Location Privacy Protection through Road Network Adaptability and User Mobility Prediction

Lara Santos
Department of Computer Science
University of Porto
Porto, Portugal
up202005830@fc.up.pt

Mariana Cunha
CRACS/INESCTEC, CISUC, and
Department of Computer Science
University of Porto
Porto, Portugal
mariana.cunha@fc.up.pt

João P. Vilela
CRACS/INESCTEC, CISUC, and
Department of Computer Science
University of Porto
Porto, Portugal
jvilela@fc.up.pt

***Abstract*—The widespread collection and sharing of location data enable a wide range of location-based services but also raise significant privacy concerns, as mobility traces can reveal highly sensitive personal information. Geo-Indistinguishability has emerged as a principled approach to location privacy by adding controlled noise to users' positions. However, existing mechanisms typically rely on fixed privacy budgets or adapt them based solely on past or current locations, while often ignoring both future mobility patterns and road network constraints.**

In this paper, we propose location privacy-preserving mechanisms that leverage the structure of road networks, as well as mobility prediction, to improve the achieved privacy-utility trade-off. To do so, we developed two novel complementary approaches that: (i) adapt the privacy budget dynamically based on the prediction of future locations, and (ii) aggregate locations according to the proximity of their predicted future positions. Experimental results show that incorporating predictability of upcoming locations enables more effective privacy budget allocation, improves utility, and increases resilience against location prediction attacks. These findings highlight prediction-aware obfuscation as a promising direction for enhancing Geo-Indistinguishability-based location privacy mechanisms.

***Index Terms*—Location Privacy, Privacy-Preserving Mechanisms, Location-Based Applications, Geo-Indistinguishability**

1. Introduction

Over the past decade, the rapid growth of smartphones, mobile applications, and IoT devices has made real-time GPS-based services ubiquitous, spanning navigation, ride-sharing, fitness tracking, and context-aware applications. In this ecosystem, location data has become a valuable asset that enables for a multitude of personalized services. However, location information is also inherently sensitive: mobility traces can reveal routines, home/work locations, and visits to sensitive points of interest that may expose political, religious, or health-related attributes [1]–[4]. This dual role - high utility and high sensitivity - places location data at the center of ongoing ethical and regulatory debates.

Recent incidents illustrate the consequences of opaque or unexpected location collection. In November 2022, a coalition of 40 U.S. state attorneys general announced a $391.5M settlement with Google regarding allegedly misleading location-tracking controls, where users could believe tracking was disabled (e.g., via "Location History") while location data remained collectable through other settings such as "Web & App Activity" [5]. In a different context, investigations into a May 2024 pro-Palestine sit-in at the University of Melbourne found that Wi-Fi connection data was used alongside other identifying sources to track participants, and later scrutiny concluded that affected individuals were not adequately informed of this surveillance use [6]. Although distinct in scope and setting, these cases reinforce a common point: when raw mobility data is collected and retained, it can be repurposed for profiling or monitoring in ways that exceed users' expectations [7]. Mechanisms that reduce precision or otherwise protect location data at the point of disclosure can, therefore, limit downstream surveillance and misuse.

To address these challenges, the literature proposes a range of Location Privacy-Preserving Mechanisms (LPPMs), including obfuscation and perturbation methods [1], [8]. A prominent approach is Geo-Indistinguishability, which adapts differential privacy to geographic data by bounding how distinguishable nearby locations are [9]. Subsequent work has explored improvements such as context-aware budget allocation [10], [11], clustering to reduce cumulative privacy loss across repeated reports [12], and approaches that constrain outputs to feasible regions (e.g., road-network-aware mechanisms) [13], [14] to preserve utility in services that depend on transportation infrastructure. Despite these advances, two limitations remain particularly relevant for real-world deployments. First, many mechanisms do not explicitly account for the predictability of human mobility. Given that adversaries can perform trajectory prediction attacks [15], robustness against prediction-based inference is critical. Second, approaches that ignore the road network structure can generate implausible outputs (e.g., off-road points), degrading service quality and potentially introducing artifacts exploitable by adversaries [13].

This work advances along two complementary directions, both grounded in a road-network-aware obfuscation design: (i) we investigate how mobility prediction can inform adaptive privacy levels, strengthening protection in highly routinary contexts [16] while avoiding unnecessary utility loss when predictability is low; and (ii) we examine

whether prediction can reduce how frequently obfuscation must be applied, which is important because repeated independent disclosures can accumulate privacy loss [17] and inadvertently aid inference. Together, these directions aim to improve the privacy-utility balance for users of location-based services under realistic constraints.

The obtained results show that incorporating predictability positively impacts LPPM performance, since (1) using location prediction as a criterion for aggregation-based protection mechanisms provides privacy benefits, and (2) leveraging location predictability as a criterion for adapting the privacy budget yields more favorable privacy-utility trade-offs.

2. Background

This section presents an overview of location privacy, including existing Location Privacy-Preserving Mechanisms (LPPMs), relevant attack models, and commonly used evaluation metrics. It also reviews mobility prediction methods and discusses how such methods can affect the design and effectiveness of location privacy mechanisms.

2.1. Location Privacy

The widespread adoption of smartphones and GPS has enabled a broad ecosystem of Location-Based Services (LBSs), including navigation, social networking, and context-aware applications [18]. To deliver personalized results, LBSs typically collect spatio-temporal information that may include a user identifier (explicit or implicit), a position (e.g., coordinates or semantic places), and timestamps. This information can be reported as isolated samples or as trajectories, and may be disclosed in real time or published afterward. The typical LBS architecture therefore comprises a positioning subsystem, wireless connectivity, an LBS provider, and, in privacy-aware deployments, an intermediary component responsible for applying LPPMs [18].

Location data is highly sensitive because it can be exploited beyond the intended service purpose, including profiling, behavioral targeting, and resale. Moreover, mobility traces enable inference of private attributes through visits to sensitive points of interest (POIs) [1]. They also facilitate re-identification due to the distinctiveness of human mobility: a small number of spatio-temporal samples can uniquely single out individuals in large datasets [19]. These risks motivate the notion of location privacy as the ability to move in public without systematic, covert recording and subsequent misuse of one's whereabouts [20]. In many threat models, the adversary is assumed honest-but-curious: the service behaves correctly but leverages received data to infer identity, habits, or future activity [1].

