Geo-Graph-Indistinguishability: Protecting Location Privacy for LBS over Road Networks

Shun Takagi, Yang Cao, Yasuhito Asano, Masatoshi Yoshikawa

Geo-Graph-Indistinguishability: Protecting Location Privacy for LBS over Road Networks *

Shun Takagi[0000−0001−7732−2807], Yang Cao[0000−0002−6424−8633], Yasuhito Asano[0000−0002−9095−0127], and Masatoshi Yoshikawa[0000−0002−1176−700X]

Graduate School of Informatics, Kyoto University Dept. of Social Informatics, Japan

Abstract. In recent years, Geo-Indistinguishability (GeoI) has been increasingly explored for protecting location privacy in location-based services (LBSs). GeoI is considered a theoretically rigorous location privacy notion since it extends differential privacy to the setting of location privacy. However, GeoI does not consider the road network, which may cause insufficiencies in terms of both privacy and utility for LBSs over a road network. In this paper, we first empirically evaluate the privacy guarantee and the utility loss of GeoI for LBSs over road networks. We identify an extra privacy loss when adversaries have the knowledge of road networks and the degradation of LBS quality of service. Second, we propose a new privacy notion, Geo-Graph-Indistinguishability (GeoGI), for protecting location privacy for LBSs over a road network and design a Graph-Exponential mechanism (GEM) satisfying GeoGI. We also show the relationship between GeoI and GeoGI to explain theoretically why GeoGI is a more suitable privacy notion over road networks. Finally, we evaluate the empirical privacy and utility of the proposed mechanism in real-world road networks. Our experiments confirm that GEM achieves higher utility for LBSs over a road network than the planar Laplace mechanism for GeoI under the same empirical privacy level.

Keywords: Location Privacy · Geo-Indistinguishability · Road Network · Differential Privacy.

1 INTRODUCTION

In recent years, the spread of smartphones and the improvement of GPS has led to a growing use of location-based services (LBSs). While such services have provided enormous benefits to individuals and society, the exposure of the user's location raises privacy issues. By using the location information, it is easy to identify sensitive personal information, such as that pertaining to home and family. Many methods for protecting location information have been proposed in the past decade. Most of these methods perturb the true location by using a location privacy-preserving mechanism before sending it to an LBS provider or sharing it with a third party. A mechanism takes a true location as input and outputs a perturbed location that follows a probability distribution over a location domain.

* This work was supported by JSPS KAKENHI Grant Numbers (S) No. 17H06099, (A) No. 18H04093, (C) No. 18K11314.

Andrés et al. [10] defined a formal notion of location privacy based on the well-known concept of differential privacy. The definition is called geo-indistinguishability (GeoI), and an output of a mechanism achieving it guarantees indistinguishability of the true location from other locations, that is, strong privacy protection. This notion is derived from differential privacy [5], which provides a rigorous guarantee of indistinguishability of two neighboring databases. One of the most appealing features of GeoI, inherited from differential privacy, is the guarantee of privacy protection against any attacker to a certain degree.

However, since the proposal of GeoI, many studies [3, 14, 25] have identified its weaknesses. Yu et al. [25] showed that GeoI did not protect location privacy against the optimal inference attack [16], and they proposed the framework that adapted GeoI to the expected inference error [17], a complementary notion of GeoI. Chatzikokolakis et al. [3] focused on the fact that GeoI did not consider privacy for semantic information (such as population density) and proposed a mechanism that guarantees the protection of this privacy by using a graph whose weight of edges contains such information. However, this graph does not consider a road network. Oya et al. [14] quantified the privacy against an adversary who knows that the true location is one of two locations. The researchers showed that an output of the mechanism achieving GeoI results in a worse quality of service to protect the user's location privacy.

In this study, we find that GeoI provides inadequate privacy guarantee and insufficient utility for some LBSs over road networks, such as the k-nearest neighbor search (e.g., searching for k restaurants nearest to the user location). In such LBSs, the quality of service can be improved by taking advantage of the road network instead of the Euclidean space. This is because objects can usually move only on a predefined set of trajectories as specified by the road network and it is natural for these LBSs to use the distance on the road network [4, 9, 15]. GeoI may not be practical for LBSs over a road network because it assumes that (1) the perturbed location can be any location in a continuous plane and (2) the distance between locations is measured by the Euclidean distance.

Due to assumption (1), GeoI may result in unexpected privacy loss. For example, as shown in Fig.1, if a user's perturbed location is unreasonable, such as a position on the sea, an adversary can realize that such a location is impossible, which may cause unexpected privacy leakage. Next, due to assumption (2), GeoI may offer inadequate utility for LBSs over a road network. Taking k-nearest neighbor search using shortest path for example, as shown in Fig.2, a user of LBS searching the nearest restaurant expects that restaurant 2 will be returned in a higher probability than restaurant 1. This is because the path to restaurant 1 is farther than to restaurant 1 because of the river. However, if the user uses a GeoI mechanism, the probabilities outputting restaurants 1 and 2 are the same since the two Euclidean distances between restaurants and the user are the same.

In this paper, we study how to protect location privacy while preserving high utility for LBS over a road network. Our contributions are threefold. First, we identify two insufficiencies of GeoI by two empirical evaluations. We quantitatively analyze the change of the inference error w.r.t. adversaries having the

Fig. 1. Perturbation of the location to the unreasonable location.

Fig. 2. Difference of the Euclidean distance and the shortest path length on a road network.

knowledge of road network or not when the location is protected by the planar Laplace mechanism that Andrés et al. [10] proposed, and show its privacy leakage. Additionally, we propose the formulation of the utility of the mechanism for LBSs over a road network. We compute its utility of the planar Laplace mechanism on a graph (PLG), which is our extension of PLM for LBSs over a road network, on two real-world road networks, which shows that it performs poorly w.r.t. this formulation. Second, we propose a new privacy notion based on differential privacy, called Geo-Graph-Indistinguishability (GeoGI) that takes a road network into consideration so that we can construct a more suitable mechanism than GeoI mechanism for protecting location privacy in LBSs over a road network. We design a Graph-Exponential Mechanism (GEM) satisfying GeoGI. We also show the relationship between GeoI and GeoGI to explain theoretically why GeoGI is more suitable privacy notion for LBSs over a road network. Third, to better understand the proposed mechanism, we empirically evaluate privacy and utility of the approach in the case of two real-world road networks in Japan; the results verify that the proposed GEM for GeoGI achieves higher utility than PLM for GeoI when both mechanisms have the same empirical privacy level.

