Hindawi
Security and Communication Networks
Volume 2021, Article ID 5574415, 12 pages
https://doi.org/10.1155/2021/5574415

Research Article

Differential Privacy Location Protection Scheme Based on Hilbert Curve

Jie Wang, Feng Wang, and Hongtao Li

College of Mathematics & Computer Science, Shanxi Normal University, Linfen 041000, China

Correspondence should be addressed to Jie Wang; 429811049@qq.com

Received 4 February 2021; Revised 26 February 2021; Accepted 31 March 2021; Published 12 April 2021

Academic Editor: Ximeng Liu

Location-based services *(LBS)* applications provide convenience for people's life and work, but the collection of location information may expose users' privacy. Since these collected data contain much private information about users, a privacy protection scheme for location information is an impending need. In this paper, a protection scheme *DPL-Hc* is proposed. Firstly, the users' location on the map is mapped into one-dimensional space by using Hilbert curve mapping technology. Then, the Laplace noise is added to the location information of one-dimensional space for perturbation, which considers more than 70% of the nonlocation information of users; meanwhile, the disturbance effect is achieved by adding noise. Finally, the disturbed location is submitted to the service provider as the users' real location to protect the users' location privacy. Theoretical analysis and simulation results show that the proposed scheme can protect the users' location privacy without the trusted third party effectively. It has advantages in data availability, the degree of privacy protection, and the generation time of anonymous data sets, basically achieving the balance between privacy protection and service quality.

1. Introduction

With the rapid development of intelligent mobile devices and wireless communication technology, location-based services (*LBS*) applications not only bring about convenience to users but also cause serious privacy and security risks to users. In *LBS*, users provide their location information to location service providers while acquiring location services, which may lead to the leakage of users' sensitive information [1, 2]; for example, the access frequency of the users to the interest points can analyze the users' preference and economic status. At the same time, if the attacker combines the users' location information and nonlocation information, more personal information of the user will be exposed [3]. With the continuous improvement of people's awareness of privacy protection, protection of the user's location information becomes an urgent problem to be solved.

At present, location privacy protection methods mainly include k-anonymity technology and (α,k)-anonymity technology. k-anonymity technology [4–6] uses a trusted third-party (*TTP*) server to expand the user's real location into an invisible geographic location area that includes other *k-1* users, making untrusted *LSP* not able to distinguish the user's real location from the geographic location of other *k-1* users and then sending the confused location information to the *LSP* (location service provider) through *TTP*. k-anonymity technology provides the basis privacy protection research. However, k-anonymity technology does not restrict the sensitive attributes in the users' data set and is vulnerable to link attacks. To solve these problems, the authors [7, 8] propose (α,k)-anonymity technology based on the k-anonymity technology. (α,k)-anonymity technology presets the threshold α, so that the proportion of sensitive attribute values in each equivalence class will not exceed this threshold, blurring the link relationship between sensitive attributes and quasi-identifiers and enhancing the anonymity effect of users' data set. (α,k)-anonymity technology can enhance users' location privacy protection by anonymizing the data set through *TTP*. *TTP* has mastered all the

knowledge of the users' *LBS* query and is prone to suffer from a single point of failure. If the attacker captures the *TTP*, then the users' location privacy will be leaked. To protect the users' location privacy without relying on a third party, differential privacy technology [9–11] was proposed. In [2, 12], differential privacy and anonymous set with k locations are used to calculate the interference location, which can resist the attack of the adversary with useful background information. In [13], a δ-location set based on differential privacy is proposed to protect the users' real location at each time point under temporal correlation. But its disadvantage is that it ignores the attack mode of combining location data with nonlocation data. If attackers combine the user's location information at different times with some nonlocation information, user's private information will be seriously exposed.