Location privacy attacks are commonly grouped into (i) identity (deanonymization) and (ii) localization attacks [18]. Identity attacks attempt to map traces to individuals or infer relationships (e.g., meeting disclosure), often using auxiliary knowledge such as demographics, home/work priors, or external datasets. Localization attacks aim to recover places and times, ranging from identifying significant locations (e.g., home) to presence/absence disclosure and continuous tracking [18]. Importantly, attackers may combine released LBS data with background knowledge (e.g., user density, social relations, mobility statistics, or side channels such as transactions) to strengthen inference [8], [18].

To mitigate these threats while retaining service utility, LPPMs transform or limit location disclosures. Prior work typically categorizes LPPMs into cryptographic approaches, anonymization-based schemes (e.g., cloaking/k-anonymity), reduced sharing, and obfuscation mechanisms [18]. Obfuscation is widely adopted in practice [8] and includes dummy-based techniques (i.e. sending multiple candidate locations) and perturbation (i.e. adding noise). Differential privacy provides a formal basis for perturbation via privacy budgets [21], [22]; its adaptation to spatial data through Geo-Indistinguishability binds indistinguishability to geographic distance [9]. However, repeated reporting can rapidly consume privacy budget and correlations across updates can weaken protection, motivating adaptive and context-aware variants (e.g., based on correlation or velocity) [10], [11], [17]. Additionally, Euclidean perturbation may yield infeasible outputs for road-network-dependent services, leading to road-aware mechanisms that incorporate shortest-path structure or graph-based formulations [13], [14]. In particular, Geo-Graph-Indistinguishability [13] is an extension of Geo-Indistinguishability that introduces the Graph-Exponential Mechanism (GEM), which takes into account the shortest-path distances within the road network. GEM outputs a distribution of obfuscated locations that lay on the road network so that the output of the obfuscation is in the road network.

Evaluation of LPPMs typically considers privacy, utility, and performance [8]. Privacy metrics include anonymity-oriented measures, Expected Estimation Error, entropy-based uncertainty, and Geo-Indistinguishability; each captures different adversarial assumptions and may under- or over-estimate privacy when background knowledge or spatial proximity is not properly modeled [8]. Utility is often quantified through Quality Loss (e.g., distance error), though application-tailored metrics can be more representative (e.g., time impact in ride-hailing) [23]. Performance is commonly assessed via computational, communication, and storage overhead [8].

A key open challenge in real-time protection is that LPPMs typically rely only on current and past observations; incorporating mobility prediction can enable more proactive and adaptive privacy-utility control by anticipating likely future locations [1].

2.2. Mobility Prediction and Its Role in Location Privacy

Human mobility exhibits substantial regularity, which makes it highly predictable and, consequently, a privacy risk. Empirical evidence shows that movement often follows shortest-path tendencies, is spatially confined around a few anchor places (e.g., home/work), and is repetitive across daily and weekly routines, with strong time-of-day and day-of-week effects [16]. Because predictability strengthens inference attacks, mobility prediction is both

a tool for attackers and an instrument that privacy mechanisms can exploit to proactively manage the privacy-utility trade-off.

2.2.1. Mobility Prediction Approaches. Mobility prediction is commonly framed as either next-location prediction or trajectory prediction [24], [25]. Next-location prediction estimates the immediate next step (or a destination within a finite candidate set), whereas trajectory prediction forecasts a sequence of future positions (multi-step). The difficulty of both tasks depends on whether movement is network-constrained (e.g., road graphs) or unconstrained, the prediction horizon, behavioral variability, and data quality [24], [25].

Classical next-location approaches include content- and distribution-based methods, typically instantiated as (higher-order) Markov models or probabilistic processes. These methods are interpretable and lightweight but often suffer from sparsity, limited generalization, and sensitivity to modeling assumptions [26], [27]. Pattern-mining techniques predict future movement by extracting frequent or periodic mobility motifs, but their reliance on stable patterns can underperform when users exhibit diverse or evolving behaviors [25], [28]. More recent approaches enrich prediction with context: preference-based models capture individual tendencies [29], social-relation models leverage ties to mitigate cold-start and sparsity [30], and time-dependent models explicitly encode periodicity and temporal dynamics [31].

State-of-the-art performance in both next-location and trajectory prediction is increasingly driven by representation learning and deep sequence modeling. Recurrent architectures (LSTM/GRU) and Transformer-based models capture long-range spatiotemporal dependencies and typically outperform traditional baselines on multi-step tasks [24], [32]. Seq2Seq designs are particularly effective for trajectory prediction because they learn compact trajectory representations and can model dependencies across multiple future steps [32]. Recent work further improves generalization via attention mechanisms and contrastive/self-supervised pre-training, yielding stronger embeddings and improved downstream prediction accuracy [33]. Semantic-based models extend this line by incorporating contextual descriptors of places and activities, improving both accuracy and interpretability when semantic annotations are available [34].

2.2.2. LPPMs and Mobility Prediction. Mobility prediction can be incorporated into LPPMs to better balance privacy and utility. A representative example is Adaptive Geo-Indistinguishability, which adjusts noise based on the correlation between past releases and the next location, using predictability as a signal to tune protection strength [10]. More generally, the integration differs depending on whether the LPPM leverages next-location or trajectory prediction. Next-location prediction is typically more accurate but still consumes privacy budget at each release; in contrast, trajectory prediction enables planning protection across a horizon, but introduces error accumulation that can degrade utility and weaken protection if the predicted path diverges.

This distinction is particularly relevant for Geo-Indistinguishability-based mechanisms, where repeated disclosures compose and can rapidly exhaust a fixed privacy budget [9], [35]. Prediction-aware LPPMs can allocate budget non-uniformly, e.g., strengthening protection for highly predictable segments or near sensitive locations, while relaxing protection when predictability is low to preserve utility. Existing approaches illustrate this direction, for instance by estimating per-point predictability (via trajectory prediction) and adapting protection accordingly to resist prediction attacks [15]. However, important challenges remain: many schemes ignore road-network constraints, potentially producing infeasible outputs for network-dependent LBSs, and some rely on Markov-family predictors that may overfit and fail to capture complex spatiotemporal dependencies [26], [33]. These limitations motivate more robust, context-aware integrations of modern mobility prediction with LPPMs.

