, user locations are encoded, perturbed using L-OSUE's two-round approach, and reported. This allows for improved accuracy in frequency estimates across both time and multiple dimensions, crucial for longitudinal location-based analysis.

3.3 LOLOHA

Unlike RAPPOR and L-OSUE, LOLOHA [3] does not make use of unary encoding; it instead improves utility by shrinking the domain size through local hashing. Like the previously cited estimators, LOLOHA implements two rounds of sanitization and memoization to provide LDP guarantees while maintaining utility for queries executed over time. In the context of location data, LOLOHA uses a hash function to reduce the domain size from k (the number of grid cells) to a smaller space before applying perturbation. This hashing is crucial for managing the evolving nature of location data over time. At each timestamp, user locations are hashed, perturbed, and reported, allowing the aggregator to estimate frequency distributions accurately as new reports are continuously collected. In this paper, we will use OLOLOHA, the optimal configuration of LOLOHA. It shrinks a domain size k to an optimal hashed size g, that is computed by

$$g = 1 + \max\left(1, \left\lfloor \frac{1 - a^2 + \sqrt{A}}{6(a - b)} \right\rfloor\right), \quad (4)$$

where $A = a^4 - 14a^2 + 12ab(1 - ab) + 12a^3b + 1$ given $a = \epsilon_\infty$ and $b = e^{\alpha\epsilon_\infty}$ for $\alpha \in (0, 1)$.

In local hashing, given an original domain size k and input values $v \in V$, a randomly chosen universal hash function H maps the original domain to a range $[1...g]$ with $g \geq 2$. Approximately k/g values $v \in V$can be mapped to the same hashed value $H(v) \in [1...g]$ due to collision. The sanitization protocol for LOLOHA is as follows:

$$\forall_{x \in [1...g]}\ Pr\left[\Psi_{LOLOHA_{(\epsilon_\infty)}}(v) = x\right] = \begin{cases} p_1 = \frac{e^{\epsilon}}{e^{\epsilon}+g-1} & \text{if x = v} \\ q_2 = \frac{1}{e^{\epsilon}+g-1} & \text{if x} \neq \text{v} \end{cases}$$

followed by a second round that outputs a report x':

$$\forall_{x' \in [1...g]}\ Pr\left[\Psi_{L-GRR_{(\epsilon_1)}}(x) = x'\right] = \begin{cases} p_2 & \text{if x' = x} \\ q_2 = \frac{1-p_2}{g-1} & \text{if x' } \neq \text{ x} \end{cases}$$

where:

$$\epsilon_1 = \ln\left(\frac{p_1p_2 + q_1q_2}{p_1q_2 + q_1p_2}\right), \text{ and}$$

$$p_2 = \frac{q_1 - e^{\epsilon_1}p_1}{-p_1e^{\epsilon_1} + gq_1e^{\epsilon_1} - q_1e^{\epsilon_1} - p_1(g - 1) + q_1}.$$

After the sanitization step, the reports, privacy budgets, and hash function seeds are sent to the aggregator, which then outputs estimates with the original domain size k.

4 The ALOG Approach

In this section, we introduce ALOG (Adaptive Longitudinal Grids for Geospatial Data using Local Differential Privacy), a novel approach for privacy-preserving location frequency estimation. ALOG employs a frequency estimation framework that leverages Local Differential Privacy to collect user location data in a privacy-preserving manner. ALOG consists of two primary entities: users, who submit encoded reports of their private locations, and the aggregator, which collects these reports and provides a frequency estimation for each cell of the adaptive grid.

A key feature of ALOG is its use of adaptive grids that evolve dynamically in response to longitudinal data, incorporating an adaptation window to improve resource usage, such as the privacy budget. By capturing correlations within users' data over time, ALOG creates adaptive grids that reduce non-uniformity errors, enhancing the accuracy and utility of spatial data analysis. This framework balances privacy, utility, and adaptability, making it an effective solution for geospatial data collection and analysis in privacy-sensitive contexts.

In the following subsections, we will define how the spatial representation is constructed using this adaptive grid. Then we will introduce three variations of ALOG, each designed to achieve private frequency estimation over time while addressing different aspects of the utility-privacy trade-off.

4.1 Adaptive Grid

The adaptive grid is a non-uniform data structure where cell dimensions are determined by the volume of reports aggregated

within each cell. This approach refines cell granularity in high-density regions based on actual user-submitted reports to address the challenges of skewed data distributions. This process ensures that the overall data distribution across the grid becomes more evenly spread, enhancing the accuracy of frequency estimations. This behavior is demonstrated in Figure 2, where the adaptive grid structure achieves a balanced density between regions with varying population concentrations.

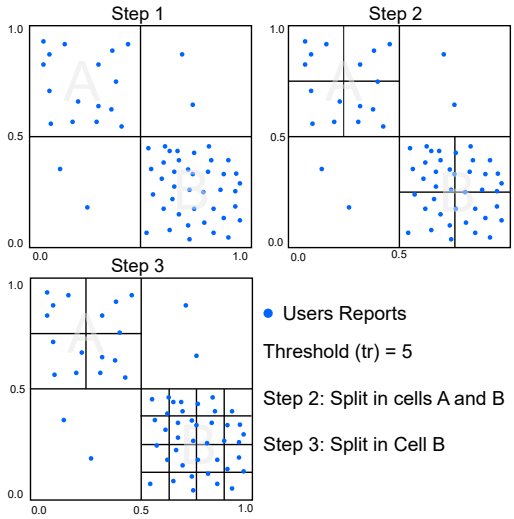


Figure 2: Illustration of the adaptive grid process: Cell A is split into four sub-cells after exceeding the threshold, and Cell B undergoes an additional split into sixteen sub-cells.

The first step in constructing the adaptive grid is to define the maximum allowable number of reports per cell, denoted as the cell threshold tr, which is determined by privacy requirements. Once the aggregator processes the users' reports for a specific timestamp t, it calculates the count c_i of reports for each cell $g \in G$. Cells with report counts exceeding the threshold tr are recursively subdivided into four uniform cells, each occupying an equal portion of the original cell's spatial area. Subsequently, the count for each newly formed cell is assigned as one-quarter of the original cell count c, allowing the adaptive grid to dynamically adjust to areas of high report density by increasing granularity. This process is repeated until all cells satisfy the threshold tr. The complete algorithm for constructing the adaptive grid is detailed in Algorithm 1.

Figure 2 illustrates the grid adaptation process. In the transition from Box 1 to Box 2, cells A and B are each split into four smaller cells. While cell A reaches a count below the threshold tr, cell B still exceeds the threshold, requiring an additional split in Box 3. As a result, all cells achieve a more balanced distribution of reports, ensuring a more uniform spatial representation. The time complexity of the adaptive grid creation is $O(k \log n)$, where k is the number of initial grid cells and n is the number of user reports, reflecting the recursive subdivision of high-density cells.

4.2 Frequency Estimator Framework

ALOG comprises two main components: the users, who send encoded reports of their private locations, and the aggregator, who is responsible for collecting these reports and estimating the frequency of each grid cell. Beyond frequency estimation,

Algorithm 1: Adaptive Grid Creation Algorithm

Input : Grid G, Estimated frequency for timestamp t, Cell threshold tr
Output : Adaptive grid with updated cell counts

Function `AdaptativeGrid` (G)**:**
  Calculate $\forall i \in G : c_i := \hat{f}_i$;
  for *each cell* i **do**
    SplitCell($\langle i, c_i \rangle$);
Function `SplitCell` $(\langle i, c_i \rangle)$**:**
  if $c_i > tr$ **then**
    Split cell i into four sub-cells of equal area;
    for *each sub-cell* j *of cell* i **do**
      Set count for sub-cell j to $c_j = \frac{c_i}{4}$;
      for *each cell* i **do**
        SplitCell($\langle i, c_i \rangle$);

the aggregator also contributes to generating and distributing the adaptive grid—a spatial representation that dynamically adjusts according to report density—to users. This adaptive grid enables refined spatial partitioning that is aligned with the population distribution, enhancing the precision of data collection and analysis.