2 Preliminary and Problem Setting

2.1 Geo-Indistinguishability [10]

In this section, we describe the definition of Geo-Indistinguishability (GeoI). Let $\mathcal{X}$ be a set of locations and let $\mathcal{Z}$ be a set of query outputs. Intuitively, a mechanism K achieving GeoI guarantees that $K(x)$ and $K(x')$ are similar to a certain degree for any two locations $x, x' \in \mathcal{X}$. This means that even if an adversary obtains an output of the mechanism, he cannot distinguish the true location from other locations to a certain degree.

The multiplicative distance $d_{\mathcal{P}}$ that expresses the distance between two probability distributions σ_1 and σ_2 on $\mathcal{S}$ is defined as $d_{\mathcal{P}}(\sigma_1, \sigma_2) = \sup_{S \in \mathcal{S}} |\ln \frac{\sigma_1(S)}{\sigma_2(S)}|$ with the convention that $|\ln \frac{\sigma_1(S)}{\sigma_2(S)}| = 0$ if both σ_1 and σ_2 are zero and ∞ if only one of them is zero. $d(x, x')$ represents the Euclidean distance between x and x'. Given $\epsilon \in \mathbb{R}^+$, ϵ-GeoI is defined as follows.

Definition 1. *(ϵ-geo-indistinguishability) The mechanism satisfies ϵ-GeoI iff* $\forall x, x' \in \mathcal{X}, Z \subseteq \mathcal{Z}$

$$d_{\mathcal{P}}(\Pr(K(x) \in Z), \Pr(K(x') \in Z)) \leq \epsilon d(x, x') \tag{1}$$

Mechanism satisfying ϵ-GeoI The authors of [10] introduced a mechanism called planar Laplace mechanism (PLM) for achieving ϵ-GeoI. The probability distribution that PLM generates is called the planar Laplace distribution and, as its name suggests, is derived from a two-dimensional version of the Laplace distribution as follows: $\Pr(PLM_\epsilon(x) = z) = \frac{\epsilon^2}{2\pi} \mathrm{e}^{-\epsilon d(x,z)}$.

2.2 Problem Statement

As we described in Section 1, some LBSs improve their services by using a road network. In this paper, we assume these LBSs, where the road network G is defined as a weighted undirected graph (V, E). Let V be a set of vertices which represent points on the road network, each of which has a coordinate on the Euclidean plane, and let E be a set of edges. The weight $w(a, b)$ of edges between connected vertices $a \in V$ and $b \in V$ represents the shortest distance of the road on the Euclidean plane connecting the two vertices a and b, which leads to $w(a, b) \geq d(a, b)$. Then, $d_s(v, v')$ represents the shortest path length between $v \in V$ and $v' \in V$, and following inequality holds. This is because $d(v, v')$ stands for the shortest distance as the crow flies while $d_s(v, v')$ stands for the shortest distance on the road network.

$$d_s(v, v') \geq d(v, v') \tag{2}$$

In these LBSs, a user who wants to receive the service sends a vertex that represents his location to an untrusted LBS provider, and the LBS provider performs the service's computations w.r.t. the vertex (using the road network) and provides the service. Furthermore, we assume that there is no trusted server. Hence, a user needs to protect privacy on his device by himself.

Quantification of Utility and Privacy Guarantee When a user uses a mechanism to protect his privacy, the quality of the service the user receives degrades. Shokri et al. [16] generally quantified this quality loss, referred to as service quality loss (SQL), in LBSs when a user uses mechanism K:

$$SQL(\pi_u, K, d_q) = \sum_{r, r'} \pi_u(r) \Pr(K(r) = r') d_q(r, r') \tag{3}$$

Here, π_u is the probability distribution representing the probability of user's location, called a prior of the user. $d_q(r, r')$ represents the metric of a degree of dissimilarity, which depends on the LBS. Thus, this means the expected value of a degree of dissimilarity between the actual location of the user and the location obfuscated by mechanism K. Shokri et al. used the Euclidean distance from a

Table 1. Summary of notation

Symbol	Meaning
$LBSs$	Location-based Services.
u, a	A user and an adversary.
$\mathcal{X}$	Set of locations of users on the Euclidean plane.
$\mathcal{Z}$	Set of query outcomes that represent the users' perturbed locations
G	Weighted undirected graph (V, E) that represents a road network.
V	Set of vertices on the Euclidean plane.
E	Set of edges. A weight is the shortest distance on a road connecting two vertices.
$\pi_u(r)$	The probability of being at location r when accessing the LBS.
$\pi_a(r)$	The adversary's knowledge about user's location that represents the probability of being at r.
K	A mechanism. Given a location, K outputs a perturbed location.
$d(x, x')$	Euclidean distance between x and x'.
$d_s(v, v')$	The shortest path length on the graph between v and v'.

natural idea that the longer the Euclidean distance between the true location and the obfuscated location is, the worse the service becomes. We call this SQL_e.

Shokri et al. also quantified the degree of a privacy protection by a mechanism. The researchers translated location privacy into adversarial error (AE) by measuring how accurately an adversary could infer the user's true location. Formally, AE can be formulated as follows:

$$AE(\pi_a, K, h, d_q) = \sum_{\hat{r}, r', r} \pi_a(r) \Pr(K(r) = r') \Pr(h(r') = \hat{r}) d_q(\hat{r}, r) \tag{4}$$

Here, π_a is the probability distribution representing the adversarial knowledge about the user's actual location, called a prior of the adversary. An inference mechanism h outputs an inferred point by drawing a point according to the probability distribution when given an obfuscated point, which stands for the inference of the adversary. Thus, $\Pr(h(r') = \hat{r})$ means the probability of estimating $\hat{r}$ as the actual location of the user when the adversary observes r'. Therefore, AE represents the expected value of a degree of dissimilarity $d_q(\hat{r}, r)$ between the user's true location r and the location $\hat{r}$ the adversary infers. As in the case of SQL, the researchers used the Euclidean distance as d_q.

The major notations in this paper are summarized in Table 1.

3 Evaluating Privacy and Utility of Geo-Indistinguishability

In this section, we empirically show two insufficiencies of GeoI caused by considering a road network. First, we describe the model of an adversary [17]. Then, we show the insufficiency of the privacy guarantee by modeling an adversary who considers the road network and quantifying the accuracy of the attack of this adversary. Next, we describe the formulation of the utility of an output of the

mechanism [16]. Then, we propose the way of applying it for LBSs over a road network, and we show by experimentations with two real-world road networks that PLM performs poorly for this formulation.

3.1 Empirical Privacy Evaluation

We assume that the adversary also uses a road network to infer the user's actual location because a road network is publicly available. In the paper [10], the authors did not consider such an adversary, and we empirically show that if the adversary considers a road network, this may lead to privacy leakage even if the mechanism satisfies GeoI, which is referred to as an insufficiency of the privacy protection.