3. Methodology

The proposed methodology integrates mobility prediction with location obfuscation through two novel approaches: Predictive Aggregation (GEM-PA) and with Adaptive Epsilon (GEM-AE). Both rely on the Graph-Exponential Mechanism (GEM) [13], to ensure that they account for the underlying road network when adding noise to a reported location, and a predictability metric derived from future trajectory estimates.

The proposed methodology consists of a mobility prediction step, followed by the incorporation of mobility prediction into the proposed Location Privacy-Preserving Mechanism, explained as follows.

3.1. Mobility Prediction and Predictability Estimation

We employ a Multi-Step Seq2Seq LSTM framework [32] to forecast the next $k = 5$ user locations. By default the proposed LSTM-based Seq2Seq is deterministic at inference: with dropout disabled and fixed weights, repeated forward passes on the same input produce the same output (up to negligible numerical noise). Due to this, to be able to quantify prediction uncertainty, we implement Monte Carlo (MC) dropout [36] by performing $T = 50$ stochastic forward passes at inference time.

For each prediction step s, we compute the coordinate-wise variances $v_{x,s}$ and $v_{y,s}$. These coordinate-wise variances are then summed, yielding a single total variance per step $V_s = v_{x,s} + v_{y,s}$.

To ensure comparability across sequences, we rescale the variance using the reciprocal of the median total variance, r_i. The final predictability score $p_s \in [0, 1]$ is defined as:

$$p_s = \frac{1}{1 + r_i V_s} \tag{1}$$

where $p_s \approx 1$ indicates low dispersion of MC predictions and $p_s \approx 0$ indicates high dispersion.

In this way, the predictability score captures the degree of confidence in forecasting the user's next position: high score correspond to highly predictable movements, while low scores reflect greater uncertainty in the trajectory.

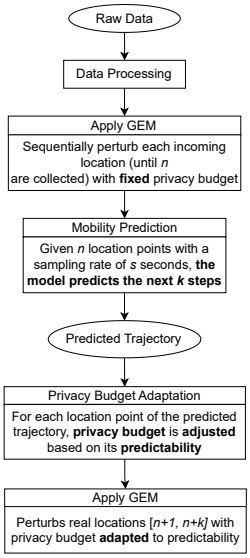

Figure 1. Overview of the Adaptive Privacy Budget (GEM-AE) proposal.

3.2. Proposed Methods

In addition to incorporating the ability to predict the next steps of a trajectory, the proposed Location Privacy-Preserving Mechanism also accounts for the underlying road network when introducing noise into the reported locations.

To achieve this, the privacy-preserving mechanism implemented in this work uses Graph-Exponential Mechanism, introduced by Takagi et al. [13]. This approach extends traditional Geo-Indistinguishability by incorporating the underlying road network, thereby achieving a more favorable balance between privacy and utility. In practice, it introduces carefully calibrated noise to reported locations while respecting the graph structure of streets and intersections.

Taking into account the structure of the road network, we propose two novel methods that incorporate the prediction of future locations. In particular, we propose (1) the GEM-AE method that dynamically adapts the privacy budget according to predicted mobility, and (2) a predictive aggregation method GEM-PA that aggregates nearby points along the trajectory to address the privacy-loss due to repeated obfuscations.

3.2.1. Adaptation of the Privacy Budget - GEM-AE. In Graph-Exponential Mechanism with Adaptive Epsilon (GEM-AE), the privacy budget ϵ is not fixed but dynamically adjusted according to the predictability of the user location where it will be applied. This method is illustrated in Figure 1.

The motivation behind this approach is that a fixed privacy budget may lead to excessive protection in some scenarios, unnecessarily degrading data utility, while in others it may provide insufficient privacy protection. To address this limitation, this work explores a variation that dynamically adjusts the privacy budget according to the predictability of the user's next trajectory points. This adaptive strategy seeks to optimize budget allocation, thereby achieving a more effective balance between privacy and utility.

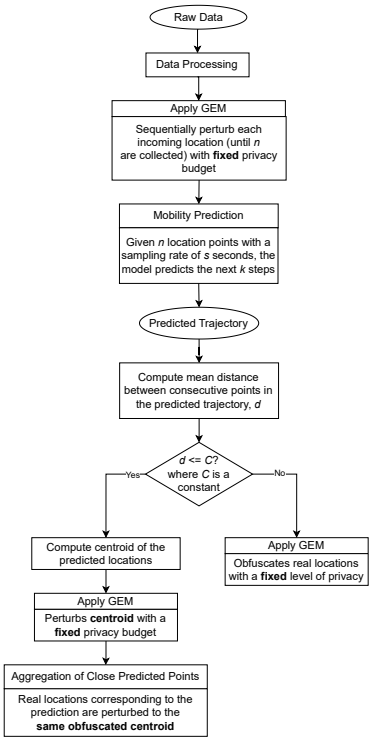

Figure 2. Overview of the Aggregation (GEM-PA) proposal.

GEM-AE has the ability to resist prediction attacks, as it takes into consideration the predictability of the locations before the user even gets there, adding more noise to a location with higher predictability and a lower privacy protection to locations that have lower predictability.

The value of ϵ is computed using an exponential decay function bounded between a minimum (ϵ_{min}) and a maximum (ϵ_{max}) value. Specifically, as the predictability risk p_i increases - meaning that the user's next location can be inferred with higher certainty - the function assigns a value of ϵ closer to ϵ_{min}, enforcing stronger privacy protection. Conversely, when predictability is low, ϵ approaches ϵ_{max}, thereby reducing the amount of injected noise and improving utility:

$$\epsilon = \epsilon_{\text{min}} + (\epsilon_{\text{max}} - \epsilon_{\text{min}}) \cdot e^{-2 \cdot p_i},$$

where p_i is the predictability of location i where $i \in \{1, ..., k\}$ and k denotes the number of predicted locations.

The exponential form of the function ensures a smooth but non-linear transition between the extremes, providing a more effective balance between privacy and utility compared to a fixed privacy budget.

3.2.2. Aggregation of Close Predicted Location Points - GEM-PA. In Graph-Exponential Mechanism with Predictive Aggregation (GEM-PA), the implemented mobility prediction model is leveraged to determine what locations should be aggregated.

To determine whether locations should be aggregated, we introduce a parameter d, defined as the maximum allowable mean distance between predicted points. Given a sequence of geographic coordinates, the haversine formula is applied to compute pairwise great-circle distances between successive points, in meters. These distances are

then averaged, yielding a single measure of the typical spacing between points in the sequence. If the mean distance is smaller than d, the five predicted points are aggregated into a single representative point by computing their centroid. This centroid is then used as the location to be obfuscated. The integration of this variation into the overall mechanism is illustrated in Figure 2.