Initially, the aggregator broadcasts a uniform grid accessible to all users, partitioning the reference space—such as a city—into square cells of equal size. On the user side, each user encodes their current location according to the grid provided by the aggregator, as detailed in Section 2.3. Utilizing a local differential privacy protocol, each user sends a noisy, obfuscated report of their location back to the aggregator. ALOG leverages the LOSUE protocol 3.2, which, following empirical analysis, has proven well-suited to this context; however, other longitudinal LDP protocols could also be employed.

Grid refinement is essential for reducing non-uniformity errors and improving data utility. However, evolving grids during anonymization can decrease memoization efficiency, leading to higher privacy budget consumption. This inefficiency arises because new cells are created with each new grid generated during the adaptation process. As a result, the previous canonical mapping of data to grid cells must be discarded, requiring the system to reinitialize the memoization for the newly created grid structure.

To mitigate this trade-off between budget consumption and grid adaptation, we propose a Grid Adaptation Window (GAW) mechanism. This mechanism ensures that grid adaptations occur only after every w consecutive timestamps, after which the updated grid is broadcast to the users. Figure 3 illustrates the grid adaptation process with a window size of 2. Despite the use of GAW, the aggregator continues performing frequency estimations at each timestamp, ensuring continuous system monitoring and analysis.

An additional benefit of the GAW is that it addresses situations where grid refinement is not always necessary. In some cases, the gain from refining the grid is not significant enough to justify the additional computational cost, as there may be little to no change between consecutive occurrences. In these situations, the GAW mechanism effectively reduces unnecessary grid adaptations, ensuring that system performance remains optimal without compromising the accuracy of the frequency oracles. The entire ALOG algorithm is detailed in Algorithm 2.

We propose three variations of the ALOG approach to determine the most effective solution for anonymizing user location

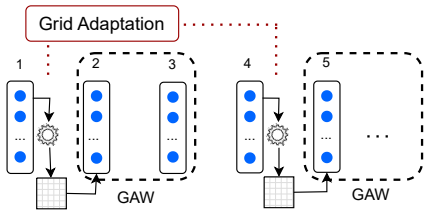


Figure 3: GAW: Grid Adaptation Window mechanism with time window of size $w = 2$

Algorithm 2: ALOG Algorithm

User-side:
Input : w *(grid adaptation window)*
Output: Report $\hat{r}_i$
for *each timestamp t* **do**
  l_i = Capture current location
  G = Receive the Grid from Server
  r_i = Encode(l_i,G)
  M = Initialize the Memoization Vector (G)
  $\hat{r}_i$ = LDP Protocol(r_i, G, M);
  if $t\%w == 0$ **then**
    M = Initialize the Memoization Vector (G);
Server-side:
Input : tr *(Cell Threshold),* w *(grid adaptation window)*
Output: Adaptive grid, Frequency Estimations
G = Initialize the initial Grid;
Process the reports to aggregate counts for each grid cell;
for *each timestamp t* **do**
  R = Receive the reports for all users
  F = Aggregate the reports to process the frequency estimation (R);
  if $t\%w == 0$ **then**
    G = Adaptive Grid Creation Algorithm (G, F, tr);

reports in the context of grid refinement and data utility. The first variation, **ALOG-2R**, implements a two-round sanitization process designed to improve both grid refinement and the utility of the data. In the first round, each user i anonymizes their location report using the initially provided uniform grid, allocating only a fraction of the available privacy budget to this step. For example, suppose a user i has a true location at (x_i, y_i), and the initially assigned grid cell is g_0. The user first encodes their location into the grid cell g_0 (e.g., by rounding their location to the nearest cell in the grid). The user then applies a Local Differential Privacy (LDP) protocol, utilizing a fraction of the total privacy budget to anonymize the encoded location. The resulting anonymized report, denoted as $\hat{r}_i$, is sent to the aggregator. This initial stage aims to provide a preliminary anonymized data sample to the aggregator.

Once the aggregator has collected the anonymized reports from all users, it proceeds to refine the spatial grid based on these reports. Specifically, the aggregator applies the Adaptive Grid Creation Algorithm (Algorithm 1) to construct a refined, adaptive grid that more accurately reflects the underlying data distribution. This adaptive grid is tailored to capture variations in the density of user locations, allowing for more precise spatial partitioning.

In the second round, the user i now anonymizes their true location using the refined grid. For instance, if the user's true location (x_i, y_i) falls in a newly created, smaller grid cell g_1 (which results from the adaptive grid refinement), the user will now encode their location into this smaller, more precise grid cell g_1. The user then applies the remaining portion of their privacy budget to anonymize the location in this refined grid cell, using the same LDP protocol as in the first stage. This refined anonymization ensures that the user's location is represented with higher precision according to the new adaptive grid. The anonymized report is then sent to the aggregator. After all users have submitted their second-round reports, the aggregator collects these updated anonymized reports and performs frequency estimation for the given timestamp. With grid refinement in the first stage and a more precise anonymization in the second stage, this two-stage process results in improved utility. By incorporating the adaptive grid refinement early on, ALOG-2R ensures that the data is processed with an optimized spatial partitioning, leading to more accurate frequency estimations while maintaining a strong privacy guarantee.

Alternatively, **ALOG-1R-a** (One-Round Sanitization — Adaptive) and **ALOG-1R-b** (One-Round Sanitization — Baseline) utilize a single-round sanitization process. In both variations, users initially report perturbed data using a uniform grid, and the adaptive grid is applied only in subsequent reports. The key distinction between ALOG-1R-a and ALOG-1R-b lies in the choice of the base grid used during the adaptation process.

In ALOG-1R-a, the base grid used for adaptation is not static. Instead, the grid from the previous adaptation step serves as the input for the Adaptive Grid Creation Algorithm. This allows for incremental refinements over time, where the grid is continuously adjusted to reflect the evolving spatial patterns of user locations. For instance, suppose in the first timestamp, a user i has a true location at (x_i, y_i), which initially maps to a grid cell g_0. After anonymizing the location using LDP and sending the anonymized data to the aggregator, the adaptive grid is refined based on the collected reports. When the user submits their report in the next timestamp, the adaptive grid generated from the prior stage is used as the base grid for further refinement. If the user's true location (x_i, y_i) now falls into a more precise, smaller grid cell g_1 in this updated grid, the user will anonymize their location based on this refined grid, improving the accuracy of the spatial partitioning with each adaptation step. This dynamic approach allows the system to adapt to changing patterns in user data, improving the overall utility while maintaining privacy.

Conversely, in ALOG-1R-b, the algorithm takes a different approach. Here, the adaptive grid is reset at every adaptation step, and the algorithm consistently uses the default uniform grid as the base grid for every iteration of the grid refinement process. In each adaptation step, the algorithm ignores any previous refinements and starts from the uniform grid, applying the Adaptive Grid Creation Algorithm as if it were the first round of processing. For example, a user i's true location (x_i, y_i) is first mapped to a grid cell g_0 of the uniform grid. After anonymizing their location using LDP and sending the anonymized report to the aggregator, the grid is adapted based on the collected data. This new adaptive grid will be used in the next timestamp for users to anonymize their location, but the grid is discarded for the next adaptation process. When the user submits their second report, the algorithm again uses the uniform grid as the base for the refinement, disregarding the previous adaptation. This results in a stable reference framework for the grid adaptation

process, which may be more predictable but does not account for changes in user spatial distributions over time.

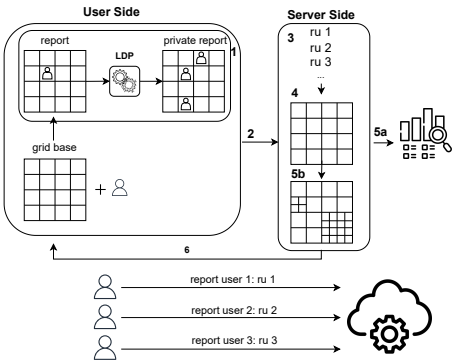


Figure 4: ALOG Workflow - Step 1: Users sanitize their own reports. Step 2: Send the sanitized reports to the aggregator. Step 3: Aggregate received reports. Step 4: Process count for each cell. Step 5a: Perform frequency estimation. Step 5b: Grid adaptation to refine spatial granularity. Step 6: Broadcast the new Grid

The workflow of ALOG is illustrated in Figure 4. The process begins with Step 1, performed on the user side, where each user encodes their location using the current base grid provided by the aggregator. This grid could either be the initial uniform grid (for ALOG-1R-b) or the refined grid from the previous moment (for ALOG-1R-a and ALOG-2R). The user applies a Local Differential Privacy (LDP) protocol to anonymize the location report in all approaches. Once anonymized, the report is transmitted securely to the aggregator, completing Step 2.