Adversarial model First, we describe the model of the adversary who tries to infer the user's actual location. Shokri et al. [17] modeled the adversary who knows the prior π_a and the mechanism that the user uses and can solve problems with any computational complexity. Although this assumption is advantageous for the adversary, showing the protection against this adversary will guarantee strong privacy. When the adversary obtains the user's obfuscated location r', he tries to infer the user's true location by the optimal inference attack. In this attack, an adversary solves the following mathematical optimization problem and obtains the optimal probability distribution and constructs the optimal inference mechanism h w.r.t. his knowledge; by using this mechanism with input r', he estimates the user's true location.

$$\begin{aligned} \underset{h}{\text{minimize}} \quad & \sum_{\hat{r},r',r} \pi_a(r) \Pr(K(r) = r') \Pr(h(r') = \hat{r}) d_p(r, \hat{r}) \\ \text{subject to} \quad & \sum_{\hat{r}} h(r')(\hat{r}) = 1, \forall r', \\ & h(r')(\hat{r}) \geq 0, \forall r', \hat{r} \end{aligned} \tag{5}$$

We model an adversary who knows a road network in this way. If an adversary knows a road network, the domain of his prior π_a is V, and d_p is d_s because we assume that the adversary also tries to improve his inference w.r.t. the shortest distance. In this setting, this is a linear programming problem because $\Pr(h(r') = \hat{r})$ represents a variable and the other terms are constant so that the objective function and the constraints are linear. We solve this problem using CBC (coin-or branch and cut)[1] solver of the PuLP library of Python.

Experiment In the following paragraph, we show that the adversary who knows a road network can attack with higher accuracy than can the adversary who does not know it. To make this easy to understand, we use a simple synthetic data illustrated in Figure 3.

[1] https://projects.coin-or.org/Cbc

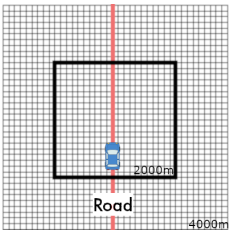

Fig. 3. A synthetic map. The dimensions are 4000 m*4000 m, and each lattice point has a coordinate. The red line indicates the road, and a user is located inside the black frame.

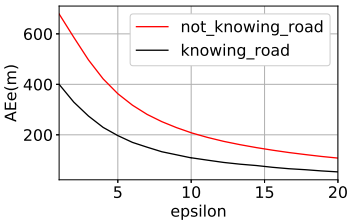

Fig. 4. Adversarial error (AE) in the scenarios of the adversary knowing or not knowing the road network.

This map consists of 1600 squares with the side length of 100 m; that is, the area dimensions are 4000 m * 4000 m, and each lattice point has a coordinate. The red line through the center represents the road where the user is considered to be, and the other area represents locations where the user must not be, such as on the sea. Thus, we assume that the user's location is determined according to a uniform distribution on the red line and inside the black frame to make an output of the mechanism planarly spread; the adversary who does not consider uses the prior given by the uniform distribution inside the black frame, and the adversary who considers the road network uses the prior given by the uniform distribution on the red line inside the black frame. Then, Fig. 4 shows AE_e of each adversary w.r.t. the privacy parameter ϵ of the mechanism. Comparing AE_e of both adversaries, it is clear that the adversary with the prior considering the road network can estimate the user's true location more accurately.

Remark 1 This results from that the PLM could obfuscate the user's location to a place where the user must not be; in the case of Fig. 3, anywhere except the red line. Thus, determining this place contributes to the accuracy of the attack. Fig. 4 shows that an adversarial error could become approximately half in this particular case, so this is an insufficiency of the privacy protection of GeoI.

3.2 Utility

If a user uses LBSs over a road network and uses a mechanism on the Euclidean plane, such as PLM, the user or the LBS provider needs to map the perturbed location to a vertex of the road network because the LBS provider presumes that the user is located at a vertex of the graph to take advantage of a road network. For example, in the LBS that searches for the nearest restaurants, the LBS provider needs to compute the shortest path length between vertices where the user and restaurants are located. If the user is located outside of the graph, the shortest path length cannot be computed. Then, it is worth noting that if the user (rather than the service provider) performed the mapping to the vertex

before the user sent the perturbed location, it would prevent the perturbed location from being the location where the user must not be and would improve the privacy protection. In this view, we propose the mechanism on the graph, which is defined as the algorithm that outputs the perturbed vertex when given a vertex. Then, we can formulate SQL on a road network as SQL_s using the shortest path length as the metric.

$$SQL_s(\pi_u, K) = SQL(\pi_u, K, d_s) = \sum_{v,v'} \pi_u(v) \Pr(K(v) = v') d_s(v, v') \tag{6}$$

In this formulation, we anticipate the lower utility of an output of the mechanism achieving GeoI because it cannot consider SQL_s due to the definition using the Euclidean distance. For example, if there is a river with no bridges so that the user cannot cross it, the opposite riverside is far away and obfuscating to the opposite riverside results in the lower utility (see Fig. 2). However, GeoI may consider the opposite riverside to be close.

Then, we can also formulate the adversarial error over a road network that we call AE_s as follows:

$$\begin{aligned} AE_s(\pi_a, K, h) &= AE(\pi_a, K, h, d_s) \\ &= \sum_{\hat{r},r',r} \pi_u(r) \Pr(K(r) = r') \Pr(h(r') = \hat{r}) d_s(\hat{r}, r) \end{aligned} \tag{7}$$

This formula expresses the expected value of the shortest path length between the actual location and the location that an adversary with the prior of π_a infers with inference mechanism h when a user uses mechanism K. Intuitively, this represents the adversarial error on a road network because we use the shortest path length d_s as the metric. Thus, when the adversary can infer the true location on the road network with a high accuracy, the formula (7) will have a small value. It can be stated that if AE_s is small, the privacy protection level of the mechanism is low.

Experiments Here, we empirically show that the mechanism satisfying GeoI may perform worse w.r.t. SQL_s than we expect, since the definition of GeoI considers SQL as SQL_e. To illustrate this, we compare SQL_s and SQL_e of the same GeoI mechanism. As we stated, we use a mechanism on a graph for LBSs over a road network. Then, we propose a natural and straightforward way of converting PLM to a planar Laplace mechanism on a graph (PLG). First, a user perturbs the location using PLM. Next, the user maps the perturbed location on the Euclidean plane to the nearest vertex on the road network. Fig. 5 is an example of an output of PLG. Formally, we formulate this mechanism as follows:

$$Pr(PLG_\epsilon(v) = w) = \int_{S_w} \Pr(PLM_\epsilon(x_v) = z) dz \tag{8}$$

Here, x_v is the coordinate of vertex v, and let S_w be a Voronoi cell of vertex w when the Voronoi diagram created from the graph on the Euclidean plane is given.