The purpose of this strategy is to prevent generating multiple obfuscated points when predicted locations are spatially close, thereby reducing redundancy and mitigating the cumulative privacy loss that can arise from multiple correlated releases of obfuscated data, which could otherwise enable an adversary to infer sensitive information more effectively.

The new Aggregation-based Graph-Exponential Mechanism relies on a fixed privacy budget. However, as discussed in Section 2, recent research has emphasized the importance of adapting the privacy budget according to specific criteria.

4. Evaluation Results

This section begins by detailing the experimental setup and the location privacy-preserving metrics used to evaluate the LPPMs. It then presents two evaluation axes: (i) the evaluation of GEM-PA in comparison with other aggregation-based methods; and (ii) the evaluation of GEM-AE in comparison with other adaptive LPPMs.

4.1. Experimental Setup

To evaluate the proposed mechanisms, our work benefited from the location privacy implementations available in Privkit [37]. Privkit is a privacy toolkit that standardizes the privacy analysis by providing Privacy-Preserving Mechanisms (PPMs), attacks, and metrics. Building on this, our methodology consisted in applying the selected LPPMs to location data, followed by the Optimal Strategy for Location Attacks [38]. The output from the LPPMs and the attack was then used to assess the methods in terms of privacy and utility using two metrics: Adversary Error and Quality Loss. These metrics are described in detail in the following section. The models proposed in this work have also been integrated into Privkit, contributing to its growing collection of LPPMs and making them readily available for future research.

To perform the experiments, we selected a real mobility dataset, Geolife [39], which was collected in a period of over three years from GPS devices. As a preliminary step, the dataset was pre-processed following the data processing methodology described in [32], resulting in trajectories with a 30-second sampling rate. For our experiments, only trajectories recorded in Beijing were retained.

Regarding the configuration of the LPPMs, five different privacy budget values were used for testing: $\epsilon \in \{0.008, 0.016, 0.032, 0.064, 0.128\}$. These values were selected based on prior studies [11], [12], thereby facilitating the comparison with related work. Additional parameters specific to the LPPMs themselves are introduced in the corresponding sections.

The Seq2Seq model, used in the mobility prediction step of the proposed methods, was trained with a learning rate of 0.006, learning rate decay of 0.9, batch size of 32, L2 penalty of 0.0001 and 0.1 gradient clipping. These parameters were obtained both by considering the parameters used in the original paper and by performing hyperparameter tuning. The prediction model was set to predict the next five steps of the trajectory.

4.2. Location Privacy Preserving Metrics

The Adversary Error and Quality Loss metrics quantify the fundamental trade-off between privacy and utility.

Adversary Error represents the privacy level by measuring the expected estimation error of an attacker. Assuming the adversary possesses prior knowledge of user distribution, typically modeled as a probability distribution π over the set of possible user locations X, and employs an optimal inference strategy [38], the Adversary Error quantifies the average geographic distance between the user's true location, x, and the attacker's best estimate, x'. Given the released perturbed location, the attacker computes the posterior probability distribution over x:

$$Pr(x|x') = \frac{\pi(x) \cdot Pr(x'|x)}{\sum_{x \in X} Pr(x'|x)}$$

Based on this posterior, the adversary selects an estimate $\hat{x}$ that minimizes the expected inference error, measured as the geographic distance, considering the geographic distance between $\hat{x}$ and x, $||\hat{x} - x||$:

$$\hat{x} = \arg\min_{\hat{x}} \sum_{x \in X} Pr(x|x') \cdot ||\hat{x} - x||$$

The Adversary Error is then defined as the expected distance between the true location and the adversary's estimate:

$$AdvError = E||\hat{x} - x|| = \sum_{\hat{x}} Pr(\hat{x}|x') \cdot ||\hat{x} - x||$$

Quality Loss assesses the utility degradation inherent in the perturbation process. It is defined as the expected geographic distance between the user's actual location x and the perturbed location x' released by the LPPM:

$$Q_{loss} = E||x - x'||$$

4.3. Aggregation of Close Location Data Points - GEM-PA

We evaluate our proposed GEM-PA method against two baselines: Clustering Geo-Indistinguishability (CGI) [12], which clusters nearby location points based on a fixed radius, and Prediction-based Clustering Geo-Indistinguishability (P-CGI), a variation of the CGI that aggregates locations according to the proximity of predicted points along a trajectory. The goal of P-CGI is to isolate the effect of prediction without considering the road network.

To isolate effects on Adversary Error, we set the aggregation radius to $r = 110$ for CGI and the mean distance threshold to $d = 20$ for prediction-based methods. These parameters were chosen to ensure comparable levels of Quality Loss across all methods.

In terms of privacy protection (Adversary Error), Figure 3 shows that GEM-PA consistently outperforms both

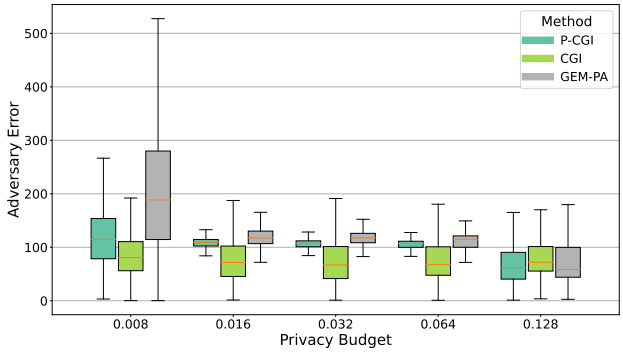

Figure 3. Comparative performance of Adversary Error of Clustering mechanisms.

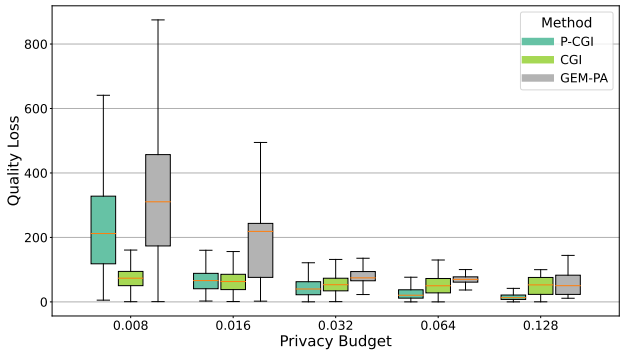

Figure 4. Comparative performance of Quality Loss of Clustering mechanisms.

