Hindawi
Security and Communication Networks
Volume 2021, Article ID 9635411, 15 pages
https://doi.org/10.1155/2021/9635411

Research Article

KLPPS: A *k*-Anonymous Location Privacy Protection Scheme via Dummies and Stackelberg Game

Dongdong Yang,[1,2] **Baopeng Ye**,[3] **Wenyin Zhang**,[4] **Huiyu Zhou**,[5] **and Xiaobin Qian**[6]

[1]*State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang 550025, China*
[2]*Guangxi Key Laboratory of Cryptography and Information Security, Guilin, China*
[3]*Information Technology Innovation Sercive Center of Guizhou Province, Guiyang, China*
[4]*School of Information Science and Engineering, Linyi University, Linyi, Shandong 276 000, China*
[5]*School of Informatics, University of Leicester, Leicester, UK*
[6]*Guizhou CoVision Science and Technology Co., Ltd, Guiyang, China*

Correspondence should be addressed to Wenyin Zhang; zhangwenyin@lyu.edu.cn

Received 13 October 2021; Revised 10 November 2021; Accepted 18 November 2021; Published 7 December 2021

Academic Editor: Xin-Yi Huang

Protecting location privacy has become an irreversible trend; some problems also come such as system structures adopted by location privacy protection schemes suffer from single point of failure or the mobile device performance bottlenecks, and these schemes cannot resist single-point attacks and inference attacks and achieve a tradeoff between privacy level and service quality. To solve these problems, we propose a *k*-anonymous location privacy protection scheme via dummies and Stackelberg game. First, we analyze the merits and drawbacks of the existing location privacy preservation system architecture and propose a semitrusted third party-based location privacy preservation architecture. Next, taking into account both location semantic diversity, physical dispersion, and query probability, etc., we design a dummy location selection algorithm based on location semantics and physical distance, which can protect users' privacy against single-point attack. And then, we propose a location anonymous optimization method based on Stackelberg game to improve the algorithm. Specifically, we formalize the mutual optimization of user-adversary objectives by using the framework of Stackelberg game to find an optimal dummy location set. The optimal dummy location set can resist single-point attacks and inference attacks while effectively balancing service quality and location privacy. Finally, we provide exhaustive simulation evaluation for the proposed scheme compared with existing schemes in multiple aspects, and the results show that the proposed scheme can effectively resist the single-point attack and inference attack while balancing the service quality and location privacy.

1. Introduction

With the rapid development of mobile devices and social networks, location-based service (LBS) has become a vital part of our daily activities in recent years. With smartphones or tablets, users can download location-based applications from Apple Store or Google Play Store. With the help of these applications, users can easily send queries to a service provider and obtain LBSs related to some points of interest. For example, users can check the bus schedule, the price information of nearby restaurants or gas stations, etc. Undoubtedly, by submitting LBS queries, users can enjoy the convenience provided by LBS. However, since the untrusted service provider has all the information about users such as where they are at which time, what kind of queries they submit, what they are doing, etc., he may track users in various ways or release their personal data to third parties. Thus, we need to take appropriate measures to protect users' privacy.

Many approaches [1–3] have been taken to address such privacy problems, where the location *k*–anonymity and location perturbation are commonly used. The existing location *k*–anonymity technology usually adopts the structure based

on a trusted third party (TTP) [3]. The TTP structure refers to introduce a trusted third party, called centralized location anonymizer, between the user and the service provider, and the usage of the location anonymizer to make the target user's information indistinguishable from that of at least $k-1$ other users, so that the probability of location leakage is therefore at most $1/k$. Specifically, to achieve $k-$ anonymity, an LBS-related query is submitted to the service provider via a centralized location anonymizer, which enlarges the queried location into a bigger cloaking region covering many other users (e.g., $k-1$) geographically. As a result, it is hard for the untrusted service provider to distinguish the user's real location from this area. However, these approaches of using $k-$ anonymity have a fatal problem. It heavily relies on the location anonymizer, which suffers from a single point of failure [4]. If the adversary gains control of it, the privacy of all users will be compromised.

In response to the problems existed in the TTP structure, some researchers have proposed a dummy location technology that can also achieve $k-$anonymity, which uses an independent system structure [2]. The independent structure only contains the user and the service provider, where the user uses the mobile terminal to generate $k-1$ dummy locations and then sends $k-1$ dummy locations combined with the user's real location to the service provider. As a result, it is hard for the untrusted service provider to distinguish the user's real location from the other dummy locations. Since the structure achieves functions such as location anonymity and filtering query results through a mobile terminal instead of a location anonymizer, there is no single point of failure caused by the location anonymizer. In 2008, with the birth of Bitcoin [5], blockchain technology has been widely used in finance, medical care, supply chain, and other fields. Blockchain [6–8], as the underlying technology of Bitcoin, realizes distributed information interaction and collective maintenance of data in a decentralized and autonomous way, with decentralization, tamper-proof, autonomy, traceability, etc. Simultaneously, the security of consensus protocols [9] and the protection of user privacy [10] in the blockchain has become a new research hotspot. Chen et al. [11] proposed a dynamic multikey fully homomorphic encryption. The decentralized characteristic of blockchain opens a new door for location privacy protection. Based on this idea, [12] proposes a distributed $k-$anonymity location privacy protection scheme based on blockchain, which can also achieve $k-$anonymity without the help of a location anonymizer.

In the LBS, the user firstly adopts approaches that are based on perturbing the information reported to the service provider, so to prevent the disclosure of one's location. Clearly, the perturbation of the information sent to the service provider leads to a degradation of service quality, and consequently, there is a trade-off between the level of privacy that the user wishes to guarantee and the service quality loss that she will have to accept; however, the adversary formulates corresponding strategies based on the privacy protection method adopted by the user and infers the real location of the user by observing the perturbation of the information. Since the relationship between users and adversaries objectively conforms to the game relationship between participants in the Stackelberg game model, introducing the Stackelberg game method into the field of location privacy protection is an important research direction. Shokri et al. [13] took the lead in introducing the Stackelberg game into location privacy protection and proposed a location privacy protection scheme based on the Stackelberg game (SG). The solution assumes that the adversary has acquired prior knowledge and allows the user and the adversary to play the game in turn: The level of privacy protection is maximized by user when the service quality loss is less than a given threshold, whereas the adversary strives to minimize the level of privacy protection based on prior knowledge and offset location. By playing the game above, this strategy can finally optimize the level of privacy protection while ensuring that the service quality loss is less than a given threshold.

Based on the analysis above, the existing location privacy protection schemes have the following shortcomings: (1) the existing location privacy protection schemes either adopt TTP structure that has a single point of failure or the independent system structure. However, users in the independent structure use mobile terminals to perform location anonymity algorithms and filter query results, which will greatly increase the client's pressure, affecting the service quality in turn. (2) On the one hand, these schemes do not fully consider the semantics, physical dispersion, and query probability of the location when selecting dummy locations. On the other hand, they do not fully consider the background knowledge that the adversary may have, which adversaries can use to infer the users' location privacy information. So they cannot effectively resist single-point attacks and inference attacks. (3) Since such schemes need to sacrifice the service quality for improving the privacy protection level, there is no trade-off between service quality and privacy protection level. Aiming at related shortcomings above, this paper comprehensively considers features such as side information, location semantics, physical dispersion of locations combined with dummy locations, $k-$anonymity technology, Stackelberg game and other ideas, and then designs a $k-$anonymous location privacy protection scheme (KLPPS) based on Stackelberg game and dummy locations, which can resist single-point attacks and inference attacks while effectively balancing service quality and location privacy. Our contributions are mainly as follows:

(1) A semitrusted third party (STTP) based location privacy protection structure is proposed. The STTP is based on the TTP structure by adding an encryption server and WiFi-AP, and stores the user's privacy information on three party servers through a certain mechanism. In the STTP, even if the adversary steals the information on the location anonymizer, he still cannot locate the user and obtain the user's complete privacy information, which effectively solves the single point of failure existed in the TTP structure. Meanwhile, the location anonymizer is responsible for implementing location anonymity algorithms and filtering query results, etc., therefore, also solve the mobile device performance bottlenecks that exist in the independent structure.

(2) We propose a dummy location selection algorithm based on location semantics and physical distance (SPDDS). Compared with existing dummy location selection algorithms, SPDDS takes into account the characteristics such as location semantic diversity, physical dispersion, query probability and offset location when selecting dummy locations, which can effectively protect users' location privacy against single-point attack. Furthermore, we propose a location anonymous optimization method based on Stackelberg game, which introduces Stackelberg game to improve the dummy location selection algorithm. More specifically, we formalize the mutual optimization of user-adversary objectives (location privacy vs. correctness of inferring location) by using the framework of Stackelberg games, and find an optimal dummy location set by solving the game equilibrium. The optimal dummy location set can resist single-point attacks and inference attacks while effectively balancing service quality and location privacy.

(3) We conduct a comprehensive experiment to evaluate the proposed scheme. Experimental results show that our scheme can effectively resist single-point attacks and inference attacks while effectively balancing service quality and location privacye when compared with other dummy-based schemes.