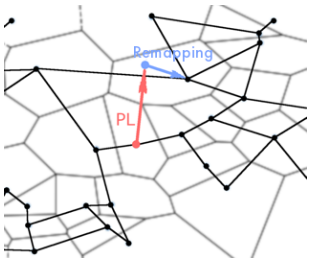

Fig. 5. Output of PLG.

Fig. 6. Tokyo(left) and Akita(right) road networks.

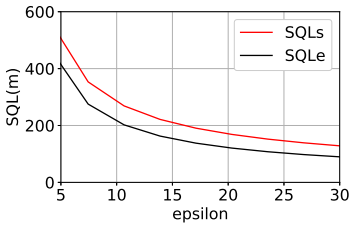

Fig. 7. SQL_s and SQL_e of PLG vs. ϵ on the Tokyo graph.

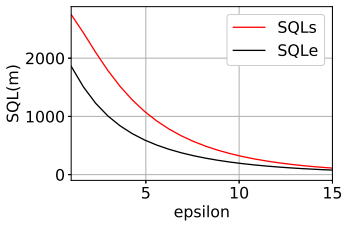

Fig. 8. SQL_s and SQL_e of PLG vs. ϵ on the Akita graph.

Theorem 1. *PLG_ϵ satisfies ϵ-GeoI on the graph.*

We refer the reader to the appendix for the proof. Thus, it is assumed that this is a straightforward way of GeoI mechanism in our setting where a user needs to output a vertex as we stated in Chapter 2.2, and because we cannot use PL, it is reasonable to use this mechanism instead of PL. We compute SQL_s and SQL_e on two road networks. We used OpenStreetMap[2] to retrieve two maps of areas of two cities, Tokyo and Akita, in Japan with dimensions of 4000 m*4000 m as in Fig. 6.

We plot the results in Fig. 7 and Fig. 8 around the range where SQL is reasonable. If the user uses the same mechanism (i.e., the same ϵ), it is observed that the utility for the LBS over a road network is worse. This outcome is caused by the difference of the Euclidean distance and the shortest path length between two vertices. Additionally, it is observed that the difference between SQL_s and SQL_e on the Akita graph is larger than that on the Tokyo graph at the same SQL_e because the difference between the two distances on the Akita graph is larger than that on the Tokyo graph.

Remark 2 GeoI constrains the mechanism to use the Euclidean distance so that the mechanism cannot improve its utility of the output for LBSs using a road network. Regardless of how hard we try to improve the utility of the mechanism output, as long as there is the constraint of GeoI, the mechanism cannot consider

[2] https://openstreetmap.jp/

a road network, and we cannot improve the mechanism. Additionally, we showed that SQL_s of PLG is worse on two graphs; however, there may be a road network that results in much worse utility than we showed. Unless the mechanism considers a road network, the mechanism cannot guarantee high utility.

4 GEO-GRAPH-INDISTINGUISHABILITY

In this section, we propose a new definition of location privacy called Geo-Graph-Indistinguishability (GeoGI) for LBSs using a road network, which is tolerant to the weaknesses of GeoI. We first formally define GeoGI, and then propose a mechanism satisfying GeoGI which is called Graph-Exponential Mechanism (GEM). Finally, we clarify the relationship between GeoI and GeoGI, and describe validity of the definition of GeoGI.

4.1 Definition

Given a graph $G = (V, E)$ representing a road network, let $\mathcal{W}$ be a set of vertices that a mechanism outputs. Then, mechanism K on the graph returns the random vertex $w \in \mathcal{W}$ according to a probability distribution when given a vertex $v \in V$. Then, given $\epsilon \in \mathbb{R}^+$, ϵ-Geo-Graph-Indistinguishability is defined as follows.

Definition 2. *(ϵ-Geo-Graph-Indistinguishability) A mechanism K on a road network $G = (V, E)$ satisfies ϵ-Geo-Graph-Indistinguishability iff $\forall v, v' \in V, \forall W \subseteq \mathcal{W}$,*

$$d_{\mathcal{P}}(\Pr(K(v) \subseteq W), \Pr(K(v') \subseteq W) \leq \epsilon d_s(v, v') \tag{9}$$

The definition can be also formulated as $\forall v, v' \in V, \forall W \subseteq \mathcal{W}, \frac{\Pr(K(v) \subseteq W)}{\Pr(K(v') \subseteq W)} \leq \mathrm{e}^{\epsilon d_s(v,v')}$. This formulation implies that GeoGI is an instance of $d_{\mathcal{X}}$-privacy [7] proposed by Chatzikokolakis et al. as are GeoI and differential privacy. The authors showed that an instance of $d_{\mathcal{X}}$-privacy guaranteed strong privacy. We refer the reader to the appendix for further details. Intuitively, this definition guarantees that for any $v, v' \in V$, the closer to v' a vertex v is w.r.t. the shortest path length, the more similar $K(v)$ and $K(v')$ are.

It is worth noting that the definition of GeoGI includes a given graph representing a road network, and this results in the privacy protection level and utility varying depending on the road network even if the privacy parameter ϵ remains the same.

4.2 Graph Exponential Mechanism

In this section, we propose a mechanism that achieves ϵ-GeoGI. Given parameter $\epsilon \in \mathbb{R}^+$ and graph $G = (V, E)$, we define GEM_ϵ for any user's location $v \in V$ and perturbed location $w \in \mathcal{W}$ as follows.

Definition 3. *GEM_ϵ takes v as an input and outputs w with the following probability.*

$$\Pr(GEM_\epsilon(v) = w) = \alpha(v)\mathrm{e}^{-\frac{\epsilon}{2} d_s(v,w)} \tag{10}$$

where α is a normalization factor and $\alpha(v) = \frac{1}{\sum_{w \in V} \mathrm{e}^{-\frac{\epsilon}{2} d_s(v,w)}}$.

The pseudocode of GEM is described in Appendix (Section 8.3) due to space limitation. This mechanism employs the idea of exponential mechanism [12] that is one of the general mechanisms for achieving differential privacy.

Theorem 2. *GEM_ϵ satisfies ϵ-GeoGI.*

We refer the reader to the appendix for the proof. This mechanism considers a road network so that high utility for LBSs over a road network can be expected. Moreover, since this mechanism satisfies GeoGI, strong privacy based on differential privacy is guaranteed.

Creating the Probability Distribution and Drawing a Random Point Because we assume that the LBS provider is untrusted and there is no trusted server, a user needs to create this distribution by himself and choose the perturbed vertex according to the distribution. In this section, we describe a method to do this and its issues caused by the number of vertices.

To create the probability distribution, (i) the user gets shortest path lengths to all vertices from the vertex where the user is located. (ii) Then the user computes $\mathrm{e}^{-\frac{\epsilon}{2}d_s}$ and based on this distribution, (iii) chooses a point.