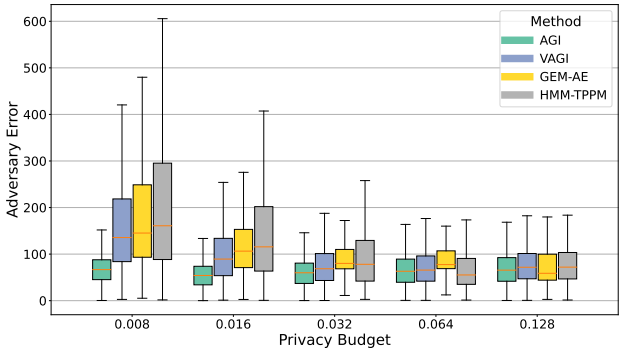

Figure 5. Comparative performance of Adversary Error of Adaptive mechanisms.

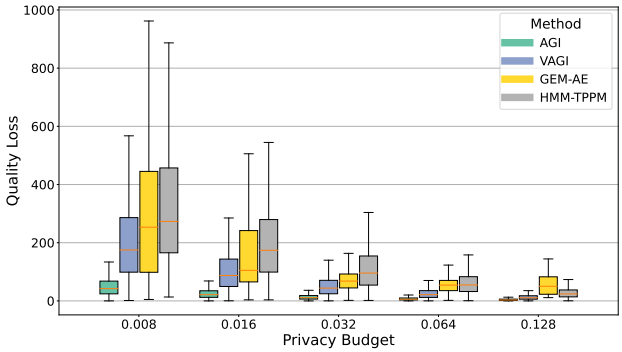

Figure 6. Comparative performance of Quality Loss of Adaptive mechanisms.

P-CGI and CGI across all privacy budgets, and more significantly for lower ϵ values (i.e. more noise). At $\epsilon = 0.008$, GEM-PA achieves a median Adversary Error of approximately 200 m, as can be seen in Figure 3, doubling the protection offered by CGI. This tendency for higher privacy protection is maintained across all tested privacy budgets, but reduced for larger values of ϵ, since less noise is added. This demonstrates that road-network-aware aggregation of predicted points effectively restricts the information available for trajectory inference.

The privacy gains of GEM-PA result in higher Utility Loss when more noise is added (i.e. $\epsilon \leq 0.016$), as shown in Figure 4. This is expected, as replacing trajectory steps with a spatial centroid introduces inherent deviation. However, as the privacy budget ϵ increases GEM-PA provides a better privacy-utility trade-off, with a higher Adversary Error for comparable Quality Loss.

4.4. Adaptation of the Privacy Budget based on Location Predictability - GEM-AE

We compare GEM-AE against three adaptive baselines: Adaptive Geo-Indistinguishability (AGI) [10], Velocity-Aware Geo-Indistinguishability (VA-GI) [11] (where the multiplier m, used to adjust the privacy and utility bounds, was set to $m = 10$), and Hidden Markov Model Trajectory Prediction Privacy Mechanism (HMM-TPPM) [15]. For both GEM-AE, AGI and HMM-TPPM, the privacy budget range was constrained by $\epsilon_{min} = \epsilon/2$ and $\epsilon_{max} = 2\epsilon$. For VA-GI, a base privacy budget ϵ is specified, and the effective budget is modulated by the multiplier m. For AGI, the window size was set to 2 and the remaining parameters were kept at their default values.

AGI adapts the privacy budget based on correlations between consecutive locations, while VA-GI adjusts it according to user speed and reporting frequency to balance privacy and utility. HMM-TPPM leverages trajectory predictability by using a Hidden Markov Model to estimate the next location and tune the privacy budget based on the discrepancy between predicted and actual positions. GEM-AE adapts the privacy budget by incorporating location predictability into an exponential function, which selects an appropriate budget from a predefined range for each location, as detailed in 3.2.

As shown in Figure 5, GEM-AE consistently outperforms AGI and VA-GI in terms of privacy protection, particularly at the lowest budget ($\epsilon = 0.008$) where it achieves a significantly higher median Adversary Error. While HMM-TPPM, the only other method leveraging predictability, occasionally reaches higher peak privacy, it suffers from extreme instability and high variance, as indicated by the long whiskers at lower privacy budgets. In contrast, GEM-AE's use of a Seq2Seq model with MC-dropout provides a more robust and concentrated protection profile.

The privacy gains of GEM-AE involve a structured trade-off. While AGI and VA-GI maintain lower Quality Loss, as seen in Figure 6, this is a direct result of their lower privacy floor. The critical advantage of GEM-AE over HMM-TPPM lies in the utility preservation; HMM-TPPM frequently produces superior Quality Loss, especially for $\epsilon \geq 0.016$. GEM-AE's exponential decay

function ensures a more graceful degradation of utility, maintaining a compact distribution of error even under strict privacy constraints.

For location-based applications, GEM-AE provides a more reliable privacy-utility balance than HMM-TPPM. By dynamically adjusting the budget only when trajectory uncertainty is low, it avoids unnecessary "over-perturbation". Our results indicate that for $0.016 \leq \epsilon \leq 0.064$, GEM-AE clearly emerges as the optimal choice, achieving superior privacy with minimal utility impact through strategic budget allocation.

4.5. Discussion

Experimental evaluation focused on privacy-utility trade-offs using point-level metrics (Adversary error and Quality loss). Overall, prediction-enhanced clustering proved consistently beneficial: GEM-PA (and prediction-based clustering variants) outperformed standard clustering approaches, indicating that prediction is a promising tool to address the privacy issue of repeated disclosure of correlated obfuscated locations. Moreover, approaches based on prediction for adaptation of the privacy budget epsilon (GEM-AE and HMM-TPPM) exhibited consistently better privacy-utility trade-offs than their counterparts, with a utility advantage for our proposed GEM-AE. Additionally, due to its formulation, GEM-AE provided protection against prediction attacks which is not guaranteed in other state-of-the-art methods.