In Step 3, the aggregator collects all received reports. Step 4 follows, where the aggregator counts the occurrences within each grid cell to estimate the spatial distribution of user locations. Step 5a is executed at every timestamp, regardless of the approach. In this step, the aggregator publishes the estimated spatial distribution for each grid cell based on the received reports. This frequency estimation is critical for maintaining real-time monitoring of the data. The distinction arises when a Grid Adaptation Window (GAW) concludes. The GAW mechanism, primarily used in ALOG-2R and ALOG-1R-a, allows the grid adaptation process to occur only after a predefined number of timestamps (denoted as w). When this window ends, Step 5b is triggered. During Step 5b, the grid cells are adjusted based on report density. This refinement aims to enhance spatial resolution and precision, tailoring the grid to represent the underlying distribution of user locations better. This adjustment will occur as follows for each approach:

- ALOG-2R: Grid refinement occurs after every w timestamp, allowing for dynamic, incremental adaptations. This means that after each window of w timestamps, the aggregator adapts the grid based on the spatial distribution observed in those reports.
- ALOG-1R-a: Similarly, the grid is adapted after every w timestamps, but in this approach, the refinement is based on the previously adapted grid. The grid evolves incrementally, becoming more precise with each new adaptation.
- ALOG-1R-b: In this approach, even when the GAW ends, the grid is reset to the initial uniform grid at every adaptation step, discarding any previous refinements. This provides a stable reference framework but does not consider previous grid adaptations.

Finally, in Step 6, the aggregator broadcasts the refined grid back to the users. This ensures that users always work with an up-to-date spatial representation of the data. The users then continue the process of encoding, anonymizing, and sending reports based on the latest grid, ensuring that the system adapts to real-time patterns.

This entire workflow is designed to repeat iteratively, maintaining a dynamic balance between spatial precision and privacy guarantees. The GAW ensures that grid adaptations occur controlled, reducing unnecessary overhead, while the different approaches (ALOG-1R-a, ALOG-1R-b, and ALOG-2R) offer flexibility depending on the need for stability versus dynamic adaptation.

On the client side, the ALOG algorithm 2 has a per-timestamp time and space complexity of $O(k)$, where k is the number of grid cells. This cost stems from encoding the location as a unary vector and applying the LDP mechanism. The time complexity of the ALOG algorithm on the server side, assuming a one-round approach, can be described in terms of the number of grid cells k, the number of user reports n, and the grid adaptation window size w. For each timestamp, the server processes n reports and performs frequency estimation over k cells, leading to a per-timestamp cost of $O(nk)$. Every w timestamps, the adaptive grid is updated using a recursive algorithm, whose worst-case cost is $O(k \log n)$ due to cell subdivisions based on report density. Thus, over a time horizon T, the amortized server-side complexity becomes $O(Tnk + wTk \log n)$, balancing continuous frequency estimation with periodic grid refinement.

4.3 Privacy and Utility Analysis

In our approach, we employ LDP to protect user privacy while estimating the frequency of each grid cell in a spatial domain. Each user perturbs their location data locally before transmitting it to a centralized aggregator. This is done using the LOSUE protocol, as Section 3.2 explains.

At each timestamp, ALOG uses at most a privacy budget of ϵ_∞, even ALOG-2R. In the case of ALOG-2R, the ϵ_∞ is divided across two stages: a fraction of ϵ_∞ is spent in the first stage, and the remaining fraction is applied in the second stage. Like other data collection models, whether using adaptive grids or not, ALOG consumes a privacy budget of ϵ_∞ for a single report, which is the single report's upper bound.

Proposition 4.1 provides a pessimistic worst-case privacy guarantee for our approach. Nevertheless, our experiments demonstrate that ALOG achieves Pareto efficiency – a state where no further improvement can be made to one objective (e.g., utility) without degrading another (e.g., privacy) – even with a slightly higher theoretical privacy budget than competing methods.

Proposition 4.1 (ALOG's Privacy Guarantee). *The ALOG algorithm guarantees, in the worst case, $\eta\epsilon_\infty$-local differential privacy, where: $\eta = \left(\sum_{r=0}^{R}(k-\bar{k})\cdot 4^r + \bar{k}\right)$, $\bar{k}$ is the number of initial grid cells satisfying the threshold condition, $R = \left\lceil \log_4 \frac{|U|}{tr} \right\rceil$ is the maximum recursion depth of the Algorithm 1, ϵ_∞ is the per-report privacy budget under Ψ, the underlying LDP mechanism.*

Proof. At initialization, the grid contains k cells, of which $\bar{k}$ satisfy the threshold tr. The remaining $k - \bar{k}$ cells are recursively split (at most R times) until all subcells meet tr, as formalized in Algorithm 1. For recursion level $r \in [0, R]$, the grid contains $(k - \bar{k}) \cdot 4^r + \bar{k}$ cells. Summing over all levels, the *total cumulative cell count* is: $\sum_{r=0}^{R} \left((k - \bar{k}) \cdot 4^r + \bar{k}\right)$. Since each user's location report is privatized via Ψ (an ϵ_∞-LDP protocol), the *longitudinal privacy cost* of ALOG, accounting for adaptive grid construction and repeated reporting, is bounded by the worst-case total cell count multiplied by ϵ_∞. This holds because:

(1) *User-side perturbation*: Each report is independently ϵ_∞-LDP.
(2) *Aggregator-side adaptation*: The grid's dynamic refinement does not access raw data, relying only on privatized counts.

The total cumulative cell count reflects the worst-case scenario where Algorithm 1 exhausts all possible recursions. Each cell is sanitized only once during the longitudinal process by employing memoization in the LDP protocol. Thus, since the upper bound on cell count is η, the maximum number of sanitizations is precisely the cumulative cell count. As each report is independently ϵ_∞-LDP and there are at most η cells in the grid, our proposed algorithm is $\eta\epsilon_\infty$-local differentially private. □

In ALOG, both the frequency estimation (Section 2.2) and grid adaptation (Section 4.1) processes operate on the perturbed reports, ensuring that individual location data remains protected. As Proposition 2.3 states, any function applied to the output of an ϵ-LDP mechanism preserves ϵ-LDP privacy guarantees, meaning that the privacy established by the initial LDP perturbation is maintained throughout these post-processing steps without introducing additional privacy risks.

The granularity of the initial grid used in LDP significantly impacts data sensitivity and privacy preservation. This relationship is analyzed through variance, a key measure of frequency estimate accuracy and reliability. The variance of LDP protocols is inversely proportional to the number of times a value v_i has been reported [3], meaning that higher variance indicates greater uncertainty and error in estimates. In comparison, lower variance suggests more accurate results. Fine-grained grids yield fewer reports per cell in low-density areas, increasing variance and making estimates less reliable. Conversely, in high-density areas, fine-grained grids can maintain utility without compromising privacy, as the number of reports per cell remains high, reducing variance.

ALOG directly addresses the variance problem by using coarser grids in low-density areas, increasing reports per cell and decreasing variance. ALOG refines the grid in high-density areas to maintain utility, and the higher density helps control variance. Due to this adaptive behavior, ALOG's grids tend to keep the density between cells close to uniformity, further enhancing the accuracy of estimates while preserving privacy.

When handling longitudinal data, ALOG mitigates privacy loss through the Grid Adaptation Window (GAW). By regulating the frequency of grid adjustments, the GAW ensures that the grid adapts only after every w timestamps, limiting the reinitialization of memorization, which is a process that consumes the privacy budget. This mechanism balances the utility gains offered by dynamic grid adaptation and the potential increase in privacy loss resulting from frequent re-randomization, thus maintaining privacy protection while enhancing utility.