Phase (i) is acceptable if the server which has enough computing power computes the all shortest lengths and sends users it in advance. This is because the shortest path length can be computed by Dijkstra's algorithm; this computational complexity of this operation depends on the data structure. if we use a naive method, it is $O(|E| + |V|^2)$, and it can be improved by using Fibonacci heap to $O(|E| + |V| \log |V|)$, where $|V|$ and $|E|$ represent the counts of edges and vertices. However, There is a problem if the user needs to compute it and the size of vertices is large because the user uses a mobile phone with limited computing power. So we have to consider the better algorithm. On a road network, a fast algorithm computing the shortest path length has been studied; we refer the reader to [1] that may be applied to our algorithm. Phase (ii) is no problem because it computational complexity is $O(|V|)$. For phase (iii), when the number of vertices is much larger than we expected, we may not be able to effectively sample the vertices according to the distribution. This problem has also been studied and is known as consistent weighted sampling (CWS): we refer the reader to [11, 23]. We believe that these studies can be applied to our algorithm and it can be computed even if the size of vertices is somewhat large.

4.3 Analyzing the relationship between GeoI and GeoGI

In this section, we describe the relationship between GeoI and GeoGI. There are two major differences: one is the domain, and the other is the distance metric.

First, we state the difference of the location domain. We can design a GeoI mechanism on the Euclidean plane; however, the same cannot be done for GeoGI because GeoGI constrains a mechanism to use a vertex on the graph due to the definition using the shortest path length. Since we exploit the mechanism for LBSs using a road network, the constraint does not pose a problem. Moreover, this constraint prevents the perturbation to a location where the user must not

be located and improve the privacy guarantees, as we stated in Section 3.1. We refer the mechanism which meets this constraint as a mechanism on the graph.

The other difference is the used metric. In this part, we assume the mechanism on a graph; otherwise, we cannot design a GeoGI mechanism. Then, the following theorem holds due to the used metric of GeoGI.

Theorem 3. *If a mechanism on the graph satisfies ϵ-GeoI, it satisfies ϵ-GeoGI.*

This is due to the definition of a graph that represents the road network. Using Inequality (2), we derive the following inequality:

$$d_{\mathcal{P}}(\Pr(K(v) \subseteq (W)), \Pr(K(v') \subseteq (W))) \leq \epsilon d(v, v') \leq \epsilon d_s(v, v') \tag{11}$$

This inequation shows that the GeoI mechanism is also the GeoGI mechanism, but the reverse is not always true. For example, PLG satisfies both GeoI and GeoGI. This means that GeoGI relaxes the restriction of GeoI. Thus, we can design a more suitable mechanism which improves the utility for a road network.

It is worth noting whether this relaxing of the definition leads to weakening of the guarantees of privacy protection. In short, GeoGI has no guarantees of privacy protection w.r.t. the Euclidean distance so that if a user uses a mechanism that satisfies GeoGI to protect the location, the adversary may easily distinguish the user's location in terms of the Euclidean distance. In what follows, we show this fact using the notion of the true probability (TP). The probability that an adversary can distinguish user's location is represented as

$$TP(\pi_u, K, h) = \sum_{\hat{v},v,w} \pi_u(v) \Pr(K(v) = w) \Pr(h(w) = \hat{v}) \delta(\hat{v}, v) \tag{12}$$

Here $\delta(\hat{v}, v)$ is a function that returns 1 if $\hat{v} = v$ holds and otherwise returns 0. TP means the expected value of a probability that an adversary can remap the obfuscated location to the true location.

We assume a set of graphs, each of which has only two vertices. The Euclidean distance between the vertices is the same for all the graphs, but the weight of the edge between them is different for each graph (Fig. 9). Next, we assume that each prior the user and the adversary have is a uniform distribution on two vertices of this graph, and we compute TP of PLG and GEM. Fig. 10 shows the change of TP when the weight, that is, the shortest path length, changes. Due to the guarantee of the Euclidean distance of GeoI, PLG does not degrade TP even if the shortest path length changes, however, since GeoGI does not have a guarantee of the Euclidean distance, GEM significantly degrades TP, which means that the adversary can know the user's true location.

GeoGI can achieve better utility than can GeoI by guaranteeing privacy protection in terms of the shortest path length instead of the Euclidean distance. This idea comes from the interpretation of privacy; in this paper, we assume that privacy and the utility can be interpreted as the shortest path length on the graph and that it should be acceptable for LBSs on a road network. Therefore, GeoGI may not be suitable for protecting location privacy if the privacy should be interpreted as the Euclidean distance, e.g., querying the weather conditions where we need to protect a wide range of locations.

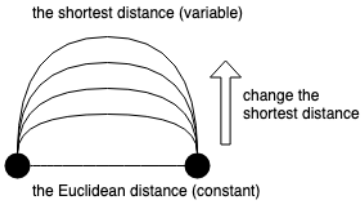

Fig. 9. Changeing the shortest path length of the graph.

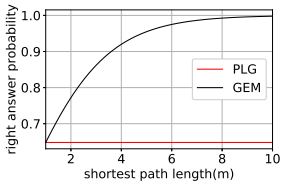

Fig. 10. Change of TP according to GEM and PLG.

4.4 Discussion

We assume that a graph representing the road network is given in GeoGI. We note that setting a different graph (road network) in the definition of GeoGI implies a different privacy model. Thus, GeoGI can be personalized. For example, a conservative user may want to use a global graph in GeoGI that covers all possible locations on Earth, while a liberal user may use a smaller graph in GeoGI that only covers the city of her residence. The graph may also depend on application scenarios in practice. For example, if the application is vehicle navigation, the graph should cover all highway road network instead of pedestrian lanes. In summary, the privacy level and utility depend on the shape of the graph, such as its density and size, and its relationships should be shown because this has the potential to lead to improvement of the privacy protection and utility. This topic is left to future research.