The observations in Section 4.4 are supported by formal statistical testing. Kruskal–Wallis tests confirm significant differences ($p < 0.001$) among all methods for both metrics at every privacy budget. To further investigate pairwise differences, Mann–Whitney U tests with Bonferroni correction were performed. In this context, smaller p-values indicate stronger evidence of a statistically significant difference after correction, while Cliff's delta ($\delta \in [-1, 1]$) quantifies the direction and magnitude of the effect, with values near 0 indicating overlapping performance and values closer to ± 1 indicating stronger dominance of one method over the other.

The comparable performance between GEM-AE and HMM-TPPM on adversary error is statistically confirmed, as shown in Table 1: no significant difference is observed at $\epsilon = 0.008$ ($p = 0.054$), and effect sizes remain negligible across most budgets, indicating largely overlapping distributions. A notable exception occurs at $\epsilon = 0.064$, where a medium effect size ($\delta = 0.373$) suggests a temporary advantage for GEM-AE. While HMM-TPPM matches GEM-AE in privacy, the significance tests also quantify the associated utility cost (Table 2): quality loss differences are negligible to small for $\epsilon \leq 0.064$ ($|\delta| \leq 0.289$), but become large at $\epsilon = 0.128$ ($\delta = 0.516$). This reinforces GEM-AE's advantage in the mid-range privacy budget window ($0.016 \leq \epsilon \leq 0.064$), where both privacy and utility are statistically comparable or favorable.

The observations from Figures 3 and 4 are also supported by statistical analysis. The differences among GEM-PA, P-CGI, and CGI are statistically significant ($p < 0.001$) for both metrics across nearly all privacy budgets (Tables 3 and 4). To quantify the magnitude of these differences beyond statistical significance, we further report in those tables the Cliff's δ as a non-parametric effect size measure.

TABLE 1. ADVERSARY ERROR SIGNIFICANCE COMPARISON - PREDICTION-BASED MECHANISM

ϵ	Pair	p-value	Cliff's δ
0.008	GEM-AE vs HMM-TPPM	0.054	-0.035 (negligible)
0.016	GEM-AE vs HMM-TPPM	0.017	-0.049 (negligible)
0.032	GEM-AE vs HMM-TPPM	< 0.001	0.063 (negligible)
0.064	GEM-AE vs HMM-TPPM	< 0.001	0.373 (medium)
0.128	GEM-AE vs HMM-TPPM	0.002	-0.065 (negligible)

TABLE 2. QUALITY LOSS SIGNIFICANCE COMPARISON - PREDICTION-BASED MECHANISM

ϵ	Pair	p-value	Cliff's δ
0.008	GEM-AE vs HMM-TPPM	< 0.001	-0.098 (negligible)
0.016	GEM-AE vs HMM-TPPM	< 0.001	-0.241 (small)
0.032	GEM-AE vs HMM-TPPM	< 0.001	-0.289 (small)
0.064	GEM-AE vs HMM-TPPM	0.039	-0.044 (negligible)
0.128	GEM-AE vs HMM-TPPM	< 0.001	0.516 (large)

For Adversary Error (Table 3), GEM-PA's advantage over CGI is substantial at restrictive privacy budgets ($\delta = -0.659$ at $\epsilon = 0.008$) and remains large through mid-range values. This confirms that the higher median values observed in Figure 3 reflect a consistent distributional shift. At $\epsilon = 0.128$, however, the effect becomes negligible ($\delta = 0.119$), indicating that the three methods converge in privacy protection as less noise is added. The comparison between P-CGI and CGI further supports the role of prediction: P-CGI tends to achieve higher Adversary Error than CGI, suggesting that trajectory prediction alone improves privacy, even without road-network-aware aggregation.

For Quality Loss (Table 4), the higher data degradation introduced by GEM-PA at strict budgets, noted in Section 4.3, is confirmed with large effect sizes against CGI ($\delta = -0.794$ at $\epsilon = 0.008$) and a small effect against P-CGI ($\delta = -0.266$). Notably, the effect size between GEM-PA and CGI decreases steadily to negligible at $\epsilon = 0.128$ ($\delta = -0.080$), quantitatively supporting the observation that GEM-PA achieves a comparable privacy-utility trade-off at relaxed privacy budgets.

Overall, these results confirm that incorporating prediction consistently improves privacy protection, particularly at strict and intermediate privacy budgets, where effect sizes indicate substantial gains over baseline methods. While this improvement often comes at the cost of increased quality loss, the gap diminishes as the privacy budget increases, leading to comparable performance across methods at higher ϵ values. Notably, GEM-PA achieves the strongest privacy guarantees under restrictive conditions, whereas GEM-AE provides a more balanced trade-off between privacy and utility, especially at intermediate privacy budget values.

5. Limitations and Future Work

The predictability of human mobility has assumed an important role in location privacy research. In this paper,

TABLE 3. ADVERSARY ERROR SIGNIFICANCE COMPARISON - AGGREGATION-BASED MECHANISMS

ϵ	Pair	p-value	Cliff's δ
0.008	CGI vs GEM-PA	< 0.001	-0.659 (large)
0.008	P-CGI vs CGI	< 0.001	0.362 (medium)
0.008	P-CGI vs GEM-PA	< 0.001	-0.411 (medium)
0.016	CGI vs GEM-PA	< 0.001	-0.704 (large)
0.016	P-CGI vs CGI	< 0.001	0.579 (large)
0.016	P-CGI vs GEM-PA	< 0.001	-0.333 (medium)
0.032	CGI vs GEM-PA	< 0.001	-0.753 (large)
0.032	P-CGI vs CGI	< 0.001	0.548 (large)
0.032	P-CGI vs GEM-PA	< 0.001	-0.528 (large)
0.064	CGI vs GEM-PA	< 0.001	-0.709 (large)
0.064	P-CGI vs CGI	< 0.001	0.540 (large)
0.064	P-CGI vs GEM-PA	< 0.001	-0.332 (medium)
0.128	CGI vs GEM-PA	< 0.001	0.119 (negligible)
0.128	P-CGI vs CGI	< 0.001	-0.194 (small)
0.128	P-CGI vs GEM-PA	< 0.001	-0.069 (negligible)