The rest of the paper is organized as follows. We discuss the related work in Section 2. Section 3 presents some preliminaries of this paper. Section 4 presents the structure of STTP and the interactive process. We present the SPDDS algorithms and a location anonymous optimization method based on Stackelberg game in Section 5. Section 6 presents the experimental process as well as results. We conclude the paper in Section 7.

2. Related Works

In this section, we first analyze the merits and drawbacks of mainstream existing location privacy protection system structure. Furthermore, we review major existing techniques for preventing location privacy leakage including privacy protection scheme based on dummy and privacy protection scheme based on Stackelberg game in Sections 2.2 to 2.3, respectively.

2.1. Location Privacy Protection System Structure. As a large body of location privacy protection technologies has been proposed, the system structure on which various privacy protection technologies depend has shown distinct category differences. As the carrier of privacy protection technology implementation, the system structure has got sufficiently researched and developed.

Currently, there are two mainstream system structures: TTP–based central server structure and independent system structure. In the TTP structure [14–16], the location anonymizer obtains the location information of all users and is responsible for implementing the location privacy protection mechanism. It is currently a more commonly used privacy protection system structure, the advantage of which is that the location anonymizer can obtain the location information of all users and assist users in filtering service data. The disadvantage is that it relies on a location anonymizer to enlarge the queried location into a bigger cloaking region, and hence the location anonymizer becomes the central point of failure. References [2, 17, 18] proposed an independent system structure, where users can protect their location privacy according to their own capabilities and knowledge. The architecture treats each user as an independent individual, allows the user's device to implement a location privacy protection mechanism, directly sends a service request to the service provider, and receives the query result. The advantage of this system structure is that the deployment is conveniently simple, and it is convenient for users to adjust the privacy protection granularity according to their privacy protection needs. However, the implementation of privacy protection algorithms has been limited by the performance of mobile devices. Meanwhile, filtering query results will also increase the burden on the mobile client, which in turn affects service quality.

2.2. Privacy Protection Scheme Based on Dummy. Location dummies are aimed to secure users' accurate location by sending $k - 1$ false locations ("dummies") together with the true location so that the probability of location leakage is reduced to $1/k$. Compared to the traditional k–anonymity, this approach sends exact locations instead of cloaked regions to a service provider, which can return a more precise query result and avoid single-point failure.

Kido et al. [19, 20] first proposed to use dummy locations to achieve anonymity without employing a central server. However, they only concentrate on reducing the communication costs. Moreover, they employ a random walk model to generate dummy locations, and it cannot resist side information attacks due to lack of considering factors such as query probability. Subsequently, although Dapeng et al. [21] proposed the ABR algorithm based on query probability, which can resist side information attacks. However, it cannot resist homogeneity attacks and location similarity attacks for not considering the physical dispersion and location semantic diversity. The UPHIF algorithm proposed by Chang et al. [22] protected location privacy to a certain extent, but did not consider the location semantic diversity, so it cannot deal with location similarity attacks. Niu et al. [23] selected dummy locations based on entropy metrics, and proposed a dummy location selection (DLS) scheme and its improved scheme (enhanced – DLS). Although the enhanced – DLS scheme can resist side information attacks and homogeneity attacks, which cannot resist location similarity attacks for lacking of considering the location semantic diversity. References [24, 25] all considered the user's location semantic diversity, which can effectively resist location similarity attacks, but all have the problem that the cloaked region is too big, affecting the service quality in turn. Although [26, 27] fully considered the location semantic diversity and physical dispersion, which can effectively resist

homogeneity attacks and location similarity attacks, but they cannot resist the side information attack for not considering the query probability.

2.3. Privacy Protection Scheme Based on Stackelberg Game. In a big data environment, an adversary can use the various data collected to infer the privacy information of the user's location [28], which is called the location inference attack. Because the traditional dummy-based privacy protection scheme cannot effectively resist this kind of inference attack, the location privacy protection mechanism based on probabilistic reasoning [29, 30] has gradually attracted the attention of scholars. Such methods are based on perturbing the real locations of a user to the service provider, in order to increase the uncertainty of the adversary about a user's true whereabouts. However, the perturbation of the information sent to the service provider leads to a degradation of service quality, and consequently, there is a trade-off between the level of privacy that the user wishes to guarantee and the service quality loss that she will have to accept. So, the Stackelberg game has become an important means of balancing the level of privacy protection and service-quality requirements in such methods.

Based on [13] and combined the ideas of k–anonymity and dummy location, Xingyou et al. [31] propose HCLS and its improved scheme (HCLS – SG). Although the HCLS – SG scheme can effectively resist inference attacks and better balance service quality and location privacy, it cannot resist location similarity attacks for not considering location semantic diversity. Bordenabe et al. [32] also introduced differential privacy on the basis of [13] and constructed a privacy protection mechanism that optimizes the quality of service. Since differential privacy does not depend on prior, this mechanism can minimize the service quality loss under the premise of satisfying location indistinguishability. Shokri [18] further proposed using two indicators of differential privacy and distortion privacy to optimize the privacy protection strategy based on the Stackelberg game. Differential privacy limits the extent of user privacy leakage, while distortion privacy measures the error rate of inferring a user's privacy. By combining these two standards, this privacy protection strategy can resist more kinds of inference attacks while ensuring privacy protection requirements.

3. Preliminaries

In this section, we first introduce some relative definitions of location privacy protection algorithm; meanwhile, we summarize the notations introduced throughout the section in Table 1 and then introduce the relative concepts of Stackelberg game.

3.1. Relative Definitions of Location Privacy Protection Algorithm

Definition 1. According to [27], location semantic tree (LST), a true structure used to represent the semantic relations between two locations within the range of a Wi-Fi access points (Wi-Fi AP), which satisfies the following requirements:

(a) Each nonleaf node stands for the category of its children nodes and each leaf node for a real location l

(b) The depth of LST, denoted as h, is equal to the maximum number of layers of categories plus 1

(c) The semantic distance $d_{\text{sem}}(l_i, l_j)$ between two locations $l_i, l_j\,(i \neq j)$ is the number of hops from leaf node n_i to leaf node n_j

Definition 2. User's privacy requirements S, represented by two-tuple (k, u) that has the following meanings:

(a) k denotes the anonymous degree of our location privacy preservation model. More specifically, each query is sent with at least $k-1$ dummy locations and its offset location (we use offset location instead of the real location), making that the probability of offset location leakage is therefore $1/k$.

(b) u represents the minimum acceptable value of semantic distance between two locations in dummy location set (DLS). In other words, it satisfies the inequality:

$$\min\left[d_{\text{sem}}\left(l_i, l_j\right)\right] \geq u. \tag{1}$$

Definition 3 (location map distance). If we let Map_{cur} represent the map information within the range of the current Wi-Fi AP. For any two locations $l_i, l_j\,(i \neq j)$, the location map distance is the physical distance between the two locations on Map_{cur}, the value of which ranges from tens of meters to hundreds.

Definition 4 (location query probability (LQP)). As shown in Figure 1, in a map divided into $m \times m$ cells with equal size. Each cell has a query probability based on the previous query history, which is denoted as

$$p_i = \frac{\text{number of queries in cells}}{\text{number of queries in whole map}}, \tag{2}$$

where $i = 1, 2, \ldots, m^2$, $\sum_{i=1}^{m^2} p_i = 1$. The depth of the color in the figure indicates LQP (the darker the color, the greater the LQP), and the white area indicates that the location has never had a location service request, so these locations may be rivers, barren mountains, and other places that are easily filtered by the adversary.

Definition 5. The probability of exposing real location (PERL), which has been used to measure the effectiveness of the algorithm against side information attacks, is computed by

$$\text{PERL} = \frac{1}{k - k'}, \tag{3}$$

where k denotes the anonymous degree and k' represents the number of dummy locations filtered by the adversary

Table 1: Summary of notations.