Table 2: Summary of Datasets Used in the Experiments

Dataset	Type	Distribution	#Trajectories
S1	Synthetic	Uniform	10,000
S2	Synthetic	Normal	10,000
GeoLife	Real-world	Real-world	6,281
Porto	Real-world	Real-world	10,000

5 Experimental Evaluation

In this section, we present the experiments conducted to empirically demonstrate that our approach not only achieves higher utility but also maintains a stronger level of privacy compared to traditional local differential privacy (LDP) protocols such as RAPPOR (Section 3.1), L-OSUE (Section 3.2), and LOLOHA (Section 3.3). We implemented our solution in Python, including our partitioning strategy and protocols.

5.1 Experimental Setup

We evaluate our approach using two synthetic datasets created for this study and two real-world mobility datasets. The synthetic datasets simulate user movement within a 100 km^2 area and allow us to assess behavior under contrasting spatial distributions.

- **S1**: Users are distributed uniformly across the region.
- **S2**: Users follow a normal distribution centered in the region.

Each synthetic dataset simulates 10,000 users moving at a constant speed of 40 km/h, generating one location report per minute over 40 timestamps. Locations are constrained by the speed to ensure realistic trajectories.

The **GeoLife** dataset[1] contains GPS trajectories from 183 users in Beijing. We extracted 6,261 trajectories of 20 timestamps each. The **Porto** dataset[2] comprises data from 442 taxis in Porto, Portugal. We selected 10,000 trajectories, also with 20 timestamps. Table 2 summarizes all datasets.

We acknowledge the use of only two real-world datasets, a limitation imposed by the scarce availability of longitudinal datasets with precise coordinate data — a key requirement for our analysis. Nevertheless, the selected datasets offer meaningful and representative insights. For each dataset, we simulate users submitting anonymized location reports using longitudinal LDP protocols. These reports are used to estimate cell frequencies over a uniform grid. We compare these results with those obtained using ALOG's three variations (ALOG-2R, ALOG-1R-a, and ALOG-1R-b) under identical conditions.

Data Utility: To assess utility, we compute the **Root Mean Squared Error (RMSE)** between the true and estimated frequency distributions over time. Let $F_i = \{f_0, \ldots, f_{|G_i|}\}$ and $\hat{F}_i = \{\hat{f}_0, \ldots, \hat{f}_{|G_i|}\}$ denote the true and estimated frequencies at timestamp t_i. The utility error is defined as:

$$\text{Utility Error} = \frac{1}{|T|} \sum_{i=1}^{|T|} \sqrt{\frac{1}{|G_i|} \sum_{j=0}^{|G_i|-1} (f_j - \hat{f}_j)^2} \tag{5}$$

This metric captures the accuracy of frequency estimates over time and across different spatial distributions.

Budget Consumption: We also evaluate the **privacy budget** consumption of each method. For each timestamp t_i, the budget consumed is at most ϵ_∞ during anonymization. The total budget

[1] https://www.microsoft.com/en-us/download/details.aspx?id=52367
[2] https://www.kaggle.com/datasets/crailtap/taxi-trajectory?select=train.csv

consumption is: $\sum_{i=1}^{|T|} \epsilon_\infty$. We report the average total budget per user to capture the cumulative privacy cost across the experiment.

5.2 Results

Grid Granularity x Utility: first, we analyze the adjustment of grid granularity, which is a key aspect of the proposed adaptive longitudinal grid framework. Specifically, we compare methods using static uniform grids with ALOG, starting from the same initial grid granularity. The relationship between grid granularity and data utility is central to understanding how the system adapts to different data distributions while balancing privacy constraints.

Theoretically, the expected utility of each ALOG variant is influenced by the trade-off between noise magnitude (from LDP perturbation) and the granularity of the grid. According to the variance of LDP frequency estimation protocols such as LOSUE, the estimation variance is inversely proportional to the number of reports per grid cell. Thus uniform grid approaches (e.g., RAPPOR) are expected to exhibit higher error in regions with skewed data distributions due to non-uniformity error.

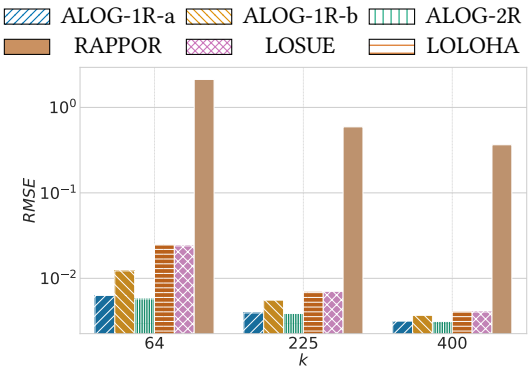
Figure 5 illustrates the effectiveness of our proposed adaptive grid approaches in enhancing utility compared to uniform grid-based methods, which consistently yielded the highest *RMSE* across all scenarios. On the x-axis of the figure, k represents the number of cells in the grid, while the y-axis shows the *RMSE* between the real and estimated frequencies. Among the adaptive grid strategies, ALOG-2R emerged as the most effective, achieving the lowest *RMSE* due to its two-round mechanism. In contrast, the ALOG-1R-b method demonstrated the poorest performance among the ALOG approaches. This is attributed to its reliance on a fixed base grid size during the grid adaptation process, which results in the loss of previously acquired knowledge and compromises its utility. For this experiment, a permanent privacy budget of 0.1 was used for each timestamp, and the adaptation window was set to 4.

Privacy Budget x Utility: analyzing the relationship between the privacy budget and utility is crucial, as it highlights the trade-off between privacy guarantees and the accuracy of data estimates. A well-calibrated privacy budget establishes a balance, minimizing the *RMSE* while ensuring adequate privacy protection. This balance is essential for practical deployments in applications involving sensitive data.

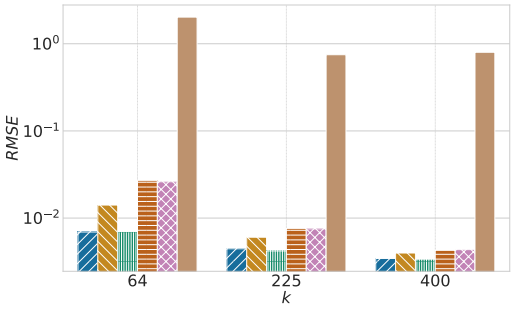
In our experiments, we varied the privacy budget values between 0 and 1, with the budget representing the upper bound of privacy for each individual. We set up a cell size of 700 meters × 700 meters, which results in an initial grid granularity of $k = 225$ for datasets S1 and S2, $k = 336$ for Geolife, and $k = 192$ for Porto. The grid adaptation window was set to $w = 4$ moments of time. This setup allows us to explore how different privacy levels affect the utility of the system while maintaining the required privacy guarantees, as shown in Table 3.

In the S1 and S2 datasets, ALOG-2R consistently achieves the lowest *RMSE* values for all privacy budgets. This demonstrates how the two-stage model effectively enhances utility while maintaining compliance with privacy requirements. ALOG-1R-a follows closely, with slightly higher *RMSE* values than ALOG-2R but significantly outperforming ALOG-1R-b.

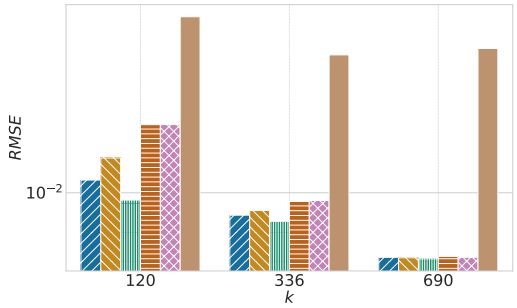
The LOLOHA and LOSUE methods, while maintaining relatively stable performance across datasets, generally exhibit higher *RMSE* values compared to ALOG-2R. Notably, RAPPOR, despite being a widely used method for privacy preservation, demonstrates the highest *RMSE* values, particularly in scenarios with small privacy budgets, indicating its limitations in achieving high utility.



