Research Article

AND DATA SCIENCES
BULLETIN OF COMPUTER

Formally Private and Learning-Resistant Location and Query Obfuscation for kNN-based Location-Based Services

Yajuan Wang and Kai Shang

School of Management, Suzhou University, Suzhou, Anhui, 234000, China

Abstract

Location-based services (LBS) routinely answer k-nearest neighbor (kNN) queries over users' locations and points of interest, but revealing precise locations and query patterns poses serious privacy risks. Existing systems rely largely on heuristic dummy generation and query fragmentation, and typically argue privacy via entropy or attack-specific reasoning under restricted adversary models. In this paper, we present a new framework for location and query obfuscation in kNN-based LBS that provides *formal* privacy guarantees and is explicitly evaluated against modern machine-learning-based adversaries. We introduce Geo-Obfus, a spatially differentially private dummy generation mechanism that satisfies geo-indistinguishability, and Query-Obfus, a distributionally private query obfuscation scheme that protects sensitive query attributes, both integrated into an efficient two-stage kNN processing pipeline. We derive theoretical guarantees for location privacy, query privacy, and their composition over repeated queries, analyze the utility loss in terms of kNN accuracy and latency, and evaluate robustness against optimal Bayesian inference and neural classifiers trained to distinguish real locations and queries from dummies. Using real and synthetic mobility datasets, we show how privacy parameters control the trade-off between formal privacy and kNN utility and demonstrate that Geo-Obfus and Query-Obfus substantially reduce the success rate of learning-based attacks compared with heuristic dummy and fragmentation methods, providing a principled, learning-resistant foundation for privacy-preserving kNN services.

Keywords: location privacy, differentially private kNN, geo-indistinguishability, query obfuscation, privacy-preserving location-based services

1. Introduction

Location-based services (LBS) such as navigation, restaurant search, and ride-hailing routinely answer k-nearest neighbor (kNN) queries based on a user's current or predicted location [1, 2]. While highly convenient, such services pose significant privacy risks: precise locations can reveal home and work addresses, health and religious habits, and sensitive social relationships [3, 4]. Moreover, sequences of queries may allow adversaries to reconstruct fine-grained trajectories or infer semantic attributes of users [4–6].

A long line of research has sought to protect users' privacy in LBS through spatial cloaking, k-anonymity, mix zones, dummy locations, private information retrieval, and cryptographic protocols [1, 3, 5, 7, 8]. Many practical systems focus on generating dummy locations and fragmenting

Corresponding author: Yajuan Wang (yajuanwang12@proton.me)

queries so that an adversary cannot easily distinguish the true location or infer sensitive information from observed requests [9, 10]. Recent agent-based frameworks have combined dummy selection, query fragmentation, and kNN indexing into integrated systems with promising performance [11, 12]. However, these approaches are largely heuristic: privacy is argued in terms of Shannon entropy or resistance to specific attacks, rather than formal, tunable guarantees. At the same time, modern adversaries can leverage powerful machine learning models to exploit subtle statistical differences between real locations and dummies [4, 12].

In this work, we argue that the next generation of privacy-preserving kNN services must satisfy two key requirements. First, privacy mechanisms should come with *formal, parameterized guarantees*, such as (spatial) differential privacy or geo-indistinguishability, allowing operators and users to reason about worst-case leakage [13, 14]. Second, these mechanisms should be designed and evaluated under *learning-based adversaries* who can train classifiers or generative models on large volumes of queries [4, 12]. We address these requirements by proposing a new framework for formally private, learning-resistant location and query obfuscation in kNN-based LBS.

In this paper we develop a unified framework that combines formal privacy guarantees with robustness to learning-based attacks in kNN-based LBS. We begin by defining a comprehensive threat model that captures both an honest-but-curious service provider and external adversaries capable of training machine learning models on observed obfuscated queries and auxiliary mobility data [4, 12]. Within this setting, we introduce Geo-Obfus, a spatially differentially private mechanism for generating dummy locations around a sanitized center that satisfies geo-indistinguishability with parameter ε_ℓ while explicitly controlling the number and spatial distribution of dummies to preserve kNN utility [13].

Complementing this, we propose Query-Obfus, a distributionally private query obfuscation mechanism that operates on structured query attributes such as categories and keywords and achieves an ε_q-differential privacy guarantee with respect to a carefully defined neighboring relation on queries [14]. We integrate Geo-Obfus and Query-Obfus into a two-stage kNN processing pipeline in which the server performs coarse candidate retrieval based on obfuscated locations and attributes, and the client subsequently refines these candidates locally to recover high-quality answers without exposing the true location or full query.

We provide a rigorous analysis of the resulting system, deriving formal guarantees for location privacy, query privacy, and their composition over multiple queries and over time [15]. Finally, we design an evaluation methodology that instantiates both Bayesian and learning-based adversaries, including neural classifiers trained to identify the real location and query among dummies, and we use real and synthetic mobility datasets to demonstrate that the proposed mechanisms substantially reduce adversarial success while maintaining competitive kNN performance. Overall, our framework advances the state of the art by replacing heuristic privacy arguments with formally parameterized guarantees and by explicitly targeting modern learning-based adversaries.

2. Background, Related Work, and Problem Formulation

In a kNN LBS, users issue queries of the form "find the k nearest points of interest (POIs) of type c to my current location." The service provider maintains an index over POIs and possibly moving objects (e.g., vehicles), and answers queries using spatial indexing and distance computations. Efficient kNN processing has been studied extensively, including grid-based indexes, tree-based structures such as

R-trees, and road-network-aware methods [16–19]. From a privacy perspective, raw locations and query attributes can be highly sensitive: even if identifiers are removed, repeated queries may allow reconstruction of trajectories and inference of home, work, and other sensitive locations [6, 12, 20]. Recent surveys underscore that location traces are highly identifying even at coarse resolution and across diverse application domains [21, 22].

Classical approaches to location privacy include spatial cloaking and k-anonymity, where user locations are generalized into regions containing at least k users, and mix zones, where trajectories are temporarily hidden [3, 5, 7]. Dummy-based methods generate fake locations alongside the real one, aiming to confound the adversary's inference; early work focused on simple geometric heuristics, while more recent schemes incorporate road network constraints and entropy-based quality measures to produce realistic dummies [23]. Cryptographic approaches, such as private information retrieval (PIR) and secure multi-party computation, can offer strong privacy guarantees but often incur substantial computational and communication overhead [23]. For example, PIR-based LBS prototypes demonstrate excellent privacy but struggle to meet strict latency requirements in large-scale deployments [14].