Symbol	Meaning
$d_{\text{sem}}(l_i, l_j)$	The semantic distance between two locations $l_i, l_j\,(i \neq j)$
$d_{\text{phy}}(l_i, l_j)$	The physical distance between two locations $l_i, l_j\,(i \neq j)$
$d_{\text{que}}(l_i, l_j)$	The query probability distance between two locations $l_i, l_j\,(i \neq j)$, which is obtained by calculating the difference between the query probabilities of two locations
θ	Representation of the semantic diversity between locations
$\varphi(l)$	Location access profile of the user (probability of being at location when accessing the LBS)
l	Actual location l of the user
l_d	Offset location output by the $f(l_d \mid l)$
DLS	Set of possible dummy locations output by the LPPM
$f(\text{DLS} \mid l)$	The location privacy protection mechanism (LPPM): probability of replacing l with DLS
$\widehat{l}$	Adversary's estimate of the user's actual location
$g(\widehat{l} \mid \text{DLS})$	Adversary's attack function: probability of estimating $\widehat{l}$ as user's actual location, if DLS is observed
$f(l_d \mid l)$	Function of generating offset location: probability of replacing l with l_d
Qos_{loss}	Expected quality loss of an LPPM with location obfuscation function $f(l_d \mid l)$
$Q_{\text{loss}}^{\max}$	Maximum tolerable service quality loss
PL	Expected location privacy of the user with profile $\varphi(l)$ using LPPM $f(\text{DLS} \mid l)$ against attack $g(\widehat{l} \mid \text{DLS})$

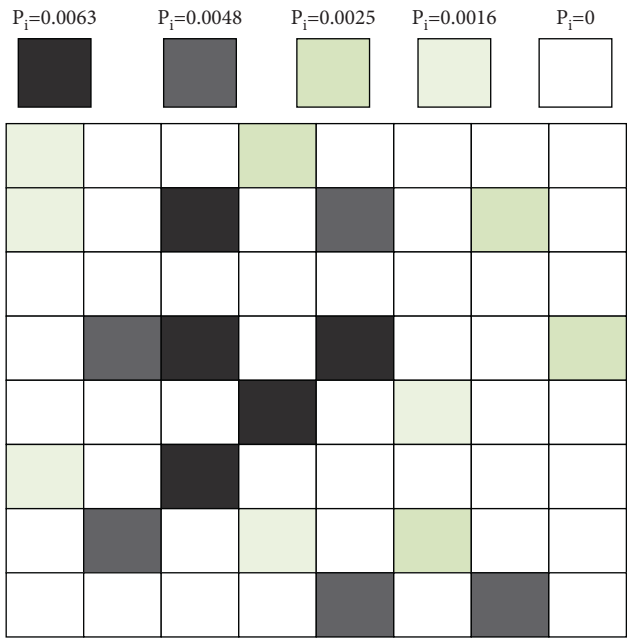

Figure 1: Location query probability diagram.

through the side information attack. The larger the PERL, the less effective the algorithm resists side information attacks; the smaller the PERL, the better the algorithm resists side information attacks.

Definition 6. Location physical dispersion (PD), which has been used to measure the effectiveness of the algorithm against location homogeneity attacks, is obtained by computing the minimum physical distance between any two locations in a DLS. The specific process is shown in (4):

$$\text{PD} = \text{Min}\left[d_{\text{phy}}\left(l_i, l_j\right)\right], \tag{4}$$

where $i, j = 1, 2, \ldots, k, i \neq j$. The greater the minimum distance between any two locations in the DLS, the greater the PD and the greater the coverage of the DLS, the better the algorithm's resistance to location homogeneity attacks.

Definition 7. θ−secure set of dummy locations. Dummy location set (DLS) consisting of $k-1$ dummy locations and the offset location, where the semantic distance between l_i and l_j satisfies

$$1 - \frac{|\text{SEM}|}{C_k^2} \geq \theta, \tag{5}$$

where $\text{SEM} = \left\{d_{\text{sem}} \mid d_{\text{sem}}(l_i, l_j) < u\right\}$, $k = |\text{DLS}|$, and C_k^2 is a combination formulas, we call DLS a θ− secure set. We use θ as a privacy protection index of location semantics in our experimental analysis in Section 6. Our aim is to achieve the maximum θ, i.e., to make it equal to 1, such that two locations in DLS belong to different categories.

Definition 8. The adversary uses background knowledge to run an inference attack on DLS in order to output estimation $\widehat{l}$ of the user's actual locations and the attack result can be denoted as $g(\widehat{l} \mid \text{DLS})$, and we define the location privacy protection mechanism (LPPM) that the adversary knows as $f(\text{DLS} \mid l)$. Then, we follow the definition in [33] and quantify the user's privacy level (PL) as the adversary's expected error in his inference attack, i.e., the expected distortion in the reconstructed event. We compute the expectation over all l, $\widehat{l}$, and DLS:

$$\text{PL} = \sum_{l, \widehat{l}, \text{DLS}} \varphi(l) f(\text{DLS} \mid l) g(\widehat{l} \mid \text{DLS}) d_{\text{phy}}(l, \widehat{l}), \tag{6}$$

PL directly reflects the adversary's inference on the user's actual location. The larger the PL, the less accurate the adversary's inference, and the better the effect of the algorithm resist the inference attack.

Definition 9. We define the process of generating the offset location l_d as $f(l_d \mid l)$. Then, following the definition in [13], the LBS response quality depends on the offset location l_d output by $f(l_d \mid l)$ and not on the user's actual location l. The distortion introduced in the offset location determines the quality of service that the user experiences. The more similar the actual and the offset location are, the higher the service

quality is. The expected quality loss (Qos_{loss}) due to $f(l_d|l)$ is computed as an average of $d_{\text{phy}}(l, l_d)$ over all l and l_d:

$$Qos_{\text{loss}} = \sum_{l,l_d} \varphi(l) f(l_d|l) d_{\text{phy}}(l, l_d). \tag{7}$$

We set a service quality threshold $Q_{\text{loss}}^{\max}$, which represents the maximum service quality loss that the user can accept. The location privacy protection scheme designed in this paper needs to guarantee $Qos_{\text{loss}} \le Q_{\text{loss}}^{\max}$ because if the service quality loss exceeds the threshold, the location service request results obtained cannot satisfy the requirements of users.

3.2. Stackelberg Game. The classic Stackelberg game is a two-player game composed of a leader and a follower [34]. The leader first determines his strategy, and after observing the leader's strategy, the follower chooses the strategy that maximizes his utility to play the game. In the field of location privacy protection, the terms "protector" and "leader," "adversary" and "follower" can be used interchangeably. For simplicity of expression, the leader (protector), denoted as Θ, is often referred to as she, whereas the follower (adversary), represented by Ψ, is referred to as he.

Definition 10. In the field of location privacy protection, the strong Stackelberg equilibrium (SSE) is generally used as the solution of the Stackelberg game. The definition of SSE is described as follows.

A strategy combination is SSE, if and only if it satisfies the following conditions:

(a) The leader's strategy is the best response:

$$\sum_{\gamma\in\Gamma} \Omega_\Theta(x, q, \gamma) \ge \sum_{\gamma\in\Gamma} \Omega_\Theta(x', q, \gamma) \; \forall x, x' \in X, q \in Q, \tag{8}$$

where $\gamma \in \Gamma$ means a particular type of follower, $x, x' \in X$ represents the leaders' mixed strategy, and $q \in Q$ represents the followers' mixed strategy.

(b) The follower's strategy is the best response:

$$\sum_{\gamma\in\Gamma} \Omega_\Psi(x, q, \gamma) \ge \sum_{\gamma\in\Gamma} \Omega_\Psi(x, q', \gamma) \; \forall x \in X, q, q' \in Q, \tag{9}$$

where $x \in X$ represents the leaders' mixed strategy, $q, q' \in Q$ represents the followers' mixed strategy.

(c) If there are multiple best responses for the followers, the followers choose the most favorable strategy for the leader to break the deadlock:

$$\sum_{\gamma\in\Gamma} \Omega_\Psi(x, q^r, \gamma) \ge \sum_{\gamma\in\Gamma} \Omega_\Psi\left(x, q^{r'}, \gamma\right) \; \forall x \in X, q^r, q^{r'} \in Q^*(x), \tag{10}$$

where $Q^*(x)$ is the follower's best response strategy set under the leader's strategy is x.

4. System Model

We first give the definition of the single-point attack mode and inference attack mode in Sections 4.1 and 4.2 respectively, and then introduce the structure of STTP in Section 4.3. Finally we present the interactive process of our scheme in Section 4.4.

4.1. Single-Point Attack Model. From the time dimension, the adversary relies on the intercepted single location-service request to infer the user's private information, which is called the single-point attack model [35]. In the model, the main attack methods of adversaries include side information attacks, homogeneity attacks and location similarity attacks.

Side information refers to information used by adversaries to filter dummy locations and help reduce anonymity, including map information and location query probability. For example, for a randomly generated dummy location set, some locations may be in a river or no man's land, and adversaries can easily filter out these locations based on the map information. Assuming that the location anonymity requirement is k, when k' of the $k-1$ dummy locations are filtered by the adversary based on the side information, the k–anonymity requirement is not satisfied, resulting in a decrease in the level of privacy protection.

Homogeneity attack means that the adversary analyzes the distance between multiple locations in a DLS to infer a user's privacy. Specifically, if the distance between any two locations is very close such as in the same building, although the DLS satisfies k–anonymity, the user's location privacy cannot be well protected because the cloaking region is too small.

The location similarity attack means that the adversary analyzes the semantic information in the cloaking region to infer a user's privacy. More specifically, if the region contains only one kind of semantic information, such as a hospital or school, the adversary can infer the user's behavior.

4.2. Inference Attack Model. In a big data environment, an adversary can use the various data collected to infer the privacy information of the user's location [28], which is called the location inference attack.

In the location inference model, the adversary has certain background knowledge such as the user's service request history records, LPPM, etc. Using the user's service request history records, the adversary can calculate the user's query probability distribution $\varphi(l)$. When the user sends a query request again, if the location query probability distribution in the anonymous set is not uniform, the adversary can infer that the user is likely to be located in a location with a higher probability. While for the LPPM, the adversary can analyze the intercepted location request, combined with the anonymity algorithm, to infer the probability that each location in the anonymous set is the user's true location, so as to make the inference attack more accurate.