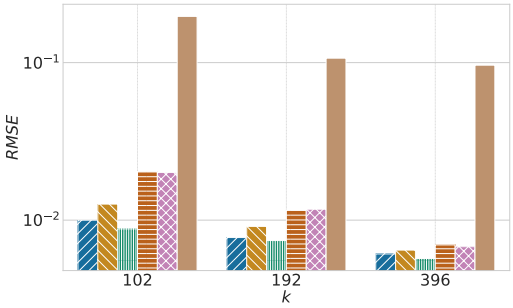
(a) S1 with $\epsilon_\infty = 0.1$



(b) S2 with $\epsilon_\infty = 0.1$



(c) Geolife with $\epsilon_\infty = 0.3$



(d) Porto with $\epsilon_\infty = 0.3$

Figure 5: RMSE variation as a function of grid size for the four datasets: S1, S2, GeoLife, and Porto. Each chart represents the error in data analysis when the initial grid size is set.

For the Geolife and Porto datasets, a similar pattern emerges. ALOG-2R continues to outperform other methods, achieving the best balance between privacy and utility. In the Geolife dataset, the *RMSE* for ALOG-2R is consistently the lowest, especially for $\epsilon = 0.05$ and $\epsilon = 0.5$. In the Porto dataset, ALOG-2R again achieves the best results, with noticeable improvements over LOLOHA and LOSUE.

In summary, Table 3 emphasizes the significance of selecting an appropriate privacy budget ϵ and algorithm based on the application requirements. The results underscore the effectiveness of ALOG-2R in achieving a superior trade-off between privacy and utility, making it a strong candidate for practical deployments in sensitive data applications.

Longitudinal Setting x Utility: another important aspect to consider is the impact of data's longitudinal nature on utility. Figure 6 illustrates how the temporal evolution of data affects the performance of privacy-preserving mechanisms, which is crucial for applications that involve dynamic and continuously changing datasets. Understanding this impact allows us to assess the robustness of adaptive grid methods, such as ALOG-2R, in maintaining low *RMSE* while addressing the challenges posed by temporal variations in data distribution. This analysis will further provide insights into how privacy budgets and grid adaptations interact over time, ensuring the methods remain effective in real-world scenarios with longitudinal data.

Figure 6 consists of four subcharts, each corresponding to a different data set. These graphs illustrate the relationship between the number of consecutive reports sent by users and the *RMSE*. In the preceding analysis, RAPPOR consistently demonstrated significantly poorer performance compared to the other methods (as seen in Figure 5 and Table 3). To provide a clearer visualization and focus on the comparative differences among the higher-performing approaches, RAPPOR was excluded from Figure 6. However, it is important to note that the trend of RAPPOR's inferior performance continues in this subsequent analysis, though not shown for visual clarity.

The ALOG approaches exhibit much better performance, with ALOG-2R and ALOG-1R-a showing very similar results, often achieving the lowest *RMSE* values. These methods leverage their adaptive mechanisms to outperform not only RAPPOR but also other approaches like LOSUE and LOLOHA, which perform moderately but fail to match the precision of ALOG methods.

The consistent superiority of ALOG-2R and ALOG-1R-a across datasets highlights their adaptability and efficiency, particularly in scenarios with a higher frequency of consecutive user requests. These results underline the importance of selecting advanced adaptive grid methods to ensure low *RMSE* in privacy-preserving systems.

Budget Consumption: we now begin a new analysis focusing on the budget privacy consumption. As shown in Table 4, the ALOG approaches demonstrate the highest budget consumption among the methods. This is expected since the grid adaptation process inherent to ALOG methods sacrifices a key efficiency mechanism available in other approaches: memoization. Memoization enables non-adaptive methods like LOLOHA, LOSUE, and RAPPOR to conserve the privacy budget by reusing previously computed information.

Despite this drawback, we now extend our analysis to explicitly evaluate privacy budget consumption using Pareto frontier analysis (Figure 7). The Pareto frontiers reveal a clear efficiency-accuracy trade-off among the evaluated methods. Methods located closer to the frontier indicate superior performance, characterized by achieving lower RMSE at minimal budget consumption.

In the synthetic datasets, LOLOHA remains close to the frontier only under conditions of very low privacy budget consumption. In contrast, adaptive methods such as ALOG-2R and ALOG-1R-a dominate all other approaches at higher budget levels, consistently achieving lower RMSE values. For the GeoLife dataset, ALOG-2R and ALOG-1R-a outperform other methods starting from privacy budget consumptions above approximately 0.35 and 0.29, respectively. Similarly, for the Porto dataset, ALOG-1R-a dominates at budget consumptions exceeding 0.9, whereas ALOG-2R surpasses all other methods once the budget exceeds 1.0.

Non-adaptive methods, including LOLOHA, LOSUE, and RAPPOR, consistently demonstrate lower budget consumption due to memoization; however, their positions farther from the Pareto frontier clearly illustrate their limitations in accuracy compared to adaptive grid-based approaches.

These results highlight that while ALOG approaches are more demanding in terms of budget consumption, their superior utility makes them highly advantageous for applications where accuracy and adaptability are critical, even in longitudinal settings.

GAW x Utility: the grid adaptation window plays a crucial role in the performance of the ALOG approaches. Since all three ALOG variants consistently outperform uniform grid approaches, it is essential to understand their behavior in terms of grid size and how it affects their utility.

Figure 8 illustrates that ALOG-2R and ALOG-1R-a maintain better utility across all tested window sizes for each dataset. However, a notable trend emerges: as the window size increases, the error for ALOG-2R and ALOG-1R-a also increases. Conversely, ALOG-1R-b demonstrates a decreasing error trend as the window size grows. This divergence arises from the distinct grid adaptation mechanisms employed by these approaches.

ALOG-1R-a and ALOG-2R experience an increase in error with larger window sizes primarily due to their reliance on the previous grid for adaptation. These methods dynamically split grid cells to increase granularity when needed but lack the ability to reduce granularity by merging cells as data density decreases. This limitation becomes more pronounced as the window size grows because larger windows lead to greater variations in data distribution over time. The inability to decrease granularity hinders their ability to adjust effectively to these abrupt changes, resulting in inefficiencies and increased error.

In contrast, ALOG-1R-b demonstrates better performance in scenarios with larger window sizes due to its use of a fixed base grid during the adaptation process. This design allows ALOG-1R-b to reset and adjust more effectively to significant changes in data distribution, making it more resilient to abrupt shifts that occur when the window size is larger. By avoiding reliance on a previous grid, ALOG-1R-b can maintain alignment with current data density, ensuring more consistent and efficient performance even under challenging conditions. This highlights the importance of adaptability in handling varying data dynamics, particularly in scenarios with large adaptation windows.

In summary, the increasing error trend for ALOG-2R and ALOG-1R-a with larger window sizes is a consequence of their

Table 3: RMSE ($\times 10^{-3}$) for $k = 225$ in S1 and S2, $k = 336$ in Geolife, and $k = 192$ in Porto, with ALOG using $w = 4$.

DataSet	Budget	ALOG-1R-a	ALOG-1R-b	ALOG-2R	LOLOHA	LOSUE	RAPPOR
S1	$\epsilon = 0.05$	4.009	5.497	**3.908**	6.948	6.924	1599.661
	$\epsilon = 0.1$	3.919	5.473	**3.866**	7.124	6.960	1068.873
	$\epsilon = 0.5$	4.125	5.417	**3.771**	6.882	6.884	153.428
	$\epsilon = 1.0$	3.998	5.406	**3.554**	6.903	6.866	67.914
S2	$\epsilon = 0.05$	4.314	5.953	**4.282**	7.391	7.506	834.638
	$\epsilon = 0.1$	4.476	5.967	**4.212**	7.516	7.536	666.694
	$\epsilon = 0.5$	4.389	6.009	**4.012**	7.522	7.488	87.250
	$\epsilon = 1.0$	4.335	5.903	**4.102**	7.749	7.489	112.595
Geolife	$\epsilon = 0.05$	7.672	8.183	**6.694**	9.053	9.162	432.922
	$\epsilon = 0.1$	**7.220**	8.151	7.513	9.042	9.028	285.988
	$\epsilon = 0.5$	7.403	8.116	**6.855**	9.001	9.190	77.613
	$\epsilon = 1.0$	**6.933**	8.154	7.044	8.937	8.996	41.224
Porto	$\epsilon = 0.05$	7.522	9.211	**7.303**	11.358	11.320	829.024
	$\epsilon = 0.1$	7.657	8.720	**7.151**	11.524	11.831	197.978
	$\epsilon = 0.5$	7.831	8.796	**6.886**	11.778	11.332	111.164
	$\epsilon = 1.0$	7.474	8.907	**6.876**	11.704	11.444	68.865