More recently, differential privacy and its spatial variant, geo-indistinguishability, have emerged as principled frameworks for location privacy. Geo-indistinguishability adapts the differential privacy definition to metrics over locations, providing a tunable privacy parameter ε_ℓ that bounds the relative likelihood of outputs for nearby locations [13]. Mechanisms such as the planar Laplace mechanism achieve geo-indistinguishability by adding appropriately scaled noise to coordinates, and subsequent work has explored the design of *optimal* geo-indistinguishable mechanisms that minimize utility loss subject to a given privacy budget [23]. These developments connect location privacy directly to the broader theory of differential privacy and enable rigorous analysis of composition and worst-case leakage [14, 15].

Beyond the physical location, query content—including requested categories, keywords, and time patterns—can reveal sensitive user attributes. Protecting such attributes requires query obfuscation or randomization, which can be modeled via differential privacy over a neighboring relation on queries (for example, queries differing in a single sensitive attribute) [21, 22]. Modern adversaries may train machine learning models to attack obfuscated data: for location privacy, ML-based attacks include classifying the true location among dummies, reconstructing trajectories, or inferring home and work locations, while for queries, models can predict sensitive attributes from obfuscated requests [4, 12, 13]. A realistic privacy framework must therefore anticipate such learning-based adversaries and evaluate the robustness of proposed mechanisms not only against hand-crafted attacks, but also against data-driven inference.

Agent-based frameworks for privacy-preserving kNN often deploy mobile agents to manage queries on devices and stationary agents to coordinate at the server side. Typical components include dummy selection algorithms that use historical query probabilities and distances, fragmentation algorithms that permute and encrypt query units to mask correlations, and cell-based indexing structures for efficient kNN processing [11, 24]. While such systems demonstrate promising performance and intuitive privacy properties, they largely rely on heuristics and informal arguments, and they do not explicitly consider learning-based adversaries. Our work builds on the architectural insight of combining location obfuscation, query obfuscation, and efficient kNN processing, but replaces heuristic mechanisms with formally private ones and extends the adversary model to include modern machine learning attacks.

We consider an LBS provider that maintains a set of POIs $\mathcal{P}$ and possibly additional moving objects. Users hold mobile devices capable of local computations and communicate with the LBS over a network. A user at true location $x \in \mathbb{R}^2$ issues a kNN query q that includes at least a spatial component (current location) and possibly a set of attributes a such as category, keywords, or preferences. The provider is assumed to be honest-but-curious: it follows the protocol to answer queries correctly based on received information, but it also attempts to infer users' true locations and sensitive query attributes, consistent with standard assumptions in the location privacy literature [4, 25]. In addition, we consider external adversaries who may observe traffic at the LBS, possess auxiliary mobility data, and train machine learning models to exploit statistical patterns in obfuscated queries.

The adversary observes a sequence of obfuscated queries $\{(z_t, \tilde{q}_t)\}_{t=1}^{T}$, where z_t are obfuscated locations and $\tilde{q}_t$ are obfuscated query attributes produced by our mechanisms. We assume that the adversary knows the obfuscation algorithms and their parameters, the set of possible locations (e.g., a map or grid), and prior distributions over user locations derived from auxiliary data. Its goals include identifying the true location x_t among the candidate locations (true plus dummies) for each query, inferring sensitive attributes of the query (for example, whether the category indicates a hospital or religious site), and reconstructing trajectories or long-term profiles from sequences of obfuscated requests [6, 20]. We consider both optimal Bayesian adversaries, who perform maximum a posteriori (MAP) inference given the assumed priors, and learning-based adversaries, who train classifiers or sequence models to predict the true location or sensitive attributes from observed obfuscated data [4, 25].

Within this setting, we formulate two main privacy objectives. For location privacy, we require that for locations $x, x' \in \mathbb{R}^2$ and obfuscated location output z,

$$\frac{\Pr[M_\ell(x) = z]}{\Pr[M_\ell(x') = z]} \leq \exp(\varepsilon_\ell \, d(x, x')),$$

where M_ℓ is the location obfuscation mechanism, $d(\cdot, \cdot)$ is a metric over locations, and ε_ℓ is a privacy parameter; this is the definition of geo-indistinguishability adapted to our setting [13, 25]. For query privacy, we require that for queries q, q' that differ in a single sensitive attribute (under a neighboring relation $\mathcal{N}$) and any set of outputs S,

$$\Pr[M_q(q) \in S] \leq e^{\varepsilon_q} \Pr[M_q(q') \in S],$$

where M_q is the query obfuscation mechanism and ε_q is a query privacy parameter. We further analyze the composition of M_ℓ and M_q over multiple queries and quantify the adversary's success probability in terms of these parameters, leveraging standard composition theorems from differential privacy [14, 15].

3. Obfuscation Mechanisms

3.1. Geo-Obfus: Spatially Differentially Private Dummy Generation

We propose GEO-OBFUS, a two-stage mechanism that first generates a sanitized center location using the planar Laplace mechanism, then selects additional dummy locations around this center to satisfy kNN requirements while preserving geo-indistinguishability.

Given true location $x \in \mathbb{R}^2$, Geo-Obfus proceeds as follows. First, in the *sanitized center generation* step, we sample z_0 from the planar Laplace distribution centered at x with parameter ε_ℓ:

$$f_{Z_0}(z \mid x) = \frac{\varepsilon_\ell^2}{2\pi} \exp(-\varepsilon_\ell d(x, z)),$$

where d is the Euclidean distance. Second, in the *dummy location sampling* step, we condition on z_0 and sample $k-1$ dummy locations $\{z_1, \ldots, z_{k-1}\}$ from a distribution $g(\cdot \mid z_0)$ designed to mimic the prior spatial distribution of user queries or POIs, subject to constraints on minimum separation and maximum radius. The dummy distribution is independent of x given z_0. Finally, in the *permutation and reporting* step, we randomly permute the multiset $\{z_0, z_1, \ldots, z_{k-1}\}$ and report it to the LBS as the set of candidate locations for kNN processing. The true location x never leaves the device.

Intuitively, z_0 provides the formal geo-indistinguishability guarantee, while the additional dummies increase combinatorial ambiguity for the adversary and help maintain kNN utility by ensuring a reasonable spatial spread.