4.3. The Structure of STTP. It can be seen from Section 2.1 that, for the current two mainstream location privacy protection system structures, the TTP structure has the

problem of a single point of failure, while the independent structure has the problem of mobile device performance bottlenecks. In view of the problems above, we have designed a semi-trusted third party (STTP) based location privacy protection structure. STTP is based on the traditional TTP structure by adding an encryption server and Wi-Fi AP and stores the user's private information in the three-party server through a certain mechanism, which results in that even if the location anonymizer has been controlled by adversaries, STTP also protects the user's private information to a certain extent. Furthermore, the location anonymizer is responsible for implementing the privacy protection algorithms and filtering query results, so there are no problems such as mobile device performance bottlenecks. STTP is shown in Figure 2, which consists of the following five entities:

User: using a mobile terminal to initiate a location service request when needed.

Wi-Fi AP: providing network support, and calculating, storing LST and LQP.

Encryption server (ES): providing encryption and decryption key pairs corresponding to the user's pseudonym.

Location anonymizer (LA): converting the user's actual location into a dummy location set, and after the service provider returns the query result, extracting appropriate service information and returning it to the user.

Service provider (SP): return the corresponding service result according to the location query request.

The proposed scheme assumes that ES, LA, and SP are "honest and curious." On the one hand, they will not disrupt the protocol process and can faithfully complete their work following the agreement; on the other hand, they all want to analyze more other sensitive information about the user from what they have mastered. Meanwhile, we further set that ES, LA, and SP cannot collude with each other, that is, they will not be controlled by an adversary simultaneously. There will be no secrets for the user if the three parties conspire, so this setting is reasonable.

4.4. Interactive Process. There are eight steps in the interactive process of the proposed scheme. The specific implementation of each step is described below (as shown in Figure 2):

(1) Before initiating a location service request, the user first requests Map_{cur}, LST, and LQP from the Wi-Fi AP.

(2) The Wi-Fi AP generates Map_{cur}, computes and stores LQP of all locations within its current coverage area, generates and saves LST by collecting location semantic information within its radio range, and then sends Map_{cur}, LST, and LQP to the user. It should be noted that for any Wi-Fi AP, the location within its coverage area is relatively stable, so LST and LQP do not need to change frequently.

(3) The user then requests the pseudonym U_{pseu} and key pair from ES. Specifically, if there are multiple service requests at the same location within the limited time t_{session}, the user only applies for the pseudonym and key pair once; when the time exceeds t_{session} or the user's real location changes, she will reapply for a new pseudonym and key, so as to achieve the effect of confusing her identity.

(4) ES generates the corresponding pseudonym U_{pseu} and RSA key pair $(\text{Key}_{\text{public}}, \text{Key}_{\text{privacy}})$, returns U_{pseu} and $\text{Key}_{\text{public}}$ to the user, and sends U_{pseu} and $\text{Key}_{\text{privacy}}$ to the SP. It should be noted that, as an example, the solution uses the classic RSA algorithm for encryption, and it can be replaced by other encryption algorithms according to actual requirements. In addition, the solution requires ES to only act as a provider of pseudonyms and keys, so ES does not store related pseudonyms and keys locally.

(5) The user first encrypts his query content Query with the public key $\text{Key}_{\text{public}}$ and then sends his current pseudonym U_{pseu}, encrypted query content Query$'$, current real location l, Map_{cur}, LQP, and LST to LA together.

(6) After receiving the information, LA performs the corresponding location anonymity algorithm that generates a dummy location set DLS to hide l and then sends U_{pseu}, Query$'$ and DLS to the SP.

(7) After receiving the location service request, the SP first searches for the corresponding private key $\text{Key}_{\text{privacy}}$ according to U_{pseu}, which is used to decrypt Query$'$, and then provides the service result Result(Query|DLS) according to Query and DLS, finally return it to LA.

(8) After receiving Result(Query|DLS), the LA first identifies the corresponding location l_d according to the U_{pseu}, and then filters out the query result Result(l_d) from Result(Query|DLS) and finally returns it to the user.

5. Proposed Scheme

In this section, we first introduce a dummy location selection algorithm based on location semantics and physical distance (SPDDS) and then present a location anonymous optimization method based on Stackelberg game.

5.1. A Dummy Location Selection Algorithm Based on Location Semantics and Physical Distance. Based on the analysis above, the final dummy location set not only needs to avoid selecting places that are easy to be filtered by adversaries, such as rivers and no man's land but also to meet the semantic diversity while making the locations as dispersed as possible. In other words, the final dummy location set needs to simultaneously satisfy (11)–(13)

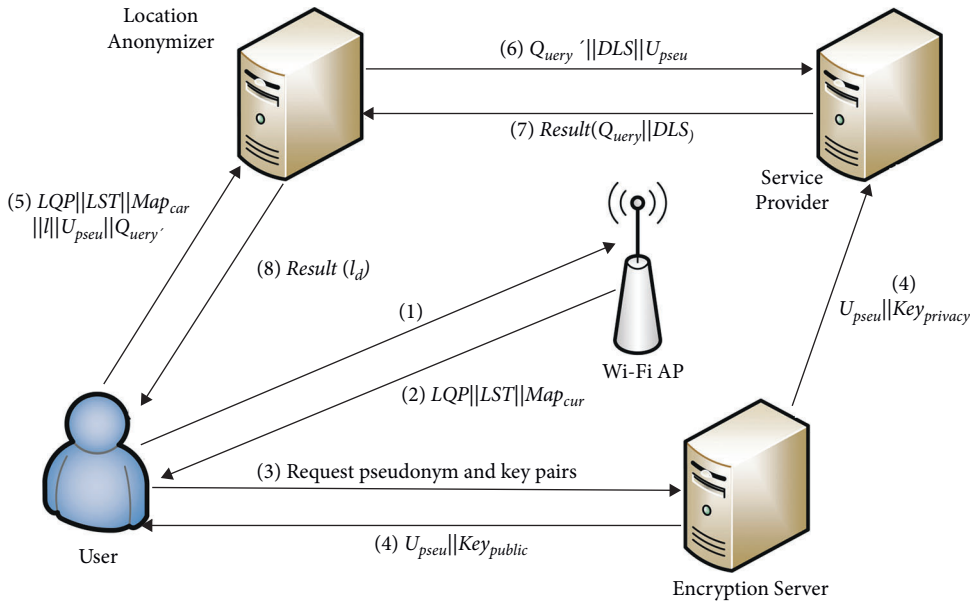

Figure 2: A semitrusted third party based location privacy protection structure.

$$\mathrm{DLS} = \arg\min\left\{\max\left[d_{\mathrm{que}}\left(l_i, l_j\right)\right]\right\}, \tag{11}$$

$$\mathrm{DLS} = \arg\max\left\{\min\left[d_{\mathrm{sem}}\left(l_i, l_j\right)\right]\right\}, \tag{12}$$

$$\mathrm{DLS} = \arg\max\left\{\min\left[d_{\mathrm{phy}}\left(l_i, l_j\right)\right]\right\}, \tag{13}$$

where $l_i, l_j \in \mathrm{DLS}, i \neq j, P_r(l_i) > 0, P_r(l_d) > 0$. It can be formulated as a multiobjective optimization problem (MOP) since three factors are considered simultaneously. However, we put forward a simpler objective formulas considering the complexity of MOP. In each dummy location set, we would like to make sure that (14) can be satisfied. Consequently, we propose a dummy location selection algorithm based on location semantics and physical distance (SPDDS).

$$\mathrm{DLS} = \arg\max\left\{\frac{\min\left[d_{\mathrm{sem}}\left(l_i, l_j\right) + r \cdot d_{\mathrm{phy}}\left(l_i, l_j\right)\right]}{\max\left[d_{\mathrm{que}}\left(l_i, l_j\right) + 1\right]}\right\}, \tag{14}$$

where $l_i, l_j \in \mathrm{DLS}, i \neq j$. $d_{\mathrm{que}}(l_i, l_j) + 1$ is to avoid the situation where the two locations have the same probability, that is, the difference between the query probability of the two locations is 0. Here, r is a controllable factor for balancing the share of semantic distance, physical distance, and query probability distance since $d_{\mathrm{sem}}(l_i, l_j) \leq 2 \cdot (h - 1)$, where h is the depth of LST and hence is usually less than 10 while $d_{\mathrm{phy}}(l_i, l_j)$, as Wi-Fi transmission distance, ranges from hundreds of meters to thousands, whereas the query probability distance is always less than 1. Consequently, we set $r = 0.03$.

Meanwhile, in order to balance service quality while the proposed algorithm can resist inference attacks effectively, we should take into account both PL and Qos_{loss}. So, we propose a location anonymous optimization method based on Stackelberg game, which introduces Stackelberg game to optimize the dummy location selection algorithm. More specifically, we formalize the mutual optimization of user-adversary objectives (location privacy vs. correctness of inferring location) by using the framework of Stackelberg games, to find an optimal dummy location set. The optimal dummy location set can resist single-point attacks and inference attacks while effectively balancing service quality and location privacy.