The rest of the paper is organized as follows: the second section describes some definitions related to location privacy. In the third section, architecture and threat model of *LBS* system are analyzed in detail, and then the specific implementation algorithm of differential privacy location protection mechanism based on Hilbert curve (*DPL-Hc*) is introduced. In the fourth section, the *DPL-Hc* scheme is evaluated, including privacy analysis, security analysis, and algorithm complexity analysis. The fifth section verifies the effectiveness of the algorithm from the availability of published data sets, the degree of privacy protection of data sets, and the generation efficiency of anonymous data sets. The sixth section summarizes the *DPL-Hc* scheme, in which privacy protection is strengthened and the balance between privacy protection degree and service quality is solved effectively. The seventh section is devoted to the references used in this paper.

2. Related Definitions

Definition 1 (Hilbert curve). Hilbert curve is used as the mapping from dimensional S-space R^S to one-dimensional space R, denoted as $H: R^S \longrightarrow R$. If point $p \in R^S$, then $H(p) \in R$; that is, $H(p)$ is the H value of point p. For point set $\{p_1, p_2, \ldots, p_n\}$,
$H\{p_1, p_2, \ldots, p_n\} = \{H(p_1), H(p_2), \ldots, H(p_n)\}$. Coding rules of the first-order, second-order, and third-order Hilbert curves are shown, respectively, in Figure 1.

Definition 2 (location differential privacy). For two data sets D and D' that differ by at most one location record, namely, $|D - D'| \leq 1$, given a differential privacy algorithm A, *Range(A)* is the range of A. Algorithm A provides ε-differential privacy, and ε is the privacy budget, which represents the degree of privacy protection. If the arbitrary position $L(L \in \text{Range}(A))$ obtained by algorithm A from arbitrary trajectory data set D sand D' satisfies the following inequality, then algorithm A satisfies ε-differential privacy.

$$\Pr[A(D) \in L] \leq \Pr[A(D') \in L] e^{\varepsilon}. \tag{1}$$

Probability $Pr(.)$ represents the risk of users' location privacy being exposed, and it is randomly controlled by algorithm A; parameter ε is the privacy budget. It can be seen from the above formula that the smaller the parameter ε is, the more similar the probability distribution of the query results returned by the-differential privacy algorithm acting on a pair of adjacent trajectory data sets is, and the more difficult it is for the attacker to distinguish this pair of adjacent trajectory data sets. In extreme cases, when $\varepsilon = 0$, the degree of privacy protection is the highest. On the contrary, the higher the value of the parameter ε, the lower the degree of privacy protection.

Definition 3 (Global Sensitivity). For any function $f: D \longrightarrow R^d$, the global sensitivity of function f is

$$S(f) = \max_{D.D'} \left\| f(D) - f(D') \right\|_1, \tag{2}$$

where $\|f(D) - f(D')\|_1$ represents the first-order norm distance of the function output values of adjacent data sets D and D', and sensitivity refers to the maximum change in the output value of the function caused by adding or deleting any record in the data set. The global sensitivity of the query function $S(f)$ is determined by the properties of the function itself and is independent of the data set.

Definition 4 (Laplace Mechanism). Given a function f, the Laplace mechanism is defined as

$$A(D) = f(D) + \langle Y_1, \ldots, Y_k \rangle, \tag{3}$$

where Y_i is a random variable of the Laplace distribution $\text{lap}(S(f)/\varepsilon)$. The location parameter of Laplace distribution is 0, the scale parameter is $b = (S(f)/\varepsilon)$, and its probability density function is

$$p(x) = \frac{1}{2b} e^{-(|x|/b)}. \tag{4}$$

The added noise is proportional to the global sensitivity $S(f)$ and inversely proportional to the privacy budgetε. Laplace mechanism is limited to the functions whose return value is real, and the exponential noise mechanism is proposed for nonnumerical query function.

Definition 5 (exponential mechanism). Given a utility function u: $(D^n \times R) \longrightarrow R$, r is an entity object in the output domain range of the available function. If the output of function u satisfies equation (5), then theε-differential privacy is satisfied.

$$A(D, u) = \left\{ r: \Pr[r \in \text{Range}] \infty e^{(\varepsilon u(D,r)/2S(u))} \right\}, \tag{5}$$

where $S(u)$ is the global sensitivity of the utility function, and the exponential mechanism returns the entity object with a probability proportional to $e^{(\varepsilon u(D,r)/2S(u))}$.