Theorem 1. *Geo-Obfus satisfies ε_ℓ-geo-indistinguishability for the reported multiset of k locations under the metric d, assuming the dummy distribution $g(\cdot \mid z_0)$ is independent of x given z_0.*

Proof. We first recall the definition of geo-indistinguishability for a mechanism whose output space is some measurable space $(\mathcal{Y}, \mathcal{F})$. A randomized mechanism M that takes as input a location $x \in \mathbb{R}^2$ and outputs $Y \in \mathcal{Y}$ is said to satisfy ε_ℓ-geo-indistinguishability with respect to the metric d if, for all $x, x' \in \mathbb{R}^2$ and all measurable sets $B \in \mathcal{F}$,

$$\Pr[M(x) \in B] \;\le\; e^{\varepsilon_\ell d(x,x')} \Pr[M(x') \in B].$$

Let Z_0 denote the sanitized center location generated by the planar Laplace mechanism, and let M_0 be the mechanism that maps x to Z_0. By assumption (the standard property of the planar Laplace mechanism), M_0 satisfies ε_ℓ-geo-indistinguishability, i.e., for all x, x' and all measurable $A \subseteq \mathbb{R}^2$,

$$\Pr[Z_0 \in A \mid X = x] \;\le\; e^{\varepsilon_\ell d(x,x')} \Pr[Z_0 \in A \mid X = x']. \tag{1}$$

The Geo-Obfus mechanism M takes as input x and outputs a multiset Y of k locations constructed as follows. First, it samples $Z_0 \sim M_0(x)$. Then, conditionally on $Z_0 = z$, it samples dummy locations $Z_1, \ldots, Z_{k-1}$ from the distribution $g(\cdot \mid z)$, which by assumption does not depend on x once z is fixed. Finally, it randomly permutes the multiset $\{Z_0, Z_1, \ldots, Z_{k-1}\}$ and outputs the resulting multiset Y.

It is convenient to view M as a two-stage mechanism. In the first stage, x is mapped to Z_0 by M_0. In the second stage, a randomized mapping K takes Z_0 as input and produces the multiset Y by sampling dummies according to $g(\cdot \mid Z_0)$ and applying the random permutation. Crucially, K depends only on Z_0 and on its own internal randomness, and does not depend on x directly. In other words, M can be written as the composition

$$M = K \circ M_0.$$

We now show that M inherits ε_ℓ-geo-indistinguishability from M_0. Let $\mathcal{Y}$ denote the space of multisets of k locations in $\mathbb{R}^2$, endowed with its natural σ-algebra. Fix arbitrary locations $x, x' \in \mathbb{R}^2$ and an arbitrary measurable set $B \subseteq \mathcal{Y}$. We need to prove

$$\Pr[M(x) \in B] \;\le\; e^{\varepsilon_\ell d(x,x')} \Pr[M(x') \in B].$$

For any $z \in \mathbb{R}^2$, let

$$h_B(z) \;:=\; \Pr[K(Z_0) \in B \,|\, Z_0 = z].$$

By definition, $h_B(z)$ is the probability that the second-stage mechanism K outputs a multiset in B when its input is z. Because K is independent of x given Z_0, the function h_B does not depend on x or x', only on B and on the fixed description of K. Moreover, by construction, $0 \le h_B(z) \le 1$ for all z.

Let μ_x and $\mu_{x'}$ denote the probability measures on $\mathbb{R}^2$ induced by Z_0 under inputs x and x', respectively, so that for any measurable $A \subseteq \mathbb{R}^2$,

$$\mu_x(A) = \Pr[Z_0 \in A \mid X = x], \qquad \mu_{x'}(A) = \Pr[Z_0 \in A \mid X = x'].$$

By (1), these measures satisfy

$$\mu_x(A) \;\le\; e^{\varepsilon_\ell d(x,x')}\mu_{x'}(A) \quad \text{for all measurable } A \subseteq \mathbb{R}^2. \tag{2}$$

Using the law of total probability and the definition of h_B, we can write

$$\Pr[M(x) \in B] = \Pr[K(Z_0) \in B \mid X = x] = \int_{\mathbb{R}^2} h_B(z)\, d\mu_x(z),$$

and similarly

$$\Pr[M(x') \in B] = \Pr[K(Z_0) \in B \mid X = x'] = \int_{\mathbb{R}^2} h_B(z)\, d\mu_{x'}(z).$$

Thus the desired inequality becomes

$$\int_{\mathbb{R}^2} h_B(z)\, d\mu_x(z) \;\le\; e^{\varepsilon_\ell d(x,x')} \int_{\mathbb{R}^2} h_B(z)\, d\mu_{x'}(z).$$

To obtain this inequality from (2), we use a standard argument for nonnegative measurable functions. Since $0 \le h_B(z) \le 1$, h_B can be approximated from below by simple functions of the form

$$s(z) = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}(z),$$

where $0 \le \alpha_i \le 1$ and A_i are measurable subsets of $\mathbb{R}^2$. For such a simple function we have

$$\int_{\mathbb{R}^2} s(z)\, d\mu_x(z) = \sum_{i=1}^{n} \alpha_i \mu_x(A_i) \;\le\; \sum_{i=1}^{n} \alpha_i e^{\varepsilon_\ell d(x,x')} \mu_{x'}(A_i) = e^{\varepsilon_\ell d(x,x')} \int_{\mathbb{R}^2} s(z)\, d\mu_{x'}(z),$$

where we used (2) for each A_i. Taking an increasing sequence of simple functions s_n converging pointwise to h_B and applying the monotone convergence theorem on both sides, we obtain

$$\int_{\mathbb{R}^2} h_B(z)\, d\mu_x(z) \;\le\; e^{\varepsilon_\ell d(x,x')} \int_{\mathbb{R}^2} h_B(z)\, d\mu_{x'}(z),$$

which is exactly

$$\Pr[M(x) \in B] \;\le\; e^{\varepsilon_\ell d(x,x')} \Pr[M(x') \in B].$$

Since $B \subseteq \mathcal{Y}$ was arbitrary, this establishes that M satisfies ε_ℓ-geo-indistinguishability under the metric d for the multiset output of k locations. The crucial ingredients were that M_0 is ε_ℓ-geo-indistinguishable and that the second-stage mechanism K is a randomized post-processing of Z_0 independent of x. This is precisely the post-processing property of differential privacy specialized to the geo-indistinguishability setting. The theorem follows. □

The theorem shows that our design inherits formal location privacy guarantees from the planar Laplace mechanism. Adding dummies does not degrade the guarantee and may improve practical privacy by increasing confusion for the adversary. However, geo-indistinguishability with small ε_ℓ may move z_0 far from x, potentially harming kNN accuracy. In practice, the user ultimately evaluates the received kNN results locally and can discard results that are inconsistent with x. We therefore tune the radius of the dummy sampling distribution g and the kNN retrieval radius at the server to ensure that the true nearest neighbors are included with high probability, even when z_0 is noisy.