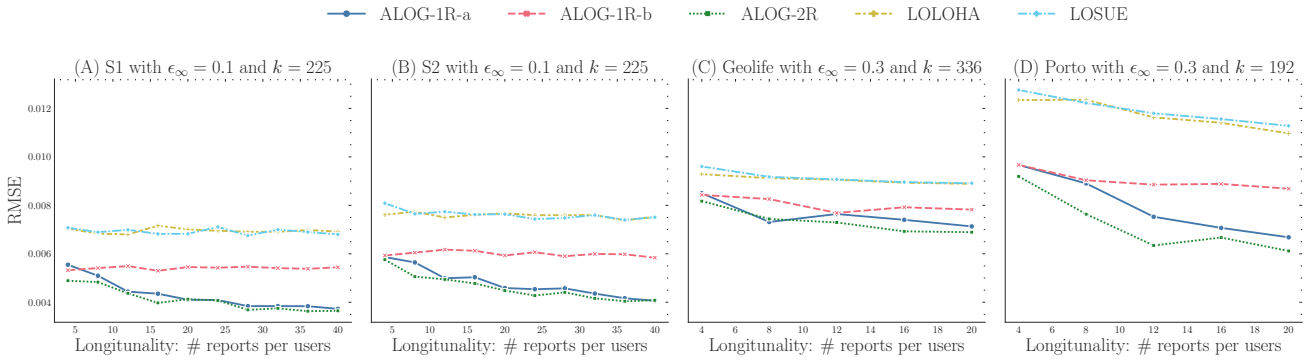


Figure 6: RMSE variation with the number of reports per user for the four datasets: S1, S2, GeoLife, and Porto.

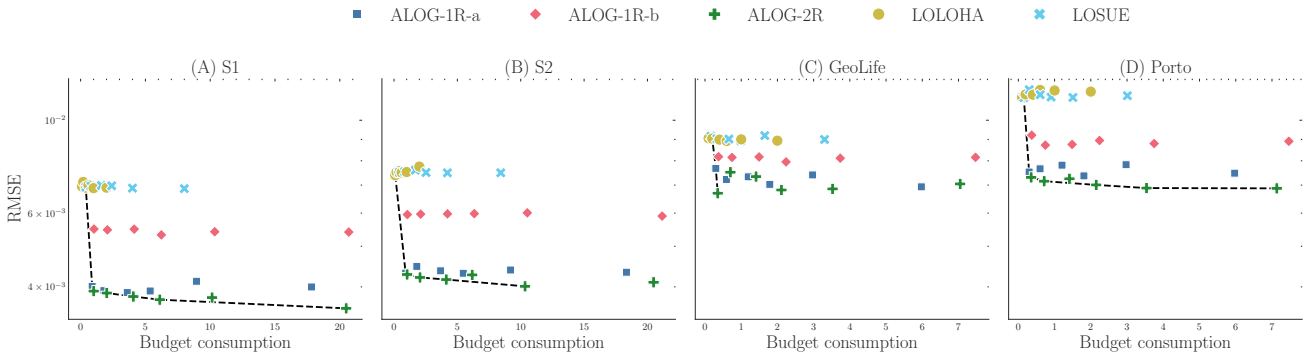


Figure 7: The Pareto frontier (black dashed line) demonstrates ALOG's superior trade-off between error (RMSE) and privacy budget consumption. Our method dominates the frontier in nearly all operational scenarios, establishing it as the preferred choice for most practical applications. Only in a narrow range of extremely low budget conditions does LOLOHA show marginally better performance.

grid adaptation process, which cannot shrink the grid when necessary. ALOG-1R-b, with its reliance on a base grid, provides better adjustment and resilience to data dynamics over longer periods, highlighting the trade-offs inherent in these adaptive strategies.

Budget partition x Utility: The final aspect of our analysis focuses on ALOG-2R, the best-performing method across datasets. A critical factor influencing its performance is the partitioning of the privacy budget between the first and second rounds. This partitioning directly affects the utility achieved by ALOG-2R, as

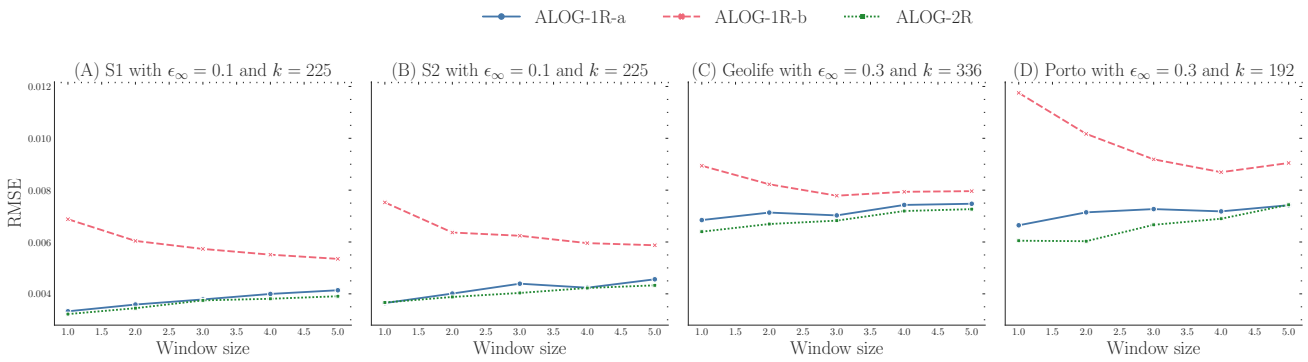


Figure 8: RMSE variation in terms of the grid adaptation window size for the four datasets: S1, S2, GeoLife, and Porto.

Table 4: Budget Consumption for Datasets S1, S2, Geolife and Porto

DataSet	ALOG 1R-a	ALOG 1R-b	ALOG 2R	LOLOHA	LOSUE	RAPPOR
S1	**1.78**	2.07	2.02	**0.2**	0.80	0.80
S2	**1.84**	2.11	2.05	**0.2**	0.84	0.86
Geolife	**1.78**	2.24	2.10	**0.59**	0.99	0.99
Porto	**1.78**	2.07	2.02	**0.2**	0.801	0.801

Table 5: RMSE as a function of budget proportion used in ALOG-2R for rounds 1 and 2 (values are in units of $\times 10^{-3}$)

DataSet	S1	S2	Geolife	Porto
$(\mathbf{r_1 = 0.1\epsilon, r_2 = 0.9\epsilon})$	3.88	4.18	7.07	6.96
$(\mathbf{r_1 = 0.3\epsilon, r_2 = 0.7\epsilon})$	**3.82**	**4.06**	7.07	**6.87**
$(\mathbf{r_1 = 0.5\epsilon, r_2 = 0.5\epsilon})$	3.92	4.08	**7.06**	7.08
$(\mathbf{r_1 = 0.7\epsilon, r_2 = 0.3\epsilon})$	3.87	4.12	7.07	7.39

it determines the extent to which each round can contribute to the grid's adaptability and the accuracy of the results.

In the first round, the budget is primarily used to initialize the grid structure and adapt it to the data distribution. This stage is crucial for ensuring that the grid reflects the overall data density and establishes a strong foundation for further refinements. The second round leverages the remaining budget to refine the grid further and address more localized variations in the data.

Table 5 reveals that for datasets S1, S2, and Porto, the optimal partitioning allocates 30% of the total budget to the first round, with the remaining 70% allocated to the second round. This configuration ensures sufficient budget for initializing the grid while allowing for meaningful refinements in the second round.