The main idea of SPDDS is to first select an offset location to replace the user's real location; secondly, selecting all locations that satisfy the semantic difference with the existing locations in the current dummy location set as the dummy location candidate set (DLCS); then, selecting an optimal location in the DLCS which refers to the location formed by satisfying (14); finally, a set consisting of an offset location and $k - 1$ dummy locations is generated. Algorithm 1 shows the formal description of the SPDDS algorithm.

5.2. A Location Anonymous Optimization Method Based on Stackelberg Game. We propose a location anonymous optimization method based on Stackelberg game, which optimizes the dummy location selection algorithm by introducing the Stackelberg game. More specifically, we formalize the mutual optimization of user-adversary objectives (location privacy vs. correctness of inferring location) by using the framework of Stackelberg game, and based on which we construct the related linear programs. We can find an optimal dummy location set by solving the linear programs, which can resist single-point attacks and inference attacks while effectively balancing service quality and location privacy.

5.2.1. Location Inference Model. In the location inference model, the adversary has certain background knowledge such as the user's service request history records, LPPM, etc. Using the user's service request history records, the adversary can calculate the user's query probability distribution $\varphi(l)$. When the user sends a query request again, if the

Input: l: user's real location
S: user's privacy requirement
Map_{cur}: map information in current Wi-Fi AP LST: location semantic tree LQP: location query probability;
Output: DLS: dummy location set;
(1) Divide the Map_{cur} as the sample space into $m \times m$ grids;
(2) Generate semantic distance matric ($S\ DM$) according to the LST and geographic distance matrix (GDM) according to the Map_{cur} and probability distance matric (PDM) according to the LQP respectively
(3) According to GDM and PDM, choose the $n-1$ locations with $p_i > 0$, which are closest to l, and then a offset location set (M_{set}) consisting of the $n-1$ locations and l has been generated;
(4) Randomly choose a location from M_{set} as offset location l_d
(5) Generate a dummy location candidate set (DLCS) for all $p_i > 0$ locations in the whole grid space;
(6) DLS = $\{l_d\}$
(7) Remove l_d from DLCS;
(8) **while** |DLS| < k **do**
(9) **if** DLCS = ϕ **then**
(10) anonymity failed;
(11) **else**
(12) max = 0; BestLoc = ϕ;
(13) **for** *each* Loc *in* DLCS **do**
(14) **if** $d_{\text{sem}}(\text{Loc}, \text{DLS.last}) \le u$ **then**
(15) anonymity failed;
(16) go back to line 13;
(17) **else**
(18) $m = d_{\text{sem}}(\text{Loc}, \text{DLS.last}) + r \cdot [d_{\text{phy}}(\text{Loc}, \text{DLS.last})]/d_{\text{que}}(\text{Loc}, \text{DLS.last})$
(19) compute the maximum m according to SDM, GDM and PDM, which is recorded with max, and then assign the corresponding Loc to BestLoc
(20) **end**
(21) **end**
(22) DLS = DLS $\cup$ {BestLoc}
(23) remove BestLoc from DLCS;
(24) **end**
(25) **end**
(26) output DLS;

Algorithm 1: A dummy location selection algorithm based on location semantic and physical distance.

location query probability distribution in the anonymous set is not uniform, the adversary can infer that the user is likely to be located in a location with a higher probability. While for the LPPM, the adversary can analyze the intercepted location request, combined with the anonymity algorithm, to infer the probability that each location in the anonymous set is the user's true location, so as to make the inference attack more accurate.

Based on the existing knowledge ($\varphi(l)$, $f(\text{DLS}|l)$, etc.), the adversary can form the posterior distribution on the true location l of the user, conditional on the anonymous set DLS:

$$P_r(l|\text{DLS}) = \frac{P_r(l, \text{DLS})}{P_r(\text{DLS})} = \frac{\varphi(l) f(\text{DLS}|l)}{\sum_l \varphi(l) f(\text{DLS}|l)}. \tag{15}$$

The adversary's objective is then to choose $\widehat{l}$ to minimize the user's conditional expected privacy, where the expectation is taken under $P_r(l|\text{DLS})$. The user's conditional expected privacy for an arbitrary $\widehat{l}$ is

$$\sum_l P_r(l|\text{DLS}) d_{\text{phy}}(l, \widehat{l}), \tag{16}$$

and for the minimizing $\widehat{l}$, it is

$$\min_{\widehat{l}} \sum_l P_r(l|\text{DLS}) d_{\text{phy}}(l, \widehat{l}). \tag{17}$$

If there are multiple values of $\widehat{l}$ that satisfy (17), then the adversary randomizes arbitrarily among them. The probability with which $\widehat{l}$ is chosen in this randomization is $g(\widehat{l}|\text{DLS})$.

5.2.2. Stackelberg Game Optimization Process. Here, we assume that the adversary has some background knowledge. Specifically, he will infer the user's actual location l as much as possible according to $\varphi(l)$, the LPPM used by the LA, the anonymous result DLS, and other background knowledge. Relatively, we can assume that all the background knowledge that LA knows will be used by the adversary, so LA can use the adversary's optimal attack strategy as a parameter to optimize the generation process of the dummy location set DLS.

We formalize the process above by using the framework of Stackelberg game. In a Stackelberg game the leader, in our case, the LA plays first by giving the dummy location set DLS according to the relative location privacy protection

algorithm $f(\mathrm{DLS}|l)$. The follower, in our case the adversary, plays next by estimating the user's true location, knowing some background knowledge.

We use the distance between the adversary's inferred location $\hat{l}$ and the user's actual location l to measure the utility of the participants in the game: the greater the distance, the greater the LA returns, indicating that the anonymous algorithm is more effective in resisting inference attacks; on the contrary, the smaller the distance, the greater the adversary returns, the more effective the adversary's attack strategy.

The game model is also an instance of a zero-sum game, as the adversary's gains (or losses) of utility is exactly balanced by the losses (or gains) of the utility of the user: the information gained (lost) by the adversary is the location privacy lost (gained) by the user.

The purpose of Stackelberg game optimization is to find SSE so that the adversary cannot obtain more benefits by optimizing the attack strategy (that is, the adversary cannot make more accurate inferences about the actual location of the user). In this paper, SPDDS optimized by Stackelberg game is denoted as SPDDS_SG.

It should be noted that DLS is the result obtained by l_d further anonymously, so $f(\mathrm{DLS}|l)$ can be further expressed by

$$f(\mathrm{DLS}|l) = f(l_d|l) \cdot f(\mathrm{DLS}|l_d). \tag{18}$$

In some cases, $f(\mathrm{DLS}|l)$ and $f(l_d|l)$ are equal, the reason is that the adversary can filter out dummy locations except l_d in such cases.

We see that, for a given DLS, the user's conditional expected privacy is given by (17). The probability that DLS is output by the LPPM is $P_r(\mathrm{DLS}) = \sum_l \varphi(l) f(\mathrm{DLS}|l)$. Hence, the user's unconditional expected privacy (averaged over all DLS) is

$$\sum_{\mathrm{DLS}} P_r(\mathrm{DLS}) \min_{\hat{l}} \sum_l P_r(l|\mathrm{DLS}) d_{\mathrm{phy}}(l,\hat{l}) = \sum_{\mathrm{DLS}} \min_{\hat{l}} \sum_l \varphi(l) f \cdot (\mathrm{DLS}|l) d_{\mathrm{phy}}(l,\hat{l}). \tag{19}$$

To facilitate the computations, we define

$$x \triangleq \min_{\hat{l}} \sum_l \varphi(l) f(\mathrm{DLS}|l) d_{\mathrm{phy}}(l,\hat{l}). \tag{20}$$

Incorporating x into (19), we write the unconditional expected privacy of the user as

$$\sum_{\mathrm{DLS}} x, \tag{21}$$

which the user aims to maximize by choosing the optimal DLS. To facilitate the computations, (20) can be transformed to a series of linear constraints:

$$x \le \sum_l \varphi(l) f(\mathrm{DLS}|l) d_{\mathrm{phy}}(l,\hat{l}) = \sum_l \varphi(l) f(l_d|l) f(\mathrm{DLS}|l_d) \cdot d_{\mathrm{phy}}(l,\hat{l}), \forall \hat{l}. \tag{22}$$

In addition, SPDDS_SG needs to conceal the user's real location on the premise of ensuring the user's service quality. In order to ensure the quality of service, we set the service quality threshold $Q_{\mathrm{loss}}^{\max}$ to limit the maximum service quality loss. The specific process is

$$\sum_{l,l_d} \varphi(l) f(l_d|l) d_{\mathrm{phy}}(l,l_d) \le Q_{\mathrm{loss}}^{\max}. \tag{23}$$

In summary, SPDDS_SG can be solved by a linear program. The final definition of linear program is

$$\begin{aligned} &\text{Maximize} \sum_{\mathrm{DLS}} x, \\ &s.t. C_1: x \le \sum_l \varphi(l) f(l_d|l) f(\mathrm{DLS}|l_d) d_{\mathrm{phy}}(l,\hat{l}), \\ &C_2: \sum_{l,l_d} \varphi(l) f(l_d|l) d_{\mathrm{phy}}(l,l_d) \le Q_{\mathrm{loss}}^{\max}, \\ &C_3: \sum_{DL\,S} f(\mathrm{DLS}|l) = 1, \\ &C_4: f(\mathrm{DLS}|l) \ge 0, \forall l, \mathrm{DLS}, \end{aligned} \tag{24}$$

where C_1 is used to maximize the utility of the adversary; C_2 reflects the service quality constraint; C_3 indicates that the sum of the generation probability of the dummy location set must be 1; C_4 indicates the probability of each candidate dummy location set is greater than zero.