Additionally, the utility and the privacy protection level depend on the vertex where the user uses the GeoGI mechanism because, as opposed to the Euclidean plane that spreads uniformly, each vertex relates differently with other vertices and because to satisfy GeoGI, the mechanism needs to vary the probability distribution depending on the vertex's relationship. This complexity obscures the mechanism performance for a user, and a user will not know how to adjust the privacy parameter. Then, we propose a way of measuring the performance of the mechanism used by a user located at a certain vertex. We formulate the mechanism's utility as SQLr_s and its privacy protection level as AEr_s as follows:

$$SQLr_s(v, K) = \sum_{v'} \Pr(K(v) = v') d_s(v, v') \tag{13}$$

$$AEr_s(v, K, h) = \sum_{v', \hat{v}} \Pr(K(r) = r') \Pr(h(v') = \hat{v}) d_s(\hat{v}, v) \tag{14}$$

When a user is located at vertex v, SQLr_s represents the expected value of the shortest path length between v and the perturbed vertex v'. AEr_s represents the expected value of the shortest path length between v and vertex $\hat{v}$ inferred by the adversary with inference mechanism h (in this case, we assume optimal inference attack). We show SQLr_s and AEr_s of the Akita graph using GEM with $\epsilon = 0.002$ in Figures 11 and 12, respectively. As we can see, the utility loss (i.e.,

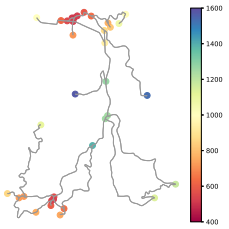

Fig. 11. SQLr_s of each vertex on Akita graph.

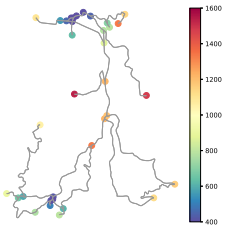

Fig. 12. AEr_s of each vertex on Akita graph.

SQLr_s) and privacy (i.e., AEr_s) differ on different locations in spite of the same privacy parameter. We can develop a tool to visualize the privacy and utility of the mechanism under different privacy parameters, which may help users to determine a proper privacy parameter. We defer this to a long version of this work.

5 EXPERIMENTS

In this section, we show that GEM outperforms the GeoI mechanism in terms of utility and privacy protection for LBSs on a road network. To demonstrate this conclusion, we performed two experiments as follows. First, since GEM, in contrast to PL, may perturb the input location to a location that is out of the road network, such as on the sea (as stated in Section 3.1), it is assumed that GEM achieves better privacy guarantee when the adversaries have the knowledge of road networks. To show this, we computed AE_e of GEM on the synthetic graph we used in section 3.1. Next, because GEM, in contrast to PLG, considers a road network (as stated in Section 3.2), it is assumed that the output quality of GEM is higher than PLG. To show this, we computed SQL_s of GEM for two cities we used in section 3.2. Additionally, we compared the results with those of PL and PLG.

5.1 Privacy Protection Level of GEM

We computed AE_e of GEM on the graph of Fig 3. As in section 3.1, we assume that the adversary knew the road network so that his prior was a uniform distribution on the red line inside the black frame. Since GEM, in contrast to PL, outputs only the locations on the red line, it is assumed that AE_e of GEM is higher than that of PLM. To fairly compare AE_e of each mechanism, we performed the comparison under the same utility SQL_e.

As is shown in Fig 13, AE_e of GEM is higher than that of PL in case of the adversary who knows a road network. This means that GE can protect user privacy more strongly than can PL because GE guarantees that the output is on the road network.

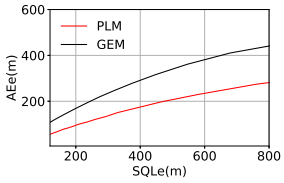

Fig. 13. AE_e when using GEM and PLM against the adversary who knows the road network.

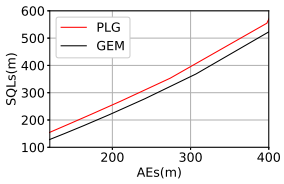

Fig. 14. SQL_s of PLG and GEM w.r.t. AE_s on the Tokyo graph.

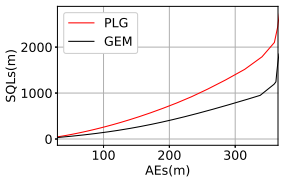

Fig. 15. SQL_s of PLG and GEM w.r.t. AE_s on the Akita graph.

5.2 Utility of GEM

We computed SQL_s of GEM on two graphs of Fig 6. Since GEM considers the road network, it is assumed that SQL_s of GEM is higher than that of PLG. To fairly compare SQL_s of each mechanism, we performed the comparison under the same AE_s. Then, we assumed that both the priors that a user and an adversary have are uniform distributions on the graph with a range of 2000 m from the centers of maps.

As we can see from Figures 14 and 15, SQL_s of GEM is lower than that of PLG. Thus, a GEM output has higher utility for LBSs using a road network than does a PLG output. Additionally, it can be said that the difference of the SQL between PLG and GEM is larger on the Akita graph than on the Tokyo graph. The reason is that the difference between the Euclidean distance and the shortest path length is larger for vertices of the Akita graph.

6 RELATED WORK

6.1 Location Privacy on a Road Network

To the best of our knowledge, this is the first study to propose the perturbation with the differential privacy approach over the road network. However, several studies explored location privacy on a road network.

Tyagi et al. [20] studied location privacy over a road network for VANET users, and they show that there are no comprehensive privacy-preserving techniques or frameworks that cover all privacy requirements or issues to maintain the desired level of location privacy.

Wang et al. [21] and Wen et al. [22] proposed the method of privacy protection for the user who wishes to receive location-based services and travels over roads. The authors use k-anonymity as the protection method and take advantage of the road network constraints.

A series of key features distinguish our solution from these studies: a) we use the differential privacy approach so that our solution has a guarantee of privacy protection against any attacker and b) we assume that there is no trusted server. We highlight these two points as advantages of our proposed method.

6.2 State-of-the-Art Privacy Models

Since GeoI was published, many related applications have been proposed. To et al. [18] developed an online framework of privacy-preserving spatial crowdsourcing service using GeoI. Tong et al. [19] proposed a framework of privacy-preserving ridesharing service based on GeoI and differential privacy approach. It may be possible to improve these applications by using GeoGI instead of GeoI. Additionally, Bordenabe et al. [13] proposed an optimized mechanism satisfying GeoI; it may be possible to apply this method to GeoGI.

According to [10], if a user uses the GeoI mechanism multiple times, this causes privacy degradation due to correlations in the data; this scenario also applies to GeoGI. This issue remains a difficult and intensely investigated problem in the field of differential privacy. There are two kinds of approaches attempting to solve this problem. The first is to develop a mechanism for multiple perturbations that satisfies existing notion, such as differential privacy and GeoI [8, 6]. Kairouz et al. [8] studied the composition theorem and proposed a mechanism that upgrades the privacy guarantee. Chatzikokolakis et al. [6] proposed a method of controlling privacy using GeoI when locations are correlated. The second approach is to propose a new privacy notion for correlated data [24, 2]. Xiao et al. [24] proposed δ-location set privacy to protect each location in a trajectory when a moving user sends locations. Cao et al. [2] proposed PriSTE, a framework for protecting spatiotemporal event privacy. We believe that these studies can be applied to our work.