Conversely, the Geolife dataset exhibits only minor differences in RMSE across the tested budget partitioning strategies. The optimal configuration, with 50% of the budget assigned to each of the two allocation stages, results in a minimal improvement in RMSE. This suggests that the specific characteristics of the Geolife data permit a wider range of effective privacy budget distributions. In summary, the results emphasize the robustness of ALOG-2R in delivering superior utility through its adaptive grid mechanism, making it a strong candidate for practical applications in privacy-preserving data analysis. The findings also underscore the importance of tuning parameters such as grid adaptation windows and budget partitioning to maximize performance across varying scenarios.

6 Conclusion

In this work, we proposed ALOG, a framework for privacy preserving frequency estimation using adaptive grids under Local Differential Privacy (LDP). Through extensive evaluations across synthetic and real-world datasets, we demonstrated the superiority of ALOG approaches, particularly ALOG-2R, in balancing utility and privacy compared to traditional methods such as RAPPOR, LOLOHA, and LOSUE. The results highlighted the effectiveness of adaptive grids in dynamically adjusting to varying data densities, significantly reducing MSE while maintaining robust privacy guarantees.

Key findings include the importance of selecting appropriate parameters, such as grid adaptation windows and privacy budget partitioning. ALOG-2R consistently delivered superior performance, with optimal budget partitioning strategies (e.g., 30%-70% for most datasets) and smaller grid adaptation windows yielding the best utility. However, a notable limitation of the current ALOG approaches is their inability to reduce grid granularity when data density decreases. This results in granular grids that can lead to increased error over extended periods.

In real-world scenarios, where the distribution of reports evolves throughout the day—such as during peak and off-peak hours in urban settings—it is critical to overcome this limitation to provide a truly adaptive model. Addressing this challenge would enable the grid to dynamically adjust to both increases and decreases in data density, ensuring optimal utility and efficient resource usage.

Future Works: While ALOG shows strong performance, further work is needed to explore several key directions. One is automating parameter tuning—using machine learning to optimize budget partitioning and window size could enhance utility and efficiency across varying datasets. Enhancing granularity adjustment through dynamic cell merging and splitting will directly overcome ALOG-2R and ALOG-1R-a's current limitations, enabling more adaptive data representation. These directions could further improve the balance among privacy, utility, and adaptability in privacy-preserving geospatial analysis.

Acknowledgements

This research was partially supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior CAPES (grant # 88882.454568/2019-01) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq (grant # 316729/2021-3). It was partially funded by Lenovo, as part of its R&D investment under Brazilian Informatics Law.

7 Artifacts

The artifacts for reproducing our experiments are available at:

https://github.com/edurdneto/ALOG

The repository includes source code, method implementations, experiment scripts, and reproduction instructions.

References

[1] Miguel E. Andrés, Nicolás Emilio Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2013. Geo-indistinguishability: differential privacy for location-based systems. In *2013 ACM SIGSAC Conference on Computer and Communications Security, CCS'13, Berlin, Germany, November 4-8, 2013*, Ahmad-Reza Sadeghi, Virgil D. Gligor, and Moti Yung (Eds.). ACM, 901–914. doi:10.1145/2508859.2516735

[2] Héber Hwang Arcolezi, Jean-François Couchot, Bechara al Bouna, and Xiaokui Xiao. 2024. Improving the utility of locally differentially private protocols for longitudinal and multidimensional frequency estimates. *Digit. Commun. Networks* 10, 2 (2024), 369–379. doi:10.1016/J.DCAN.2022.07.003

[3] Héber Hwang Arcolezi, Carlos Pinzón, Catuscia Palamidessi, and Sébastien Gambs. 2023. Frequency Estimation of Evolving Data Under Local Differential Privacy. In *Proceedings 26th International Conference on Extending Database Technology, EDBT 2023, Ioannina, Greece, March 28-31, 2023*, Julia Stoyanovich, Jens Teubner, Nikos Mamoulis, Evaggelia Pitoura, Jan Mühlig, Katja Hose, Sourav S. Bhowmick, and Matteo Lissandrini (Eds.). OpenProceedings.org, 512–525. doi:10.48786/EDBT.2023.44

[4] Yogesh Bali, Vijay Pal Bajiya, Jai Prakash Tripathi, and Anuj Mubayi. 2024. Exploring data sources and mathematical approaches for estimating human mobility rates and implications for understanding COVID-19 dynamics: a systematic literature review. *Journal of Mathematical Biology* 88, 6 (2024), 1–39.

[5] Wenxuan Bao, Adu Gong, Tong Zhang, Yiran Zhao, Boyi Li, and Shuaiqiang Chen. 2023. Mapping population distribution with high spatiotemporal resolution in Beijing using Baidu heat map data. *Remote Sensing* 15, 2 (2023), 458.

[6] Raef Bassily and Adam D. Smith. 2015. Local, Private, Efficient Protocols for Succinct Histograms. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, Rocco A. Servedio and Ronitt Rubinfeld (Eds.). ACM, 127–135. doi:10.1145/2746539.2746632

[7] Xiang Cheng, Luoyang Fang, Liuqing Yang, and Shuguang Cui. 2017. Mobile big data: The fuel for data-driven wireless. *IEEE Internet of things Journal* 4, 5 (2017), 1489–1516.

[8] Graham Cormode, Samuel Maddock, and Carsten Maple. 2021. Frequency Estimation under Local Differential Privacy [Experiments, Analysis and Benchmarks]. *CoRR* abs/2103.16640 (2021). arXiv:2103.16640 https://arxiv.org/abs/2103.16640

[9] Rachel Cummings, Damien Desfontaines, David Evans, Roxana Geambasu, Yangsibo Huang, Matthew Jagielski, Peter Kairouz, Gautam Kamath, Sewoong Oh, Olga Ohrimenko, et al. 2023. Advancing differential privacy: Where we are now and future directions for real-world deployment. *arXiv preprint arXiv:2304.06929* (2023).

[10] Teddy Cunningham, Graham Cormode, Hakan Ferhatosmanoglu, and Divesh Srivastava. 2021. Real-World Trajectory Sharing with Local Differential Privacy. *Proc. VLDB Endow.* 14, 11 (2021), 2283–2295. doi:10.14778/3476249.3476280

[11] José Serafim da Costa Filho and Javam C. Machado. 2023. FELIP: A local Differentially Private approach to frequency estimation on multidimensional datasets. In *Proceedings 26th International Conference on Extending Database Technology, EDBT 2023, Ioannina, Greece, March 28-31, 2023*, Julia Stoyanovich, Jens Teubner, Nikos Mamoulis, Evaggelia Pitoura, Jan Mühlig, Katja Hose, Sourav S. Bhowmick, and Matteo Lissandrini (Eds.). OpenProceedings.org, 671–683. doi:10.48786/EDBT.2023.56

[12] Statista Research Department. 2015. Share of U.S. smartphone Owners Using Geosocial and Location-Based Services from 2011 to 2015. https://www.statista.com/statistics/224949/mobile-geosocial-and-location-based-service-usage-by-age/. Accessed: Nov 18, 2024.

[13] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. 2017. Collecting Telemetry Data Privately. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. Neural Information Processing Systems Foundation, 3571–3580. https://proceedings.neurips.cc/paper/2017/hash/253614bbac999b38b5b60cae531c4969-Abstract.html

[14] Cynthia Dwork. 2006. Differential Privacy. In *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 4052)*, Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener (Eds.). Springer, 1–12. doi:10.1007/11787006_1

[15] Cynthia Dwork and Aaron Roth. 2014. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* 9, 3-4 (2014), 211–407. doi:10.1561/0400000042

[16] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*, Gail-Joon Ahn, Moti Yung, and Ninghui Li (Eds.). ACM, 1054–1067. doi:10.1145/2660267.2660348

[17] Daeyoung Hong, Woohwan Jung, and Kyuseok Shim. 2021. Collecting Geospatial Data with Local Differential Privacy for Personalized Services. In *37th IEEE International Conference on Data Engineering, ICDE 2021, Chania, Greece, April 19-22, 2021*. IEEE, 2237–2242. doi:10.1109/ICDE51399.2021.00230

[18] Ilyoung Hong and Jin-Kyu Jung. 2017. What is so "hot" in heatmap? Qualitative code cluster analysis with foursquare venue. *Cartographica: The International Journal for Geographic Information and Geovisualization* 52, 4 (2017), 332–348.