SPDDS_SG solves the objective function under the constraints in (24) and obtains the optimal dummy location set, which can resist single-point attacks and inference attacks while effectively balancing service quality and location privacy.

6. Simulations and Results

In this section, we use Python software to simulate the experiment. First, we give the relevant parameters of the experiment. Furthermore, we simulate the experimental results and analysis of the proposed scheme.

6.1. Simulation Setup. Our scheme is implemented in MATLAB and performed on a Windows 10 PC with an Intel Core i5-8500 CPU, a 3.00 GHz processor and a 8.00 GB main memory. We use a real road map of Guangzhou from Google Maps, since Guangzhou as a provincial capital in southern China is a big city with enough users in LBS and its central urban area has been covered by Wi-Fi APs in 2016. The coverage area of each Wi-Fi AP is about $700 \sim 800\,m$, the sample space is divided into 13×13 cells with equal size, and a total of 13 559 sample trajectories are used as historical data to calculate the historical query probability of each cell. Besides, all locations in our experiments are divided into 6 categories semantically as follows: Education and Science, Administration and Housing, Medical care, Shopping malls, Public places, Catering and Entertainment. The value ranges of the main parameters u and k of the experiment are $3 \le u \le 7$ and $2 \le k \le 30$, respectively.

6.2. Results and Analysis. We first evaluate the effectiveness of our proposed scheme in resisting single-point attacks from three assessment metrics as follows: (1) PERL. As is shown in

Definition 5, it reflects the effectiveness of the algorithm in resisting side information attack. (2) PD. As is shown in Definition 6, the larger the PD, the more dispersed the dummy locations in the DLS, and the better the effectiveness of the algorithm in resisting homogeneity attack. (3) θ. As is shown in Definition 7, it refers to the level of semantic diversity in the anonymous set, which reflects the effectiveness of the algorithm in resisting location similarity attack; Next, evaluating the effectiveness of the scheme in resisting inference attack while balancing location privacy and service quality from two assessment metrics as follows: (1) PL. As is shown in Definition 8, the larger the PL is, the better the effect of LPPM against inference attacks is. (2) Qos_{loss}. As is shown in Definition 9, it reflects the effectiveness of the algorithm in balancing location privacy and service quality.

6.2.1. Effectiveness of the Scheme against Single-point Attacks. *(1) PERL vs k.* In Figure 3(a), we compare the PERL of KLPPS, Max Min Dist DS [27], Simp Max Min Dist DS [27], enhanced − DLS [23], and HCLS − SG [31] schemes. As we can see, the PERL of the five schemes shows a downward trend with the increase of k, which means that the larger the k, the more difficult it is for adversaries to filter out invalid locations in the anonymous set through side information attacks, the better the effect of the scheme against side information attacks. The PERL of the KLPPS, enhanced − DLS, and HCLS − SG is lower than that of the Max Min Dist DS and Simp Max Min Dist DS. And that of the KLPPS, enhanced − DLS, and HCLS − SG are basically the same. The reason is that the KLPPS, enhanced − DLS and HCLS − SG all consider the query probability and avoid selecting locations with low access probability such as lakes and forests to form an anonymous set; whereas the Max Min Dist DS and Simp Max Min Dist DS do not consider the query probability, so there will be cases where invalid locations are selected, and thereby the adversary can filter out ones through side information attacks. In summary, the KLPPS scheme can effectively resist side information attacks.

(2) P D vs k. Figure 3(b) shows the PD comparison chart of KLPPS, Max Min Dist DS, Simp Max Min Dist DS, enhanced − DLS, and HCLS − SG schemes. As we can see, the PD of KLPPS, HCLS − SG, enhanced − DLS, and Max Min Dist DS are close when $k \le 4$; at $k \ge 5$, the PD of Max Min Dist DS is slightly larger than that of KLPPS, HCLS − SG and enhanced − DLS. Under the same value of k, the PD of KLPPS, HCLS − SG, and enhanced − DLS is slightly larger than that of Simp Max Min Dist DS. In additional, with the increase of k, the PD of the five schemes are both reduced gradually. The reason for this is obvious: it becomes harder to maintain a high level of dispersion with more and more dummies. In summary, Max Min Dist DS has the largest PD, KLPPS, HCLS − SG, enhanced − DLS, and Simp Max Min Dist DS decrease in order, which means that the Max Min Dist DS is better in resisting homogeneity attacks than the other four schemes, but the KLPPS scheme is also acceptable.

(3) θ vs k. Figure 3(c) shows the value of θ comparison between KLPPS, Max Min Dist DS, Simp Max Min Dist DS, enhanced − DLS, and HCLS − SG schemes. As shown in Figure 3(c), with the increases of k, the value of θ of KLPPS, Max Min Dist DS, and Simp Max Min Dist DS schemes hardly change and close to the maximum value 1. However, that of enhanced − DLS and HCLS − SG schemes is always at a relative low. The reason is that the KLPPS, Max Min Dist DS, and Simp Max Min Dist DS schemes all consider the semantic information of the location when selecting dummy locations, thereby ensuring semantic diversity, while the enhanced − DLS and HCLS − SG schemes only consider the query probability of each location point instead of considering the situation that each location point may have the same semantic information. Moreover, the location points with higher query probability are often in hotspot areas, between which the semantic information is very similar and therefore not satisfying the semantic diversity. Consequently, the enhanced − DLS and HCLS − SG schemes behave such badly in semantic diversity that they cannot resist location similarity attacks. In summary, the KLPPS scheme can effectively resist location similarity attacks.

The experimental results above show that the KLPPS scheme can effectively resist homogeneity attacks, location similarity attacks, and side information attacks simultaneously compared with the Max Min Dist DS, Simp Max Min Dist DS, enhanced − DLS, and HCLS − SG schemes, thereby effectively resisting single-point attacks.

6.2.2. Effectiveness of the Scheme against Inference Attacks and Balances PL and Qos_{loss}. Combining the location inference model and (6), it can be seen that the adversary can perform inference attacks, the purpose of which is to choose $\hat{l}$ based on existing knowledge to minimize the expected user privacy. (25) defines this attack strategy:

$$\hat{l} = \underset{\hat{l}}{\operatorname{argmin}}\, \text{PL}. \tag{25}$$

Combining (6) and (25), we can construct the following linear program to find the optimal $\hat{l}$:

$$\begin{aligned} &\text{minimize} \sum_{l,\hat{l},\text{DLS}} \varphi(l) f(\text{DLS}|l) g(\hat{l}|\text{DLS}) d_{\text{phy}}(l,\hat{l}), \\ &s.t.\ C_1: \sum_{\hat{l}} g(\hat{l}|\text{DLS}) = 1,\ \forall \text{DLS}, \\ &C_2: g(\hat{l}|\text{DLS}) \ge 0,\ \forall \text{DLS}, \hat{l}. \end{aligned} \tag{26}$$

We use the model defined by (26) to run inference attacks on KLPPS, HCLS [31], SG [13], and HCLS − SG to make a comparison from two aspects of PL and Qos_{loss}, evaluating the effectiveness of the KLPPS scheme.

(1) PL. The definition of PL is shown in (6), the larger the PL, the better the effect of LPPM against inference attacks. As shown in (24), the preset service quality loss threshold $Q_{\text{loss}}^{\max}$ and anonymous degree k have a greater impact on PL,

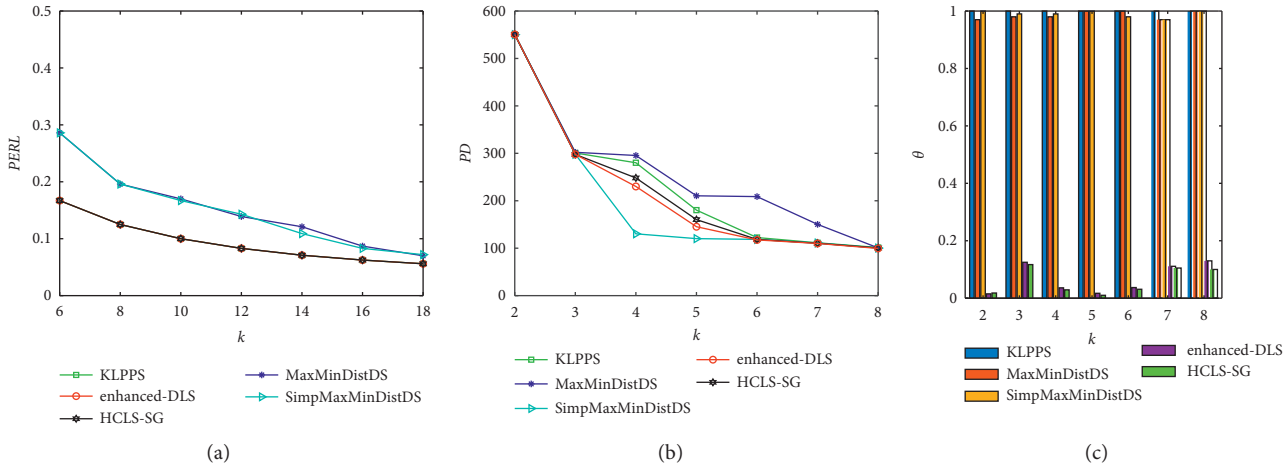

Figure 3: KLPPS scheme against single-point attack. (a) PERL vs. k. (b) PD vs. k. (c) θ vs. k.

so we evaluate the effectiveness of the KLPPS scheme against inference attacks from the two assessment metrics of $Q_{\text{loss}}^{\max}$ and k.