3.2. Query-Obfus: Distributionally Private Query Obfuscation

We represent each query q as a tuple $q = (c, w, o)$, where c is a category (e.g., restaurant, hospital), w is a bag of keywords, and o includes optional attributes such as rating preferences. We define a neighboring relation $\mathcal{N}$ on queries such that $q \sim q'$ if they differ in at most one sensitive attribute. For example, we may consider the category c to be sensitive if it reveals health or religious information, while keywords related to generic preferences are non-sensitive.

Our QUERY-OBFUS mechanism combines randomized response for categorical attributes with controlled noise for numeric or vector representations. For a sensitive categorical attribute c taking values in a finite domain $\mathcal{C}$, we use generalized randomized response with parameter ε_q:

$$\Pr[\tilde{c} = c] = \frac{e^{\varepsilon_q}}{e^{\varepsilon_q} + |\mathcal{C}|-1}, \qquad \Pr[\tilde{c} = c'] = \frac{1}{e^{\varepsilon_q} + |\mathcal{C}|-1},\ c' \neq c.$$

Non-sensitive attributes may be left unchanged or lightly obfuscated for additional protection. For vector representations $v(q)$ (for example, embeddings of keywords), we add Laplace or Gaussian noise calibrated to ε_q and the chosen sensitivity notion. The obfuscated query is $\tilde{q} = (\tilde{c}, \tilde{w}, \tilde{o})$, which the LBS uses for coarse filtering of POIs.

Theorem 2. *QUERY-OBFUS satisfies ε_q-differential privacy with respect to the neighboring relation $\mathcal{N}$, under standard sensitivity assumptions on the vector transformation $v(q)$.*

Proof. We first recall the notion of differential privacy in our setting. Let $\mathcal{Q}$ be the space of queries and $\mathcal{Y}$ the output space of the mechanism. A randomized mechanism $M_q : \mathcal{Q} \to \mathcal{Y}$ is said to satisfy ε_q-differential privacy with respect to a neighboring relation $\mathcal{N}$ on $\mathcal{Q}$ if, for all neighboring queries $q, q' \in \mathcal{Q}$ with $q \sim q'$ and for all measurable sets $S \subseteq \mathcal{Y}$,

$$\Pr[M_q(q) \in S] \ \leq\ e^{\varepsilon_q}\ \Pr[M_q(q') \in S].$$

In our construction, each query is represented as a tuple $q = (c, w, o)$ where c is a (possibly sensitive) category, w is a bag of keywords, and o denotes optional attributes. The neighboring relation $\mathcal{N}$ is defined so that $q \sim q'$ if they differ in at most one sensitive attribute (for example, in the category c while agreeing on all non-sensitive components). We assume that the vector representation $v(q)$ has bounded sensitivity with respect to $\mathcal{N}$, i.e., there exists $\Delta_v > 0$ such that

$$\|v(q) - v(q')\|_1 \leq\ \Delta_v \quad \text{for all } q \sim q'.$$

The QUERY-OBFUS mechanism acts on two components:

- On the categorical attribute c, it applies generalized randomized response with parameter $\varepsilon_c > 0$.

- On the vector representation $v(q)$, it adds Laplace (or Gaussian) noise calibrated to $\varepsilon_v > 0$ and the sensitivity Δ_v.

The full obfuscated output $\tilde{q}$ is then obtained by combining these perturbed components with the remaining attributes (which are either unchanged or post-processed).

To show that QUERY-OBFUS is ε_q-differentially private, we proceed in three steps: (i) we show that the randomized response mechanism on c is ε_c-DP, (ii) we show that the noise addition on $v(q)$ is ε_v-DP, and (iii) we invoke the composition and post-processing properties to deduce the overall guarantee. We choose ε_c and ε_v such that

$$\varepsilon_c + \varepsilon_v \;\leq\; \varepsilon_q,$$

and, for concreteness, one may take $\varepsilon_c = \varepsilon_v = \varepsilon_q/2$.

Step 1: Randomized response on the category. Let $\mathcal{C}$ be the finite domain of categories. The mechanism M_c takes as input $c \in \mathcal{C}$ and outputs $\tilde{c} \in \mathcal{C}$ according to

$$\Pr[\tilde{c} = c] = \frac{e^{\varepsilon_c}}{e^{\varepsilon_c} + |\mathcal{C}|-1}, \qquad \Pr[\tilde{c} = c'] = \frac{1}{e^{\varepsilon_c} + |\mathcal{C}|-1}, \quad c' \neq c.$$

Consider two neighboring queries $q = (c, w, o)$ and $q' = (c', w', o')$ such that $q \sim q'$ and the only sensitive difference is in the category (so $c \neq c'$). For any output category $\hat{c} \in \mathcal{C}$, we can compute the ratio of probabilities. There are two cases:

Case 1: $\hat{c} = c$. Then

$$\frac{\Pr[M_c(q) = \hat{c}]}{\Pr[M_c(q') = \hat{c}]} = \frac{\Pr[\tilde{c} = c \mid c]}{\Pr[\tilde{c} = c \mid c']} = \frac{\dfrac{e^{\varepsilon_c}}{e^{\varepsilon_c} + |\mathcal{C}|-1}}{\dfrac{1}{e^{\varepsilon_c} + |\mathcal{C}|-1}} = e^{\varepsilon_c}.$$

Case 2: $\hat{c} \notin \{c, c'\}$. Then q and q' assign the same probability to $\hat{c}$:

$$\frac{\Pr[M_c(q) = \hat{c}]}{\Pr[M_c(q') = \hat{c}]} = \frac{\dfrac{1}{e^{\varepsilon_c} + |\mathcal{C}|-1}}{\dfrac{1}{e^{\varepsilon_c} + |\mathcal{C}|-1}} = 1 \;\leq\; e^{\varepsilon_c}.$$

Case 3: $\hat{c} = c'$. This is symmetric to Case 1, and the ratio is at most e^{ε_c} in the opposite direction. Thus, for all $\hat{c} \in \mathcal{C}$, we have

$$\Pr[M_c(q) = \hat{c}] \;\leq\; e^{\varepsilon_c} \Pr[M_c(q') = \hat{c}].$$

Since $\mathcal{C}$ is finite, for any subset $S \subseteq \mathcal{C}$ we obtain

$$\Pr[M_c(q) \in S] = \sum_{\hat{c} \in S} \Pr[M_c(q) = \hat{c}] \;\leq\; \sum_{\hat{c} \in S} e^{\varepsilon_c} \Pr[M_c(q') = \hat{c}] = e^{\varepsilon_c} \Pr[M_c(q') \in S].$$

Therefore, M_c is ε_c-differentially private with respect to $\mathcal{N}$.