3. Differential Privacy Location Protection Mechanism Based on the Hilbert Curve

3.1. System Architecture and Threat Model Analysis. The system architecture used in this paper is shown in Figure 2. The system architecture is mainly composed of four parts:

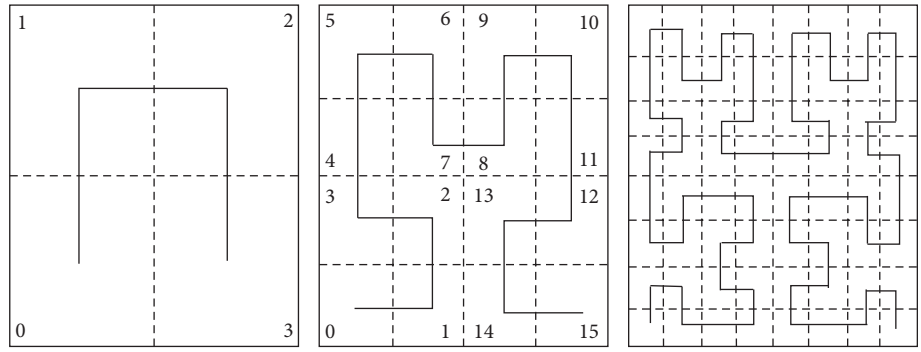

Figure 1: Coding rules of the first-order, second-order, and third-order Hilbert curves.

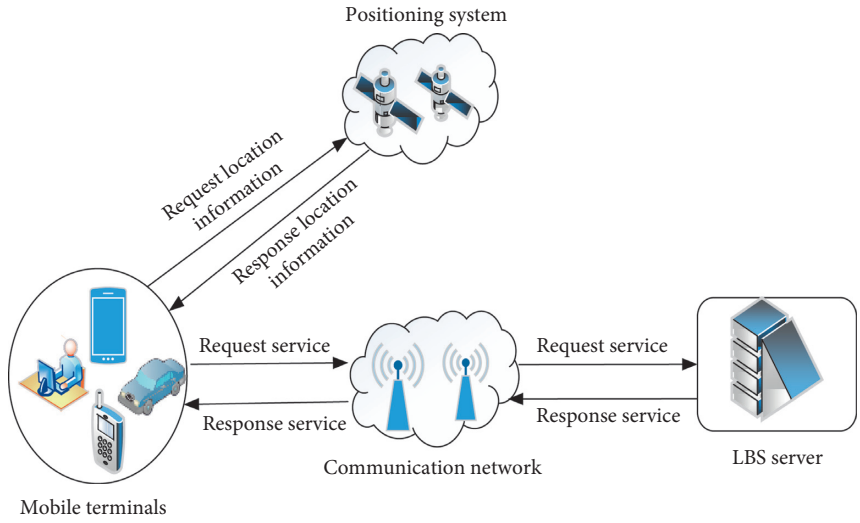

Figure 2: Basic system architecture of *LBS*.

positioning system, mobile terminal, communication network, and *LBS* server. The mobile user holds mobile positioning devices (smartphones and vehicle-mounted mobile terminals) to obtain personal accurate geographic location information through the Internet or *GPS* and other positioning technologies and then sends the query request to the *LBS* server through Wi-Fi and other communication networks. After the *LBS* server receives the query request from the user, it sends the service information to the user's mobile terminal as a response to complete a coherent process of request service and response service.

User makes a query request through the mobile terminal, and *LBS* may be exposed to the threat of privacy leakage in the process of responding to the request. Several types of attackers in the system are shown in Figure 3. Attackers include untrusted *LSP* itself, internal attackers, and external attackers in the system. Generally speaking, *LSP* is honest but curious, which can provide query service for users honestly according to the agreement. However, in order to improve its own commercial interests, the service provider collects user's location information by observing the received disturbance location. In addition, there may also be attackers within the organization that provides the service, sending users' disturbed locations to other external organizations for their own benefit; external attackers can be individuals or organizations by eavesdropping on users' data or attacking the server to access to the server.