(a) PL vs $Q_{\text{loss}}^{\max}$

We compare the PL of KLPPS, HCLS, SG, and HCLS − SG schemes under different $Q_{\text{loss}}^{\max}$ in Figure 4(a). As shown in the figure, we can draw the following 3 conclusions. First, the PL of KLPPS, SG, and HCLS − SG is significantly better than HCLS, which results from that HCLS does not consider the inference attack strategy of the adversary. Secondly, the KLPPS and HCLS − SG behave better than SG, the reason is that the former two use dummy location anonymity, increasing the difficulty of the adversary's inference. Finally, with the increase of $Q_{\text{loss}}^{\max}$, the PL of the four schemes all increase, however the growth trend of PL of the four schemes slows down when reaching a certain level, which is related to the query probability distribution of user's location, indicating that the influence of $Q_{\text{loss}}^{\max}$ on PL is limited.

(b) PL vs k

We compare the PL of KLPPS, HCLS, SG, and HCLS − SG schemes under different k in Figure 4(b). As we can see, three conclusions have been drawn below. First of all, as the value of k increases, the PL of KLPPS, HCLS, and HCLS − SG are significantly improved while not the SG. The reason is that SG only provides offset locations and does not consider dummy location anonymity; Secondly, the KLPPS and HCLS − SG behave better in improving PL than that of HCLS, which results from that the former two all consider the adversary's inference attack strategy, while the latter does not, so HCLS is less effective in resisting inference attacks than that of the KLPPS and HCLS − SG; Finally, the growth trend of of the four schemes slows down when reaching a certain level, which results from that the four schemes all use the offset location instead of the real location to protect users' privacy. More specifically, the choice of l_d will make the service quality loss Qos_{loss} is gradually approaching $Q_{\text{loss}}^{\max}$ with the increase of k, and the four schemes are all maximized PL under the premise of ensuring that $Qos_{\text{loss}} \le Q_{\text{loss}}^{\max}$, which means that when Qos_{loss} gradually approaches $Q_{\text{loss}}^{\max}$, the growth trend of PL slows down until $Qos_{\text{loss}} = Q_{\text{loss}}^{\max}$, reaching the maximum value.

(2) Qos_{loss}. Qos_{loss} is closely related to PL. Specifically, in some cases, users allow losing a certain service quality in exchange for higher privacy. In the experiment, we set the maximum service quality loss that the user can accept as $Q_{\text{loss}}^{\max} = \{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5\}$ and analyze the relationship between Qos_{loss} and PL on this condition. Figure 5 shows the experimental results. First of all, it can be seen from the figure that the Qos_{loss} and PL of the three schemes all increase with the increase of $Q_{\text{loss}}^{\max}$. The reason for this is obvious: in the first place, we can see the PL will increase to a certain extent with the increase of $Q_{\text{loss}}^{\max}$ from Figure 4(a). In the second place, from the definition of $Q_{\text{loss}}^{\max}$ and Qos_{loss}, it is obvious that Qos_{loss} will increase to a certain extent with the increase of $Q_{\text{loss}}^{\max}$. Secondly, under the same $Q_{\text{loss}}^{\max}$, the PL of KLPPS and HCLS − SG is significantly higher than that of HCLS, but Qos_{loss} is also higher than HCLS to a certain extent, which results from that both KLPPS and HCLS − SG will make full use of the limited maximum service quality loss to optimize the selection of dummy location set, so as to improve location privacy as much as possible while ensuring that the loss of service quality does not exceed the constraints of service quality, thereby effectively balancing service quality and location privacy. In addition, under the same $Q_{\text{loss}}^{\max}$, the PL of KLPPS is slightly larger than that of HCLS − SG while Qos_{loss} is slightly smaller than that of HCLS − SG, indicating that KLPPS is better than HCLS − SG in balancing service quality and location privacy.

The experimental results above show that the scheme can effectively resist inference attacks while effectively balancing service quality and location privacy compared with the SG, HCLS, and HCLS − SG schemes.

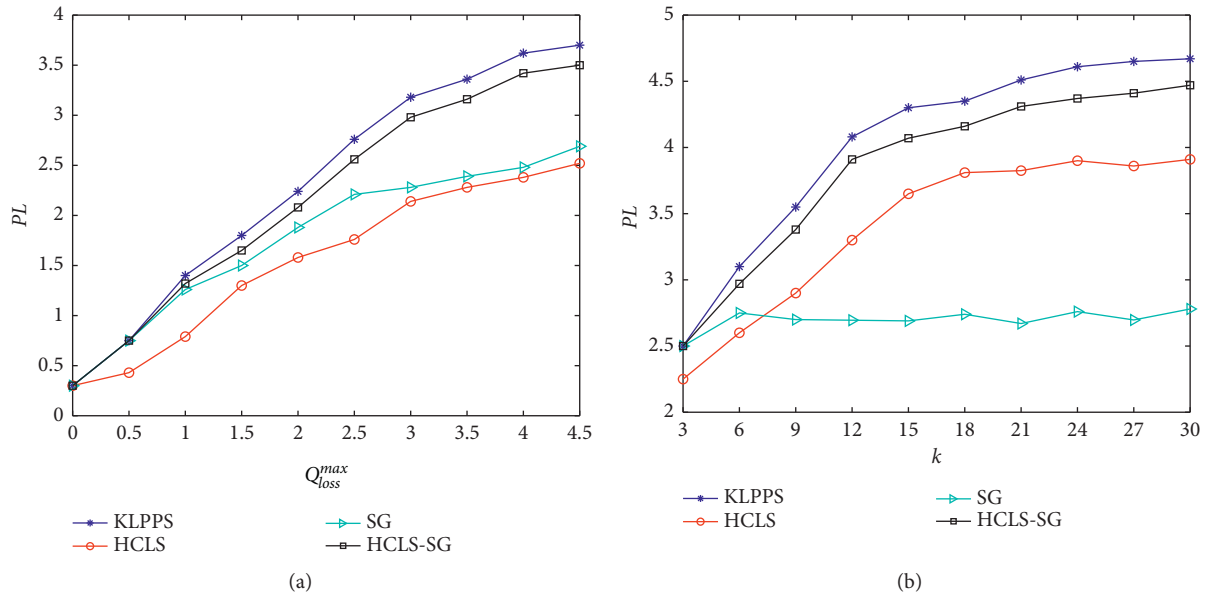

Figure 4: KLPPS scheme against inference attack. (a) PL vs Q_{loss}^{max}. (b) PL vs k.

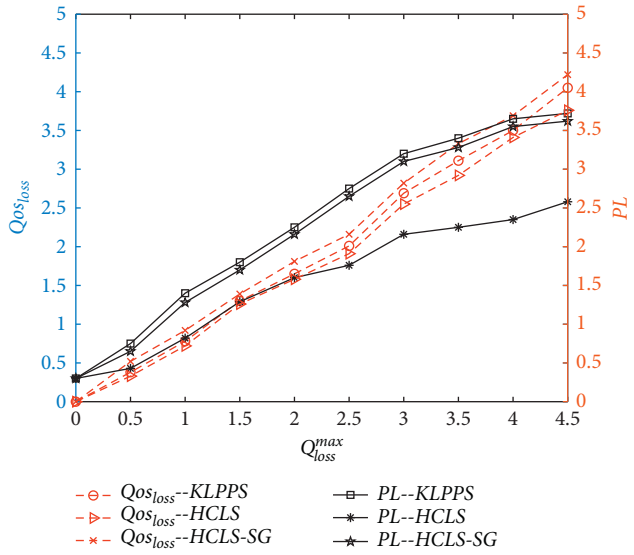

Figure 5: KLPPS scheme balances PL and Qos_{loss}.