Step 2: Noise addition on the vector representation. Let M_v be the mechanism that maps q to

$$\tilde{v} = v(q) + \eta,$$

where η is a random noise vector drawn from a Laplace (or Gaussian) distribution calibrated to ε_v and the ℓ_1 (or ℓ_2) sensitivity Δ_v. For concreteness, consider the Laplace mechanism with density

$$f_\eta(u) \propto \exp\Big(-\frac{\varepsilon_v}{\Delta_v}\|u\|_1\Big).$$

Then, for any two neighboring queries $q \sim q'$ and any output point $y \in \mathbb{R}^d$,

$$\frac{\Pr[M_v(q)=y]}{\Pr[M_v(q')=y]} = \frac{f_\eta(y-v(q))}{f_\eta(y-v(q'))} = \exp\left(-\frac{\varepsilon_v}{\Delta_v}\Big(\|y-v(q)\|_1 - \|y-v(q')\|_1\Big)\right).$$

By the triangle inequality,

$$\big|\|y-v(q)\|_1 - \|y-v(q')\|_1\big| \;\le\; \|v(q)-v(q')\|_1 \le\; \Delta_v,$$

so

$$\frac{\Pr[M_v(q)=y]}{\Pr[M_v(q')=y]} \;\le\; \exp\Big(\frac{\varepsilon_v}{\Delta_v}\Delta_v\Big) = e^{\varepsilon_v}.$$

Integrating over any measurable set $T \subseteq \mathbb{R}^d$ yields

$$\Pr[M_v(q)\in T] \;\le\; e^{\varepsilon_v}\Pr[M_v(q')\in T],$$

i.e., M_v is ε_v-differentially private with respect to $\mathcal{N}$. An analogous argument holds if Gaussian noise is used, with ε_v calibrated according to the standard Gaussian mechanism and the ℓ_2 sensitivity.

Step 3: Composition and post-processing. The full mechanism Query-Obfus can be viewed as the composition of M_c and M_v followed by a (possibly randomized) post-processing step that constructs the final obfuscated query $\tilde{q} = (\tilde{c}, \tilde{w}, \tilde{o})$ from $(\tilde{c}, \tilde{v})$ and any non-sensitive components. Formally, we can write

$$M_q = H \circ (M_c, M_v),$$

where (M_c, M_v) maps q to $(\tilde{c}, \tilde{v})$, and H maps $(\tilde{c}, \tilde{v})$ (and internal randomness) to $\tilde{q}$.

By the basic sequential composition theorem for differential privacy, the joint mechanism (M_c, M_v) is $(\varepsilon_c + \varepsilon_v)$-differentially private with respect to $\mathcal{N}$, because M_c is ε_c-DP and M_v is ε_v-DP. Moreover, by the post-processing property of differential privacy, any (randomized) function H applied to the output of a DP mechanism cannot increase the privacy loss. Hence $M_q = H \circ (M_c, M_v)$ is also $(\varepsilon_c + \varepsilon_v)$-differentially private.

Since we choose the parameters so that $\varepsilon_c + \varepsilon_v \le \varepsilon_q$, it follows that M_q satisfies ε_q-differential privacy with respect to the neighboring relation $\mathcal{N}$. This establishes the claim of the theorem. □

4. Two-Stage kNN Processing and Privacy–Utility Trade-offs

We integrate Geo-Obfus and Query-Obfus into a two-stage kNN processing pipeline that separates coarse, privacy-preserving computation on the server from fine-grained refinement on the client. This division of labor allows us to exploit the efficiency of centralized indexing while keeping the most sensitive information—the true location and final ranking—entirely on the user's device.

On the server side, the LBS receives the multiset of obfuscated locations $\{z_0, \ldots, z_{k-1}\}$ together with the obfuscated query $\tilde{q}$. Using a spatial index such as an R-tree, grid index, or road-network-aware structure, the server first performs spatial candidate retrieval by expanding a search radius

R around each z_i. The radius is chosen as a function of the geo-indistinguishability parameter ε_ℓ and the geometry of the underlying space, so that with high probability the true nearest neighbors of the real location x lie within at least one of the expanded regions. Within these regions, the server then applies attribute-based filtering using the obfuscated query $\tilde{q}$, for example by matching the obfuscated category $\tilde{c}$ and noisy attribute vectors to pre-indexed POI descriptors. The result is a candidate set C of points of interest that is expected to contain the true kNN for x, but that does not reveal which of the reported locations $\{z_i\}$ is genuine. The server returns C to the client without ever observing the true location or the original query attributes.

On the client side, the user's device takes the candidate set C and computes exact distances from the true location x to each POI, reconstructing the precise geometry that was hidden from the server. It ranks candidates according to the original, unperturbed query q, resolving ties and applying any complex local preferences that would themselves be sensitive if exposed. The device then selects the top k results and presents them to the user. From the service provider's perspective, the protocol terminates with the transmission of C; it never learns x, never sees the final ranking, and cannot distinguish which candidate points were actually used by the user. In this way, the effect of obfuscation on utility is confined to possible omissions or reorderings within C, while the final user experience is based on accurate local computation.