TABLE 4. QUALITY LOSS SIGNIFICANCE COMPARISON - AGGREGATION-BASED MECHANISMS

ϵ	Pair	p-value	Cliff's δ
0.008	CGI vs GEM-PA	< 0.001	-0.794 (large)
0.008	P-CGI vs CGI	< 0.001	0.708 (large)
0.008	P-CGI vs GEM-PA	< 0.001	-0.266 (small)
0.016	CGI vs GEM-PA	< 0.001	-0.662 (large)
0.016	P-CGI vs CGI	< 0.001	0.037 (negligible)
0.016	P-CGI vs GEM-PA	< 0.001	-0.649 (large)
0.032	CGI vs GEM-PA	< 0.001	-0.517 (large)
0.032	P-CGI vs CGI	< 0.001	-0.222 (small)
0.032	P-CGI vs GEM-PA	< 0.001	-0.647 (large)
0.064	CGI vs GEM-PA	< 0.001	-0.416 (medium)
0.064	P-CGI vs CGI	< 0.001	-0.517 (large)
0.064	P-CGI vs GEM-PA	< 0.001	-0.849 (large)
0.128	CGI vs GEM-PA	< 0.001	-0.080 (negligible)
0.128	P-CGI vs CGI	< 0.001	-0.670 (large)
0.128	P-CGI vs GEM-PA	< 0.001	-0.789 (large)

we designed two novel road-network-aware obfuscation approaches that incorporate a mobility prediction model to (i) adapt the privacy budget based on the prediction of future locations, and (ii) aggregate locations according to the proximity of their predicted future positions. Although the mobility prediction model is user-independent, it is location dependent, as it is trained based on the mobility patterns of the place where the model is applied. While our results relied on real-world data from the Beijing area, we leave for future work the analysis with other real-life datasets or synthetic data that reflect realistic mobility patterns to observe the methods behavior in different places.

Future work also involves extending the evaluation from point-level to trajectory-level privacy, including attacks such as map matching to capture end-to-end trace disclosure [12]. Additionally, predictability should be explored beyond sequence models alone, including richer predictability measures (e.g., time-of-day effects) and semantic/context-aware adaptations that increase protection near sensitive destinations (e.g., hospitals or places of worship). A natural departure for this work is the devising privacy-preserving path alternatives that guarantee successful location-based services, while better preserving user privacy.

6. Conclusion

The large-scale collection of mobility data by modern services has made location privacy a central concern, as raw traces can be repurposed for profiling, monitoring, or leakage beyond users' expectations. This work investigates how mobility prediction and road-network awareness can be jointly leveraged to improve the privacy-utility trade-off of Location Privacy-Preserving Mechanisms (LPPMs).

We propose two novel approaches that take the road-network into consideration through the Graph-Exponential Mechanism (GEM) and complement it with prediction of future locations for adaptive epsilon/privacy budget (GEM-AE) as well as prediction for aggregation of locations (GEM-PA). GEM with Adaptive Epsilon (GEM-AE) adapts the privacy budget as a function of per-point predictability, allocating stronger protection when mobility is easier to infer. GEM with Predictive Aggregation (GEM-PA) reduces redundant disclosures by aggregating predicted future locations into a single representative point when they are sufficiently close, limiting the privacy loss that arises from frequent reporting. To support this design, we adopted a multi-user, multi-step Seq2Seq predictor [32], which generalizes across users and provides a flexible horizon for integrating prediction into obfuscation decisions.

From the performed evaluation, we conclude that incorporating predictability into both budget allocation and clustering improves the behavior of LPPMs under realistic mobility correlations. In particular, GEM-AE achieves a more favorable privacy-utility trade-off than existing adaptive mechanisms, matching the privacy guarantees of HMM-TPPM while consistently incurring lower or comparable quality loss across most privacy budgets. Similarly, GEM-PA provides substantial gains in privacy at low and intermediate ϵ values, as evidenced by large effect sizes in adversary error compared to baseline methods, while maintaining competitive utility as the privacy budget increases. These results demonstrate that leveraging trajectory predictability enables more efficient allocation of privacy resources and reduces redundant disclosures, leading to more robust protection against inference attacks.

Acknowledgment

This work was supported by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, under the support UID/50014/2025 (https://doi.org/10.54499/UID/50014/2025) - INESC TEC, and within the scope of the research unit UID/00326 - Centre for Informatics and Systems of the University of Coimbra (https://doi.org/10.54499/UID/00326/2025).

References

[1] V. Primault, A. Boutet, S. B. Mokhtar, and L. Brunie, "The long road to computational location privacy: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2772–2793, 2019.

[2] L. Franceschi-Bicchierai, "Redditor cracks anonymous data trove to pinpoint Muslim cab drivers," *Mashable*, 2015. [Online]. Available: https://mashable.com/archive/redditor-muslim-cab-drivers

[3] Y. Gu, Y. Yao, W. Liu, and J. Song, "We know where you are: Home location identification in location-based social networks," in *Proc. Int. Conf. Computer Communication and Networks (ICCCN)*, 2016, pp. 1–9.

[4] T.-R. Hu, J.-B. Luo, H. Kautz, and A. Sadilek, "Home location inference from sparse and noisy data: Models and applications," *Frontiers of Information Technology & Electronic Engineering*, vol. 17, no. 5, pp. 389–402, 2016.

[5] Oregon Department of Justice, "Google: AG Rosenblum announces largest AG consumer privacy settlement in U.S. history," 2022. [Online]. Available: https://www.doj.state.or.us/media-home/news-media-releases/largest-ag-consumer-privacy-settlement-in-u-s-history/. Accessed: Sep. 13, 2025.

[6] N. Sadrolodabaee, "University of Melbourne breached privacy by tracking protesters' Wi-Fi location," *SBS News*, 2025. [Online]. Available: https://www.sbs.com.au/news/article/melbourne-university-privacy-palestinian-protests-wi-fi-data/xeeuuxn64. Accessed: Sep. 13, 2025.

[7] R. Mendes, A. Brandão, J.P. Vilela, and A. Beresford, "Effect of User Expectation on Mobile App Privacy: A Field Study," in *IEEE International Conference on Pervasive Computing and Communications (PerCom), 2022, pp. 207-214.*

[8] H. Jiang, J. Li, P. Zhao, F. Zeng, Z. Xiao, and A. Iyengar, "Location privacy-preserving mechanisms in location-based services: A comprehensive survey," *ACM Computing Surveys*, vol. 54, no. 1, art. no. 4, pp. 1–36, Jan. 2022.

[9] M. E. Andrés *et al.*, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proc. ACM SIGSAC Conf. Computer and Communications Security (CCS)*, Berlin, Germany, 2013, pp. 901–914.

[10] R. Al-Dhubhani and J. M. Cazalas, "An adaptive geo-indistinguishability mechanism for continuous LBS queries," *Wireless Networks*, vol. 24, no. 8, pp. 3221–3239, 2018.