7 CONCLUSION AND FUTURE WORK

In this paper, by evaluating privacy and utility of PL, we have shown that the definition of GeoI is insufficient for LBSs over a road network to protect privacy and output the useful perturbed location. The core of our proposal is a new notion of privacy that we call GeoGI, which takes the place of GeoI for such LBSs, and a mechanism GEM, based on the exponential mechanism, to perturb the user location. We have shown how GeoGI relates to GeoI and that GeoGI is a more suitable privacy definition for such LBSs w.r.t. privacy protection and utility. We also have shown the effectiveness of our proposed approach by comparing GEM with PLG in the example of two cities in Japan.

In the future, we aim to extend the privacy model to several graphs. Although in this paper, we represented a road network as an undirected graph, it should be represented as a directed graph because of the existence of one-way roads, and this may degrade the utility. Additionally, we need to consider the movement mode such as walking, driving, and flying. Finally, we need to pay attention to the fact that multiple perturbations of correlated data such as trajectory data may degrade the level of protection even if the mechanism satisfies GeoGI as in case of GeoI and differential privacy. This topic has been intensely studied, and we believe that it can be applied to GeoGI. We plan to solve these problems in future research.

References

1. Akiba, T., Iwata, Y., Kawarabayashi, K.i., Kawata, Y.: Fast shortest-path distance queries on road networks by pruned highway labeling. Proceedings of the Sixteenth Workshop on Algorithm Engineering and Experiments (ALENEX) pp. 147–154 (2014)
2. Cao, Y., Xiao, Y., Xiong, L., Bai, L.: Priste: From location privacy to spatiotemporal event privacy. arXiv preprint arXiv:1810.09152 (2018)
3. Chatzikokolakis, K., Palamidessi, C., Stronati, M.: Constructing elastic distinguishability metrics for location privacy. Proceedings on Privacy Enhancing Technologies (2), 156–170 (2015)
4. Cho, H.J., Chung, C.W.: An efficient and scalable approach to CNN queries in a road network. Proceedings of the 31st international conference on Very large data bases pp. 865–876 (2005)
5. Dwork, C.: Differential privacy. Encyclopedia of Cryptography and Security pp. 338–340 (2011)
6. K. Chatzikokolakis, C.P., Stronati, M.: A predictive differentially-private mechanism for mobility traces. Privacy Enhancing Technologies pp. 21–41 (2014)
7. K. Chatzikokolakis, M. E. Andrés, N.E.B., Palamidessi, C.: Broadening the scope of differential privacy using metrics. Privacy Enhancing Technologies pp. 82–102 (2013)
8. Kairouz, P., Oh, S., Viswanath, P.: The composition theorem for differential privacy. IEEE Transactions on Information Theory **63**(6), 4037–4049 (2017)
9. Kolahdouzan, M., Shahabi, C.: Voronoi-based k nearest neighbor search for spatial network databases. Proceedings of the Thirtieth international conference on Very large data bases-Volume 30 pp. 840–851 (2004)
10. M. E. Andrés, N. E. Bordenabe, K.C., Palamidessi, C.: Geo-indistinguishability: Differential privacy for location-based systems. in Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security pp. 901–9134 (2013)
11. Manasse, M., McSherry, F., Talwar, K.: Consistent weighted sampling (June 2010)
12. McSherry, F., Talwar, K.: Mechanism design via differential privacy. 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS) pp. 94–103 (Oct 2007)
13. N. E. Bordenabe, K.C., Palamidessi, C.: Optimal geo-indistinguishable mechanisms for location privacy. in Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, New York, NY, USA pp. 251–262 (2014)
14. Oya, S., Troncoso, C., Pérez-González, F.: Is geo-indistinguishability what you are looking for? Proceedings of the 2017 on Workshop on Privacy in the Electronic Society pp. 137–140
15. Papadias, D., Zhang, J., Mamoulis, N., Tao, Y.: Query processing in spatial network databases. Proceedings of the 29th international conference on Very large data bases pp. 802–813 (2003)
16. Shokri, R., Theodorakopoulos, G., Boudec, J.Y.L., Hubaux, J.P.: Quantifying location privacy. Proceedings of the IEEE symposium on security and privacy pp. 247–262 (2011)
17. Shokri, R., Theodorakopoulos, G., Troncoso, C., Hubaux, J.P., Boudec, J.Y.L.: Protecting location privacy: optimal strategy against localization attacks. Proceedings of the 2012 ACM conference on Computer and communications security pp. 617–627

18. To, H., Ghinita, G., Shahabi, C.: A framework for protecting worker location privacy in spatial crowdsourcing. Proceedings of the VLDB Endowment **7**(10), 919–930 (2014)
19. Tong, W., Hua, J., Zhong, S.: A jointly differentially private scheduling protocol for ridesharing services. IEEE Transactions on Information Forensics and Security **12**(10), 2444–2456 (2017)
20. Tyagi, A.K., Sreenath, N.: Location privacy preserving techniques for location based services over road networks. Proceedings of International Conference on Communications and Signal Processing (ICCSP) pp. 1319–1326 (April 2015)
21. Wang, T., Liu, L.: Privacy-aware mobile services over road networks. Proceedings of the VLDB Endowment **2**(1), 1042–1053 (2009)
22. Wen, J., Li, Z.: A method of location privacy protection in road network environment. 2018 International Conference on Smart Materials, Intelligent Manufacturing and Automation (SMIMA) **173**(03048)
23. Wu, W., Li, B., Chen, L., Zhang, C., Yu, P.S.: Improved Consistent Weighted Sampling Revisited. arXiv:1706.01172 [cs] (Jun 2017)
24. Xiao, Y., Xiong, L.: Protecting locations with differential privacy under temporal correlations. Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security - CCS '15 pp. 1298–1309
25. Yu, L., Liu, L., Pu, C.: Dynamic differential location privacy with personalized error bounds. Proceedings on Netw. Distrib. Syst. Security Symp. (NDSS) (2017)

8 APPENDIX

8.1 Proofs

Theorem 1. *PLG_ϵ on graph $G = (V, E)$ satisfies ϵ-GeoI on graph G.*

Proof. This proposition can be formulated as follows for all vertices $v, v' \in V, W \subseteq \mathcal{W}$,

$$d_p(\Pr(PLG(v) \subseteq W), \Pr(PLG(v') \subseteq W)) \leq \epsilon d(v, v') \tag{15}$$

Furthermore, we derive

$$\forall v, v' \in V, w \in W, \Pr(PLG(v) = w) \leq \mathrm{e}^{\epsilon d(v,v')} \Pr(PLG(v') = w) \tag{16}$$

Since PL_ϵ satisfies ϵ-GeoI, for all $z \in \mathcal{Z}$, we derive

$$\Pr(PL_\epsilon(x_v) = z) \leq \mathrm{e}^{\epsilon d(x_v, x_{v'})} \Pr(PL_\epsilon(x_{v'}) = z) \tag{17}$$

By the theorem of integral inequality we obtain

$$\begin{aligned} \int_{S_w} \Pr(PL_\epsilon(x_v) = z)dz &\leq \int_{S_w} \mathrm{e}^{\epsilon d(x_v, x_{v'})} \Pr(PL_\epsilon(x_{v'}) = z)dz \\ &= \mathrm{e}^{\epsilon d(x_v, x_{v'})} \int_{S_w} \Pr(PL_\epsilon(x_{v'}) = z)dz \end{aligned} \tag{18}$$

Using (8) and (18), we obtain

$$\Pr(PLG(v) = w) \leq \mathrm{e}^{\epsilon d(v,v')} \Pr(PLG(v') = w) \tag{19}$$

This concludes the proof.