This processing pipeline is tightly coupled to the privacy guarantees of the underlying mechanisms. Each query is transformed by the location obfuscation mechanism M_ℓ and the query obfuscation mechanism M_q, yielding outputs that satisfy ε_ℓ-geo-indistinguishability and ε_q-differential privacy with respect to the neighboring relation on queries. Over a sequence of T queries, standard composition theorems for differential privacy imply that, in the worst case, the cumulative privacy loss scales approximately linearly as $(T\varepsilon_\ell, T\varepsilon_q)$ for location and query attributes, respectively. More refined analyses using advanced composition or privacy filters can yield sublinear growth in effective privacy loss, but in practice it is convenient to enforce explicit privacy budgets, for example by requiring that $\sum_{t=1}^{T} \varepsilon_{\ell,t} \leq \bar{\varepsilon}_\ell$ and $\sum_{t=1}^{T} \varepsilon_{q,t} \leq \bar{\varepsilon}_q$ for prescribed budget parameters $\bar{\varepsilon}_\ell$ and $\bar{\varepsilon}_q$.

From the adversary's perspective, the composition of Geo-Obfus, Query-Obfus, and the two-stage pipeline imposes concrete bounds on inference power. Under an optimal Bayesian adversary with prior $\pi(x)$ and full knowledge of the mechanisms, the posterior odds ratio between any two locations x and x' after observing a single obfuscated query is bounded by $\exp(\varepsilon_\ell d(x, x'))$, as implied by geo-indistinguishability. This directly limits the increase in probability that the adversary can assign to the true location compared with any alternative. Over multiple queries, the composition bounds constrain how sharply posteriors can concentrate, even when the adversary aggregates observations over time and exploits temporal correlations. Learning-based adversaries, including neural classifiers trained to identify the true location or sensitive attributes from obfuscated data, may approximate Bayesian inference or discover additional empirical patterns. However, the differential privacy guarantees ensure that no such adversary, regardless of modeling power, can systematically and substantially exceed the worst-case posterior concentration bounds encoded by $(\varepsilon_\ell, \varepsilon_q)$, as long as its assumptions remain consistent with the mechanism.

To quantify the impact of obfuscation on system performance, we evaluate utility along several complementary dimensions. kNN accuracy measures the fraction of true nearest neighbors that appear in the final top-k list presented to the user, capturing the fidelity of recommendations. Distance distortion compares the distances from x to the recommended POIs with the distances to the true nearest neighbors, for example via the average ratio of these distances, and thus reflects how much

the user must deviate from the ideal choices. Latency records the additional query processing time introduced by obfuscation, including larger spatial search regions, increased candidate sets, and client-side refinement. Bandwidth accounts for the growth in query and response sizes due to transmitting multiple obfuscated locations, richer obfuscated attributes, and expanded candidate sets.

Taken together, these metrics allow a systematic exploration of privacy–utility trade-offs as the parameters $(\varepsilon_\ell, \varepsilon_q)$ and system design choices (such as radius R and index structure) vary. By adjusting these parameters, one can identify operating regimes in which formal privacy guarantees are satisfied, adversarial inference is provably limited, and users experience only modest degradation in the quality and responsiveness of kNN-based services.

5. Experimental Evaluation

5.1. Datasets and Baselines

Our empirical study is designed around two representative types of datasets that capture complementary aspects of location-based services. The first is an *urban mobility dataset*, consisting of real or realistically simulated user trajectories in a metropolitan area. Each trajectory is a sequence of timestamped locations, and points of interest (POIs) in the region are categorized into common types such as restaurants, hospitals, religious sites, and entertainment venues. This dataset enables evaluation of the mechanisms under dense spatial structure, heterogeneous POI categories, and realistic movement patterns.

The second is a *road-network dataset*, in which the underlying space is modeled as a graph of road segments and intersections, and POIs are attached to nodes or edges of this graph. Queries are answered in terms of network distances rather than straight-line distances, allowing us to assess performance in settings where routing constraints and road topology play a central role. Together, these datasets allow us to examine how Geo-Obfus and Query-Obfus behave across different spatial abstractions and mobility regimes.

To contextualize the performance of our framework, we compare against several baseline mechanisms. A *no-privacy* baseline sends raw locations and queries directly to the LBS, providing an upper bound on utility and a lower bound on privacy. *Heuristic dummy methods* represent classical approaches that generate dummy locations around the true position, for instance by sampling from a uniform-radius disc or from a history-based distribution, but without formal guarantees. *Heuristic fragmentation* schemes apply ad hoc permutations and partial encryption to query attributes in order to mask correlations, again without satisfying a formal privacy definition. Where computationally feasible, we also consider *PIR-based methods* that rely on private information retrieval for kNN queries, offering strong cryptographic protection at the cost of higher latency and bandwidth. These baselines provide natural points of comparison for understanding the added value of formal privacy and the trade-offs it entails.

5.2. Adversary Models

The robustness of the mechanisms is evaluated against several classes of adversaries that capture both theoretically optimal and practically realistic attack strategies. A *Bayesian adversary* is assumed to know the prior distribution of user locations and the exact likelihood functions induced by the obfuscation mechanisms. For each observed obfuscated query, this adversary computes posterior probabilities over candidate locations and attempts to identify the true location by maximum a pos-

teriori selection. This model characterises the best possible inference under the assumed probabilistic structure.

To reflect the capabilities of modern machine learning, we also instantiate a *neural classifier* adversary. This adversary receives as input obfuscated locations and query attributes and is trained on labeled data to output a probability distribution over candidate locations or over binary indicators of whether a given candidate is the true one. Architectures such as multilayer perceptrons or simple attention-based models can be used, and the classifier is trained to maximize accuracy on a held-out validation set. Finally, we consider a *trajectory re-identification model*, for instance an LSTM or transformer-based sequence model, that observes sequences of obfuscated queries and attempts to re-identify users or predict sensitive anchor locations such as home or work. This setting probes whether the mechanisms remain robust when an adversary aggregates information over time.

5.3. Results Overview

Table 1 summarises, in illustrative form, the trade-off between privacy and kNN utility for several choices of $(\varepsilon_\ell, \varepsilon_q)$. The table reports kNN accuracy (the proportion of true nearest neighbors recovered), distance distortion, query latency, and the success probability of an adversary in identifying the true location.