[11] R. Mendes, M. Cunha, and J. P. Vilela, "Velocity-aware geo-indistinguishability," in *Proc. ACM Conf. Data and Application Security and Privacy (CODASPY)*, 2023, pp. 141–152.

[12] M. Cunha, R. Mendes, and J. P. Vilela, "Clustering geo-indistinguishability for privacy of continuous location traces," in *Proc. Int. Conf. Computing, Communications and Security (ICCCS)*, 2019, pp. 1–8.

[13] S. Takagi, Y. Cao, Y. Asano, and M. Yoshikawa, "Geo-graph-indistinguishability: Location privacy on road networks with differential privacy," *IEICE Transactions on Information and Systems*, vol. 106, no. 5, pp. 877–894, 2023.

[14] Y. Wang, Y. Xia, J. Hou, S.-M. Gao, X. Nie, and Q. Wang, "A fast privacy-preserving framework for continuous location-based queries in road networks," *Journal of Network and Computer Applications*, vol. 53, pp. 57–73, 2015.

[15] S. Qiu, D. Pi, Y. Wang, and Y. Liu, "Novel trajectory privacy protection method against prediction attacks," *Expert Systems with Applications*, vol. 213, p. 118870, 2023.

[16] E. M. R. Oliveira, A. C. Viana, C. Sarraute, J. Brea, and I. Alvarez-Hamelin, "On the regularity of human mobility," *Pervasive and Mobile Computing*, vol. 33, pp. 73–90, 2016.

[17] R. Mendes, M. Cunha, and J. Vilela, "Impact of frequency of location reports on the privacy level of geo-indistinguishability," *Proc. Privacy Enhancing Technologies*, vol. 2020, no. 2, pp. 379–396, 2020.

[18] B. Liu, W. Zhou, T. Zhu, L. Gao, and Y. Xiang, "Location privacy and its applications: A systematic study," *IEEE Access*, vol. 6, pp. 17606–17624, 2018.

[19] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific Reports*, vol. 3, art. no. 1376, 2013.

[20] A. J. Blumberg and P. Eckersley, "On locational privacy, and how to avoid losing it forever," *Electronic Frontier Foundation*, vol. 10, no. 11, pp. 1–7, 2009.

[21] C. Dwork, "Differential privacy," in *Automata, Languages and Programming*, Berlin, Heidelberg: Springer, 2006, pp. 1–12.

[22] M. E. Gursoy, L. Liu, S. Truex, L. Yu, and W. Wei, "Utility-aware synthesis of differentially private and attack-resilient location traces," in *Proc. ACM SIGSAC Conf. Computer and Communications Security (CCS)*, 2018, pp. 196–211.

[23] V. Shejwalkar *et al.*, "Revisiting utility metrics for location privacy-preserving mechanisms," in *Proc. Annual Computer Security Applications Conf. (ACSAC)*, San Juan, Puerto Rico, 2019, pp. 313–327.

[24] A. Graser, A. Jalali, J. Lampert, A. Weißenfeld, and K. Janowicz, "MobilityDL: A review of deep learning from trajectory data," *GeoInformatica*, pp. 1–33, 2024.

[25] R. Wu, G. Luo, J. Shao, L. Tian, and C. Peng, "Location prediction on trajectory data: A review," *Big Data Mining and Analytics*, vol. 1, no. 2, pp. 108–127, 2018.

[26] M. Chen, X. Yu, and Y. Liu, "MPE: A mobility pattern embedding model for predicting next locations," *World Wide Web*, vol. 22, pp. 1–19, 2019.

[27] M. Li, A. Ahmed, and A. J. Smola, "Inferring movement trajectories from GPS snippets," in *Proc. ACM Int. Conf. Web Search and Data Mining (WSDM)*, 2015, pp. 325–334.

[28] L. Chen, M. Lv, and G. Chen, "A system for destination and future route prediction based on trajectory mining," *Pervasive and Mobile Computing*, vol. 6, no. 6, pp. 657–676, 2010.

[29] M. Chen, Y. Zuo, X. Jia, Y. Liu, X. Yu, and K. Zheng, "CEM: A convolutional embedding model for predicting next locations," *IEEE Trans. Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3349–3358, Jun. 2021.

[30] C. Yong, N. Xie, H. Xu, X. Chen, and D.-H. Lee, "A multi-context aware human mobility prediction model based on motif-preserving travel preference learning," *IEEE Trans. Intelligent Transportation Systems*, early access, 2023.

[31] H. Yao *et al.*, "Modeling spatial-temporal dynamics for traffic prediction," *arXiv preprint arXiv:1803.01254*, 2018.

[32] C. Wang, L. Ma, R. Li, T. Durrani, and H. Zhang, "Exploring trajectory prediction through machine learning methods," *IEEE Access*, vol. 7, pp. 117000–117012, 2019.

[33] B. Yan, G. Zhao, L. Song, Y. Yu, and J. Dong, "PreCLN: Pretrained-based contrastive learning network for vehicle trajectory prediction," *World Wide Web*, vol. 26, no. 4, pp. 1853–1875, Nov. 2023.

[34] D. Yao, C. Zhang, J. Huang, and J. Bi, "SERM: A recurrent model for next location prediction in semantic trajectories," in *Proc. ACM Conf. Information and Knowledge Management (CIKM)*, Singapore, 2017, pp. 2411–2414.

[35] K. Chatzikokolakis, C. Palamidessi, and M. Stronati, "Constructing elastic distinguishability metrics for location privacy," *arXiv preprint arXiv:1503.00756*, 2015.

[36] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Machine Learning (ICML)*, New York, NY, USA, 2016, pp. 1050–1059.

[37] M. Cunha *et al.*, "Privkit: A toolkit of privacy-preserving mechanisms for heterogeneous data types," in *Proc. ACM Conf. Data and Application Security and Privacy (CODASPY)*, Porto, Portugal, 2024, pp. 319–324.

[38] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: Optimal strategy against localization attacks," in *Proc. ACM Conf. Computer and Communications Security (CCS)*, 2012, pp. 617–627.

[39] Y. Zheng, X. Xie, W.-Y. Ma, *et al.*, "GeoLife: A collaborative social networking service among user, location and trajectory," *IEEE Data Engineering Bulletin*, vol. 33, no. 2, pp. 32–39, 2010.