Theorem 2. *GEM_ϵ satisfies ϵ-GeoGI.*

Proof. This proposition can be formulated for all vertices $v, v' \in V, W \subseteq \mathcal{W}$:

$$d_p(\Pr(GEM(v) \subseteq W), \Pr(GEM(v') \subseteq W)) \leq \epsilon d_s(v, v') \tag{20}$$

The ratio of $\Pr(GEM(v) = w)$ and $\Pr(GEM(v') = w)$ is expressed as follows:

$$\frac{\Pr(GEM(v) = w)}{\Pr(GEM(v') = w)} = \frac{\alpha(v)\mathrm{e}^{-\frac{\epsilon}{2}d_s(v,w)}}{\alpha(v')\mathrm{e}^{-\frac{\epsilon}{2}d_s(v',w)}} = \frac{a(v)}{a(v')}\mathrm{e}^{\frac{\epsilon}{2}(d_s(v',w)-d_s(v,w))} \tag{21}$$

When $-d_s(v, w) + d_s(v', w)$ has the maximum value for $w \in W$, (21) reaches the maximum value too. Due to the triangle inequality, the inequality $\forall w \in W, -d_s(v, w) + d_s(v', w) \leq -d_s(v, w) + d_s(v', v) + d_s(v, w) = d_s(v, v')$ holds, and the following inequality is derived:

$$\frac{\Pr(GEM(v) = w)}{\Pr(GEM(v') = w)} \leq \frac{\alpha(v)}{\alpha(v')}\mathrm{e}^{\frac{\epsilon}{2}d_s(v,v')} \tag{22}$$

Next, we show that the following inequality holds:

$$\frac{\alpha(v)}{\alpha(v')} < \mathrm{e}^{\frac{\epsilon}{2}d_s(v,v')} \tag{23}$$

The inequality 23 is expressed as follows for any $v, v' \in V, w \in W$ of any graph G:

$$\sum_{w \in V} \mathrm{e}^{-\frac{\epsilon}{2}d_s(v',w)} - \mathrm{e}^{\frac{\epsilon}{2}d_s(v,v')} \sum_{w \in V} \mathrm{e}^{-\frac{\epsilon}{2}d_s(v,w)} < 0 \tag{24}$$

Using the triangle inequality, we have $\forall w \in W, d(v, w) - d(v, v') \leq d_s(v', w)$, $\mathrm{e}^{-\frac{\epsilon}{2}d_s(v',w)} \leq \mathrm{e}^{-\frac{\epsilon}{2}(d_s(v,w)-d_s(v,v'))}$. Therefore,

$$\sum_{w \in V}(\mathrm{e}^{-\frac{\epsilon}{2}d_s(v',w)} - \mathrm{e}^{-\frac{\epsilon}{2}(d_s(v,w)-d_s(v,v'))}) \leq \sum_{V \setminus v}(\mathrm{e}^{-\frac{\epsilon}{2}d_s(v',V)} - \mathrm{e}^{-\frac{\epsilon}{2}d_s(v',V)}) < 0$$

Using (22) and (23), we obtain

$$\frac{\Pr(GEM(v) = w)}{\Pr(GEM(v') = w)} < \mathrm{e}^{\frac{\epsilon}{2}d_s(v,v')}\mathrm{e}^{\frac{\epsilon}{2}d_s(v,v')} = \mathrm{e}^{\epsilon d_s(v,v')} \tag{25}$$

8.2 $d_{\mathcal{X}}$-privacy

As we stated in section 4, GeoGI is an instance of $d_{\mathcal{X}}$-privacy: due to this characterization, we can give two characterizations of GeoGI that mathematically show the guarantee of strong privacy protection. In this section, we stated the characterizations of GeoGI.

Hiding function The first characterization uses the concept of a hiding function $\phi : V \to V$. For any hiding function and a secret location $v \in V$, when an attacker who has a prior distribution that expresses the user's location information obtains each output $w \sim K(v), w' \sim K(\phi(v))$ of a mechanism satisfying ϵ-GeoGI, the following inequality holds for the multiplicative distance between its two posterior distributions:

$$d_{\mathcal{P}}(p(v|w), p(v|w')) \leq 2\epsilon d_s(\phi) \tag{26}$$

Let $d_s(\phi(v)) = sup_{v \in V} d_s(v, \phi(v))$ be the maximum distance between an actual vertex and its hidden version. This inequality guarantees that the adversary's conclusions are the same (up to $2\epsilon d_{\mathcal{X}}(\phi)$) regardless of whether ϕ has been applied or not.

Informed attacker The other characterization is shown by the multiplicative distance between the prior distribution and its posterior distribution that is derived by obtaining an output of the mechanism. By measuring its distance, we can determine how much the adversary has learned about the secret. We assume that an adversary (informed attacker) knows that the vertex v where the user is located in N. When the adversary obtains an output of the mechanism. The following inequality holds for the multiplicative distance between his prior distribution $\pi_{|N}(v) = \pi(v|N)$ and its posterior distribution $p_{|N}(v|w) = p(v|w, N)$:

$$d_{\mathcal{P}}(\pi_{|N}, p_{|N}(v|w)) \leq \epsilon d_s(N) \tag{27}$$

Let $d_s(N) = max_{v,v' \in N} d_s(v, v')$ be the maximum distance between vertices in N. This inequality guarantees that when $d_s(N)$ is small, the adversary's prior distribution and its posterior distribution are similar. In other words, the more the adversary knows about the actual location, the less he cannot learn about the location from an output of the mechanism.

8.3 Pseudocode of GEM

Algorithm 1: Graph Exponential Mechanism (GEM).

Input: v, G, ϵ.
Output: Sanitized location w of input v.

initialization;
Compute shortest distances to all other vertices from v by Dijkstra's algorithm and calculate $\mathrm{e}^{-\frac{\epsilon}{2} d_s}$;
Normalize to make a distribution ;
Draw random vertex w according to the distribution;
return w.