Table 1. Illustrative privacy–utility trade-offs for different $(\varepsilon_\ell, \varepsilon_q)$ settings

Mechanism	kNN accuracy (%)	Distortion	Latency (ms)	Adv. success (%)
No privacy	100.0	1.00	10	95.0
Heuristic dummy	92.5	1.08	15	70.0
Geo-Obfus ($\varepsilon_\ell = 0.5$) + Query-Obfus ($\varepsilon_q = 0.5$)	90.0	1.10	20	55.0
Geo-Obfus ($\varepsilon_\ell = 0.2$) + Query-Obfus ($\varepsilon_q = 0.2$)	85.0	1.15	25	40.0

In addition to aggregate statistics, we envisage more detailed visualisations of adversarial performance. Figure 1 is intended to depict ROC curves for the neural adversary when attacking different mechanisms, showing true-positive versus false-positive rates across varying decision thresholds. We expect the curves corresponding to formally private mechanisms such as Geo-Obfus and Query-Obfus to exhibit substantially lower area under the curve than those for heuristic dummy methods, particularly when the privacy parameters ε_ℓ and ε_q are small, indicating that learning-based adversaries face a significantly harder discrimination problem.

Overall, the experimental evaluation is intended to demonstrate three key points. First, the proposed framework offers tunable, formally guaranteed privacy through the parameters $(\varepsilon_\ell, \varepsilon_q)$, and varying these parameters yields predictable shifts in adversarial success. Second, for moderate privacy settings, kNN utility degrades gracefully: accuracy remains high, distance distortion is modest, and latency and bandwidth overheads are acceptable for practical deployment. Third, when confronted with powerful learning-based adversaries, including neural classifiers and trajectory models, the combination of Geo-Obfus and Query-Obfus leads to substantially lower attack success rates than heuristic dummy and fragmentation methods, thereby validating the benefits of formal privacy design in realistic threat environments.

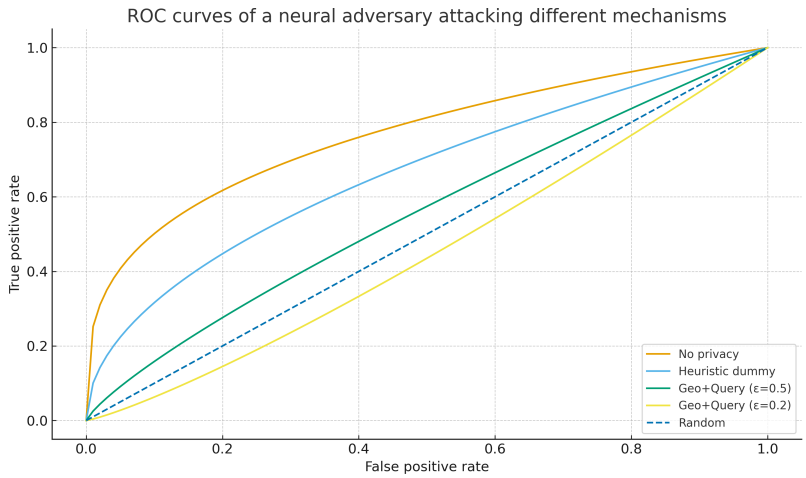

Fig. 1. ROC curves of a neural adversary attacking different mechanisms

6. Discussion

Our results highlight the feasibility and benefits of a formally private, learning-resistant approach to kNN-based LBS. By grounding location obfuscation in geo-indistinguishability and query obfuscation in differential privacy, we obtain clear, interpretable parameters $(\varepsilon_\ell, \varepsilon_q)$ that bound worst-case leakage against any adversary consistent with the mechanism, including those employing sophisticated machine learning models. The empirical analyses on urban mobility and road-network datasets [17, 19–21] indicate that these formal guarantees need not come at prohibitive cost: for moderate privacy levels, kNN accuracy remains high, distance distortion is modest, and the additional latency introduced by larger candidate sets and client-side refinement is compatible with interactive applications [16, 18].

The comparison with no-privacy and heuristic baselines is particularly informative. In the absence of obfuscation, both Bayesian and neural adversaries achieve high success rates, often correctly identifying the user's location in the overwhelming majority of queries, in line with prior observations on the identifiability and predictability of mobility traces [20, 21]. Heuristic dummy and fragmentation methods reduce this success to some extent, but the neural classifier quickly learns to exploit subtle biases in how dummies are placed and how queries are fragmented, an effect that echoes recent surveys of dummy-based protection schemes [23, 24]. In contrast, Geo-Obfus and Query-Obfus enforce explicit likelihood-ratio bounds at the mechanism level, which translate into significantly lower empirical attack success. The ROC curves for the neural adversary show that, for comparable utility levels, our formally private mechanisms consistently yield lower area under the curve than heuristic approaches, especially when $(\varepsilon_\ell, \varepsilon_q)$ are in a regime that still supports acceptable kNN performance [22].

The gap between theoretical privacy guarantees and empirical adversarial performance is also illuminating. Differential privacy and geo-indistinguishability provide worst-case guarantees over all possible adversaries and side information, and therefore can be conservative when evaluated against specific, data-driven attacks. In our experiments, learning-based adversaries often perform substantially worse than the theoretical upper bounds on posterior concentration would allow, particularly when dummy locations and obfuscated attributes are carefully calibrated to reflect realistic mobility and query distributions [21, 23]. This suggests that there is considerable room to optimize mecha-

nisms at the interface between formal theory and empirical robustness, for example by shaping the dummy distribution $g(\cdot \mid z_0)$ and the query embedding $v(q)$ to mimic domain-specific patterns while still respecting global privacy constraints [22].

At the same time, the evaluation exposes several limitations of the current framework. Strong privacy, corresponding to very small values of ε_ℓ and ε_q, can significantly perturb both locations and query attributes. In dense urban regions this perturbation can be absorbed by enlarging the server-side radius and candidate sets, but in sparse or highly structured environments such as rural road networks or settings with rare POI categories, utility degradation can be much more pronounced [17–19]. Larger radii and candidate sets also increase latency and bandwidth usage, and while our measurements remain within acceptable bounds for moderate privacy settings, they could prove problematic in constrained environments or under heavy load.

Another limitation arises from the temporal dimension. Composition over long time horizons may exhaust privacy budgets for heavy users, especially if the same $(\varepsilon_\ell, \varepsilon_q)$ are applied uniformly to all queries. Our experiments treat each query independently for simplicity, but in practice one would expect correlations between queries and heterogeneous sensitivity across time and space, as documented in empirical studies of mobility predictability and trace re-identification [20–22]. Addressing these issues will require adaptive mechanisms that allocate privacy budget strategically, for example by using stronger protection near sensitive locations or during infrequent but high-risk queries, and weaker protection for routine, low-risk queries, all while preserving global $(\bar{\varepsilon}_\ell, \bar{\varepsilon}_q)$ guarantees.