7. Conclusion

There are some problems such as single point of failure and without the ability to effectively resist single-point attack and inference attack, etc., in the traditional k–anonymous location privacy protection schemes. To solve the problems above, by analyzing the merits and drawbacks of the existing location privacy protection system architecture, we propose a semitrusted third party based location privacy protection architecture that can tackle performance bottleneck of mobile device and single point of failure. Then, comprehensively considering the characteristics of side information, semantic diversity, and physical dispersion of locations combined with the ideas of dummy location technology and offset location, a dummy location selection algorithm based on location semantics and physical distance is proposed to effectively resist single-point attacks. Finally, we propose a location anonymous optimization method based on Stackelberg game to optimize the dummy selection algorithm. Specifically, we formalize the mutual optimization of user-adversary objectives (location privacy vs. correctness of inferring location) by using the framework of Stackelberg games, to find an optimal dummy location set. The optimal dummy location set can resist single-point attacks and inference attacks while effectively balancing service quality and location privacy. The experimental results further verify the effectiveness of the proposed scheme. However, our work still has the following shortcomings. First, in LBS, more and more people use continuous query services such as navigation services, etc., while our scheme can be only applied in snapshot query scenario not the continuous query scenario. Secondly, in different application scenarios, users have different requirements for privacy protection level and service quality, so it needs to be improved as much as possible in terms of balancing data validity and privacy levels. The next work hopes to improve our scheme to make it suitable for continuous query scenarios. Meanwhile, aiming at users with different needs in different scenarios, on the basis of further balancing service quality and privacy protection level, we design a personalized location privacy protection scheme.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Foundation of National Natural Science Foundation of China (grant number: 61 962 009); Major Scientific and Technological Special Project of Guizhou Province (grant number 20 183 001); Science and Technology Support Plan of Guizhou Province (grant number [2020]2Y011); and Foundation of Guangxi Key Laboratory of Cryptography and Information Security (grant number GCIS202118).

References

[1] Y. Chen, J. Sun, Y. Yang, T. Li, X. Niu, and H. Zhou, "Psspr: a source location privacy protection scheme based on sector phantom routing in wsns," *International Journal of Intelligent Systems*, 2021.

[2] R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar, "Preserving user location privacy in mobile data management infrastructures," in *Proceedings of the International Workshop on Privacy Enhancing Technologies*, pp. 393–412, Springer, Cambridge, UK, June 2006.

[3] M. Gruteser and D. Grunwald, "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, New York, NY, USA, May 2003.

[4] S. Wang, Q. Hu, Y. Sun, and J. Huang, "Privacy preservation in location-based services," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 134–140, 2018.

[5] C. S. Wright, "Bitcoin: a peer-to-peer electronic cash system," *SSRN Electronic Journal*, 2008.

[6] T. Li, Y. Chen, Y. Wang et al., "Rational protocols and attacks in blockchain system," *Security and Communication Networks*, vol. 2020, Article ID 8839047, 11 pages, 2020.

[7] T. Li, Z. Wang, G. Yang, Y. Cui, Y. Chen, and X. Yu, "Semi-selfish mining based on hidden Markov decision process," *International Journal of Intelligent Systems*, vol. 36, no. 7, pp. 3596–3612, 2021.

[8] T. Li, Z. Wang, Y. Chen, C. Li, Y. Jia, and Y. Yang, "Is semi-selfish mining available without being detected?" *International Journal of Intelligent Systems*, 2021.

[9] G. Yang, Y. Wang, Z. Wang, Y. Tian, X. Yu, and S. Li, "Ipbsm: an optimal bribery selfish mining in the presence of intelligent and pure attackers," *International Journal of Intelligent Systems*, vol. 35, no. 11, pp. 1735–1748, 2020.

[10] F. Li, Z. Liu, T. Li, H. Ju, H. Wang, and H. Zhou, "Privacy-aware PKI model with strong forward security," *International Journal of Intelligent Systems*, 2020.

[11] Y. Chen, S. Dong, T. Li, Y. Wang, and H. Zhou, "Dynamic multi-key FHE in asymmetric key setting from LWE," *IEEE Transactions on Information Forensics and Security*, 2021.

[12] L. Hai, L. XingHua, L. Bin et al., "Distributed k-anonymity location privacy protection scheme based on blockchain," *Chinese Journal of Computers*, vol. 42, no. 5, pp. 942–960, 2019.

[13] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy," in *Proceedings of the 2012 ACM conference on Computer and communications security*, pp. 617–627, Association for Computing Machinery, New York, NY, USA, October 2012.

[14] J. Chen, K. He, Q. Yuan, M. Chen, R. Du, and Y. Xiang, "Blind filtering at third parties: an efficient privacy-preserving framework for location-based services," *IEEE Transactions on Mobile Computing*, vol. 17, no. 11, pp. 2524–2535, 2018.

[15] H. Jiang, J. Li, P. Zhao, F. Zeng, Z. Xiao, and A. Iyengar, "Location privacy-preserving mechanisms in location-based services," *ACM Computing Surveys*, vol. 54, no. 1, pp. 1–36, 2021.

[16] S. Zhang, G. Wang, M. Z. A. Bhuiyan, and Q. Liu, "A dual privacy preserving scheme in continuous location-based services," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 4191–4200, 2018.

[17] L. Yu, L. Liu, and C. Pu, "Dynamic differential location privacy with personalized error bounds," in *Proceedings of the Network and Distributed System Security Symposium*, San Diego, CA, USA, January 2017.

[18] R. Shokri, "Privacy games: optimal user-centric data obfuscation," *Proceedings on Privacy Enhancing Technologies*, vol. 1, no. 2, pp. 299–315, 2015.

[19] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," in *Proceedings of the ICPS'05. Proceedings. International Conference on Pervasive Services, 2005*, pp. 88–97, IEEE, Santorini, Greece, July 2005.

[20] H. Kido, Y. Yanagisawa, and T. Satoh, "Protection of location privacy using dummies for location-based services," in *Proceedings of the 21st International conference on data engineering workshops (ICDEW'05)*, p.1248, IEEE, Tokyo, Japan, April 2005.

[21] Z. Dapeng, S. Guangxuan, J. Yuanyuan, and W. Xiaoling, "Query probability-based location privacy protection approach," *Journal of Computer Applications*, vol. 37, no. 2, pp. 347–351, 2017.

[22] L. Chang, Z. Xing, Y. Fei, L. Wanjie, and L. Shuai, "Fake location generation scheme based on user preference selection," *COMPUTER ENGINEERING AND DESIGN*, vol. 40, no. 4, pp. 914–919, 2019.

[23] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Achieving k-anonymity in privacy-aware location-based services," in *Proceedings of the IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pp. 754–762, IEEE, Toronto, ON, Canada, May 2014.

[24] C. Hui and Q. Xiaolin, "Location-semantic-based location privacy protection for road network," *Journal on Communications*, vol. 37, no. 8, pp. 67–76, 2016.

[25] Z. Haiyan, Z. Kaizhong, W. Yonglu, and L. Rui, "Semantic diversity location privacy protection method in road network environment," *Computer Engineering and Applications*, vol. 56, no. 7, pp. 102–108, 2020.

[26] Z. Yongbing, Z. Qiuyu, L. Zongyi, D. Hongxiang, and Z. Moyi, "A k-anonymous location privacy protection method of dummy based on approximate matching," *Control and Decision*, vol. 35, no. 1, pp. 55–64, 2020.

[27] S. Chen and H. Shen, "Semantic-Aware dummy selection for location privacy preservation," in *Proceedings of the 2016 IEEE Trustcom/BigDataSE/ISPA*, pp. 752–759, IEEE, Tianjin, China, August 2016.

[28] W. Lu and M. XiaoFeng, "Location privacy preservation in big data era: a survey," *Journal of Software*, vol. 25, no. 4, pp. 693–712, 2014.

[29] Q. A. Arain, I. Memon, Z. Deng, M. H. Memon, F. A. Mangi, and A. Zubedi, "Location monitoring approach: multiple mix-zones with location privacy protection based on traffic flow over road networks," *Multimedia Tools and Applications*, vol. 77, no. 5, pp. 5563–5607, 2018.

[30] X. Zheng, Z. Cai, J. Li, and H. Gao, "Location-privacy-aware review publication mechanism for local business service systems," in *Proceedings of the IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pp. 1–9, IEEE, Atlanta, GA, USA, May 2017.

[31] X. XingYou, B. ZhiHong, L. Jie, and Y. RuiYun, "A location cloaking algorithm based on dummy and stackelberg game," *Chinese Journal of Computers*, vol. 42, no. 10, pp. 2216–2232, 2019.

[32] N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Optimal geo-indistinguishable mechanisms for location privacy," in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pp. 251–262, Association for Computing Machinery, New York, NY, USA, November 2014.

[33] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Proceedings of the 2011 IEEE symposium on security and privacy*, pp. 247–262, IEEE, Oakland, CA, USA, May 2011.

[34] W. Zhen, Y. Yong, A. Bo, L. MingChu, and W. FeiYue, "An overview of security games," *Journal of Command and Control*, vol. 1, no. 2, pp. 121–149, 2015.

[35] W. Sheng, L. Fenghua, N. Ben, S. Zhe, and L. Hui, "Research progress on location privacy-preserving techniques," *Journal on Communications*, vol. 37, no. 12, pp. 124–141, 2016.