[19] Haichao Huang, Zhi-heng Chen, Bo-wen Li, Qing-hai Ma, and Hong-Di He. 2024. FeSTGCN: A frequency-enhanced spatio-temporal graph convolutional network for traffic flow prediction under adaptive signal timing. *Appl. Intell.* 54, 6 (2024), 4848–4864. doi:10.1007/S10489-024-05401-5

[20] Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2011. What can we learn privately? *SIAM J. Comput.* 40, 3 (2011), 793–826.

[21] Jongwook Kim. 2024. Improving Data Utility in Privacy-Preserving Location Data Collection via Adaptive Grid Partitioning. *Electronics* 13, 15 (2024), 3073.

[22] Jong Wook Kim, Dae-Ho Kim, and Beakcheol Jang. 2018. Application of Local Differential Privacy to Collection of Indoor Positioning Data. *IEEE Access* 6 (2018), 4276–4286. doi:10.1109/ACCESS.2018.2791588

[23] Ala Eddine Laouir and Abdessamad Imine. 2024. DiApprox: Differential Privacy-based Online Range Queries Approximation for Multidimensional Data. In *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing, SAC 2024, Avila, Spain, April 8-12, 2024*, Jiman Hong and Juw Won Park (Eds.). ACM, 337–344. doi:10.1145/3605098.3636070

[24] Àlex Miranda-Pascual, Patricia Guerra-Balboa, Javier Parra-Arnau, Jordi Forné, and Thorsten Strufe. 2023. SoK: Differentially Private Publication of Trajectory Data. *Proc. Priv. Enhancing Technol.* 2023, 2 (2023), 496–516. doi:10.56553/POPETS-2023-0065

[25] Hafsa Ouchra, Abdessamad Belangour, and Allae Erraissi. 2023. An overview of GeoSpatial Artificial Intelligence technologies for city planning and development. In *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. IEEE, IEEE, 1–7.

[26] Michael P Peterson. 2014. *Mapping in the Cloud*. Guilford Publications.

[27] Wahbeh H. Qardaji, Weining Yang, and Ninghui Li. 2013. Differentially private grids for geospatial data. In *29th IEEE International Conference on Data Engineering, ICDE 2013, Brisbane, Australia, April 8-12, 2013*, Christian S. Jensen, Christopher M. Jermaine, and Xiaofang Zhou (Eds.). IEEE Computer Society, 757–768. doi:10.1109/ICDE.2013.6544872

[28] Wahbeh H. Qardaji, Weining Yang, and Ninghui Li. 2013. Differentially private grids for geospatial data. In *29th IEEE International Conference on Data Engineering, ICDE 2013, Brisbane, Australia, April 8-12, 2013*, Christian S. Jensen, Christopher M. Jermaine, and Xiaofang Zhou (Eds.). IEEE Computer Society, 757–768. doi:10.1109/ICDE.2013.6544872

[29] Zhan Qin, Yin Yang, Ting Yu, Issa Khalil, Xiaokui Xiao, and Kui Ren. 2016. Heavy Hitter Estimation over Set-Valued Data with Local Differential Privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*, Edgar R. Weippl, Stefan Katzenbeisser, Christopher Kruegel, Andrew C. Myers, and Shai Halevi (Eds.). ACM, 192–203. doi:10.1145/2976749.2978409

[30] Aleksandar Rikalovic, Gerson Antunes Soares, and Jelena Ignjatić. 2018. Spatial analysis of logistics center location: A comprehensive approach. *Decision Making: Applications in Management and Engineering* 1, 1 (2018), 38–50.

[31] Shashi Shekhar, Michael R Evans, James M Kang, and Pradeep Mohan. 2011. Identifying patterns in spatial information: A survey of methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1, 3 (2011), 193–214.

[32] Peipei Shi, Yinghui Xiao, and Qingming Zhan. 2020. A study on spatial and temporal aggregation patterns of urban population in Wuhan City based on Baidu heat map and POI data. *International Review for Spatial Planning and Sustainable Development* 8, 3 (2020), 101–121.

[33] Peipei Shi, Yinghui Xiao, and Qingming Zhan. 2020. A study on spatial and temporal aggregation patterns of urban population in Wuhan City based on Baidu heat map and POI data. *International Review for Spatial Planning and Sustainable Development* 8, 3 (2020), 101–121.

[34] Reka Solymosi, David Buil-Gil, Vania Ceccato, Eon Kim, and Ulf Jansson. 2023. Privacy challenges in geodata and open data. *Area* 55, 4 (2023), 456–464.

[35] Ekin Tire and Mehmet Emre Gursoy. 2024. Answering Spatial Density Queries Under Local Differential Privacy. *IEEE Internet Things J.* 11, 10 (2024), 17419–17436. doi:10.1109/JIOT.2024.3357570

[36] Jayakrishnan Unnikrishnan and Farid Movahedi Naini. 2013. De-anonymizing private data by matching statistics. In *51st Annual Allerton Conference on Communication, Control, and Computing, Allerton 2013, Allerton Park & Retreat Center, Monticello, IL, USA, October 2-4, 2013*. IEEE, 1616–1623. doi:10.1109/ALLERTON.2013.6736722

[37] Israel C. Vidal, André Luis Mendonça, Franck Rousseau, and Javam de Castro Machado. 2020. ProTECting: An Application of Local Differential Privacy for IoT at the Edge in Smart Home Scenarios. In *XXXVIII Brazilian Symposium on Computer Networks and Distributed Systems, SBRC 2020, Rio de Janeiro, Brazil (virtual), December 7-10, 2020*, Marcelo Gonçalves Rubinstein and Anelise Munaretto (Eds.). Sociedade Brasileira de Computação (SBC), Fortaleza, Ceará, Brazil, 547–560. doi:10.5753/SBRC.2020.12308

[38] Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. 2017. Locally Differentially Private Protocols for Frequency Estimation. In *26th USENIX Security Symposium, USENIX Security 2017, Vancouver, BC, Canada, August 16-18, 2017*, Engin Kirda and Thomas Ristenpart (Eds.). USENIX Association, 729–745. https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/wang-tianhao

[39] Jianhao Wei, Yaping Lin, Xin Yao, and Jin Zhang. 2019. Differential privacy-based location protection in spatial crowdsourcing. *IEEE Transactions on Services Computing* 15, 1 (2019), 45–58.

[40] Yan Yan, Zichao Sun, Adnan Mahmood, Fei Xu, Zhuoyue Dong, and Quan Z Sheng. 2022. Achieving differential privacy publishing of location-based statistical data using grid clustering. *ISPRS International Journal of Geo-Information* 11, 7 (2022), 404.

[41] Mengmeng Yang, Taolin Guo, Tianqing Zhu, Ivan Tjuawinata, Jun Zhao, and Kwok-Yan Lam. 2024. Local differential privacy and its applications: A comprehensive survey. *Comput. Stand. Interfaces* 89 (2024), 103827. doi:10.1016/J.CSI.2023.103827

[42] Yutong Ye, Min Zhang, and Dengguo Feng. 2021. Collecting Spatial Data Under Local Differential Privacy. In *17th International Conference on Mobility, Sensing and Networking, MSN 2021, Exeter, United Kingdom, December 13-15, 2021*. IEEE, 120–127. doi:10.1109/MSN53354.2021.00032

[43] Jia Yu, Zongsi Zhang, and Mohamed Sarwat. 2018. GeoSparkViz: a scalable geospatial data visualization framework in the apache spark ecosystem. In *Proceedings of the 30th International Conference on Scientific and Statistical Database Management, SSDBM 2018, Bozen-Bolzano, Italy, July 09-11, 2018*, Dimitris Sacharidis, Johann Gamper, and Michael H. Böhlen (Eds.). ACM, 15:1–15:12. doi:10.1145/3221269.3223040

[44] Youwen Zhu, Yiran Cao, Qiao Xue, Qihui Wu, and Yushu Zhang. 2024. Heavy Hitter Identification Over Large-Domain Set-Valued Data With Local Differential Privacy. *IEEE Transactions on Information Forensics and Security* 19 (2024), 414–426. doi:10.1109/TIFS.2023.3324726