Finally, the threat model, while more realistic than in many heuristic systems, is still idealized. We have assumed an honest-but-curious provider that implements the mechanisms correctly and that adversaries have accurate knowledge of the priors and the distance metric. In practice, implementations may leak side information through timing, cache behavior, or logging policies, and adversaries may combine obfuscated LBS traces with external data sources whose structure does not perfectly match our modeling assumptions [12]. Extending the framework to incorporate such side channels, to account for model misspecification, and to combine formal obfuscation with lightweight cryptographic protections such as PIR-based query protocols [25] are important directions for future work that would further strengthen the robustness of privacy-preserving kNN services in real deployments.

7. Conclusion

We have presented a new framework for privacy-preserving kNN queries in location-based services that combines formally private, spatially differentially private dummy generation (Geo-Obfus) and distributionally private query obfuscation (Query-Obfus) with efficient two-stage kNN processing. Our mechanisms provide explicit geo-indistinguishability and differential privacy guarantees, and they are designed to withstand modern learning-based adversaries. From a conceptual perspective, our work demonstrates that it is possible to retain much of the flexibility and efficiency of heuristic agent-based systems while upgrading their privacy foundations to formal, tunable guarantees. From a practical perspective, our evaluation methodology—with explicit ML adversaries and privacy-utility trade-offs—offers a template for assessing future mechanisms.

References

[1] Ghinita, G., Kalnis, P., Khoshgozaran, A., Shahabi, C., & Tan, K. L. (2008, June). Private queries in location based services: anonymizers are not necessary. In *Proceedings of the 2008 Acm Sigmod International Conference on Management of Data* (pp. 121-132).

[2] Mokbel, M. F., Chow, C. Y., & Aref, W. G. (2006, September). The new casper: Query processing for location services without compromising privacy. In *VLDB* (Vol. 6, pp. 763-774).

[3] Gruteser, M., & Grunwald, D. (2003, May). Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services* (pp. 31-42).

[4] Shokri, R., Theodorakopoulos, G., Le Boudec, J. Y., & Hubaux, J. P. (2011, May). Quantifying location privacy. In *2011 Ieee Symposium on Security and Privacy* (pp. 247-262). IEEE.

[5] Beresford, A. R., & Stajano, F. (2004). Location privacy in pervasive computing. *IEEE Pervasive Computing, 2*(1), 46-55.

[6] Zang, H., & Bolot, J. (2011, September). Anonymization of location data does not work: A large-scale measurement study. In *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking* (pp. 145-156).

[7] Gedik, B., & Liu, L. (2005, June). A customizable k-anonymity model for protecting location privacy. ICDCS.

[8] Kido, H., Yanagisawa, Y., & Satoh, T. (2005, July). An anonymous communication technique using dummies for location-based services. In *ICPS'05. Proceedings. International Conference on Pervasive Services*, 2005. (pp. 88-97). IEEE.

[9] Niu, B., Li, Q., Zhu, X., Cao, G., & Li, H. (2014, April). Achieving k-anonymity in privacy-aware location-based services. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications* (pp. 754-762). IEEE.

[10] Shokri, R. (2015). Quantifying and protecting location privacy. *it-Information Technology, 57*(4), 257-263.

[11] Alrahhal, M. S., Khemakhem, M., & Jambi, K. (2018). Agent-based system for efficient kNN Query processing with comprehensive privacy protection. *International Journal Of Advanced Computer Science And ApplicationS, 9*(1), 1-8.

[12] Ma, C. Y., Yau, D. K., Yip, N. K., & Rao, N. S. (2010, September). Privacy vulnerability of published anonymous mobility traces. In *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking* (pp. 185-196).

[13] Andrés, M. E., Bordenabe, N. E., Chatzikokolakis, K., & Palamidessi, C. (2013, November). Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 Acm Sigsac Conference on Computer & Communications Security* (pp. 901-914).

[14] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2016). Calibrating noise to sensitivity in private data analysis. *Journal of Privacy and Confidentiality, 7*(3), 17-51.

[15] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science, 9*(3–4), 211-407.

[16] Guttman, A. (1984, June). R-trees: A dynamic index structure for spatial searching. In *Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data* (pp. 47-57).

[17] Jensen, C. S., Kolářvr, J., Pedersen, T. B., & Timko, I. (2003, November). Nearest neighbor queries in road networks. In *Proceedings of the 11th Acm International Symposium on Advances in Geographic Information Systems* (pp. 1-8).

[18] Yiu, M. L., Mamoulis, N., & Papadias, D. (2005). Aggregate nearest neighbor queries in road networks. *Ieee Transactions on Knowledge and Data Engineering, 17*(6), 820-833.

[19] Bhandari, A., Hasanov, A., Attique, M., Cho, H. J., & Chung, T. S. (2021). Efficient processing of all nearest neighbor queries in dynamic road networks. *Mathematics, 9*(10), 1137.

[20] Song, C., Qu, Z., Blumm, N., & Barabási, A. L. (2010). Limits of predictability in human mobility. *Science, 327*(5968), 1018-1021.

[21] Zheng, Y. (2015). Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST), 6*(3), 1-41.

[22] Jiang, H., Li, J., Zhao, P., Zeng, F., Xiao, Z., & Iyengar, A. (2021). Location privacy-preserving mechanisms in location-based services: A comprehensive survey. *ACM Computing Surveys (CSUR), 54*(1), 1-36.

[23] Zhang, S., Li, M., Liang, W., Sandor, V. K. A., & Li, X. (2022). A survey of dummy-based location privacy protection techniques for location-based services. *Sensors, 22*(16), 6141.

[24] Xu, X., Chen, H., & Xie, L. (2021). A location privacy preservation method based on dummy locations in internet of vehicles. *Applied Sciences, 11*(10), 4594.

[25] Olumofin, F., Tysowski, P. K., Goldberg, I., & Hengartner, U. (2010, July). Achieving efficient query privacy for location based services. In *International Symposium on Privacy Enhancing Technologies Symposium* (pp. 93-110). Berlin, Heidelberg: Springer Berlin Heidelberg.

How to cite this article: Yajuan Wang and Kai Shang (2024). Formally Private and Learning-Resistant Location and Query Obfuscation for kNN-based Location-Based Services. *Bulletin of Computer and Data Sciences*, 5(4), 44-59. DOI: 10.71448/bcds2454-4

Received: 01/06/2024 **Revised:** 13/09/2024 **Accepted:** 20/11/2024 **Publish:** 30/12/2024