When the user sends out a query request, differential perturbation mechanism is implemented on the user's mobile. Add Laplace noise to the user's geographic location and provide the disturbed location to the *LSP*. When the *LSP* responds to user's query and returns the query results to user, it does not infer user's real geographical location combined with some background information, which largely protects users' location privacy.

3.2. Implementation Mechanism and Algorithm. In this paper, interest points are defined as the real interest points in geographic space. Each interest point L can be approximately expressed as $L(x, y, z)$, where (x, y) is the location coordinate of the interest point L and z represents the semantic location of the interest point L. The privacy protection of users is to achieve context sensing. The *DPL-Hc* scheme obtains context (such as the same density) through the distribution of users' interest points. The optimal data clustering and distance preserving characteristics of the Hilbert curve make two adjacent points in two-dimensional

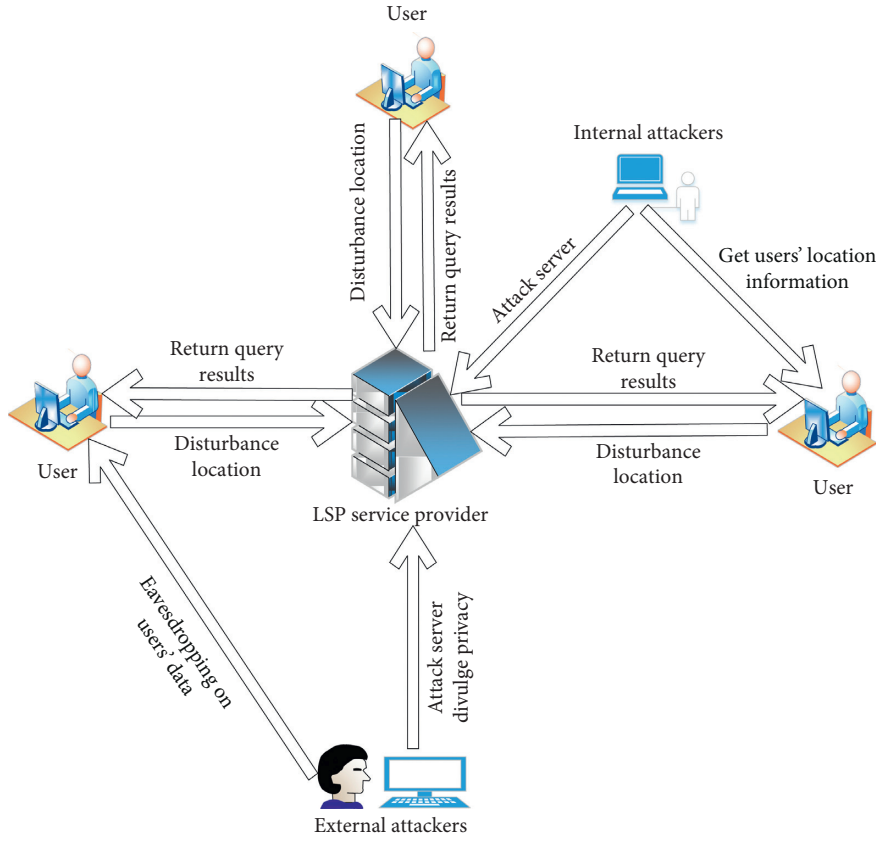

Figure 3: Threat model.

space more likely to be adjacent in one-dimensional space; that is, all interest points in one-dimensional space have homogeneous context. When users request *LBS* query service, the users' two-dimensional space interest points are first mapped to the one-dimensional space; then the Laplace noise is added to $l = (l_1, l_2, \ldots, l_k)$ locations containing the users' real location; the disturbance location from the l_k location points set is almost the same and the attacker cannot distinguish the users' real location. As shown in Figure 4, the third-order Hilbert curve coding rule is adopted in this paper, and the circle points $P_i\,(i = 1, \ldots, 15)$ in the figure are the interest points after projection. The number in each atomic unit represents the Hilbert value of that atomic unit.

In this paper, we first use the quadtree index structure to index the users' location and divide the area containing all users' locations into a quadtree, and index to obtain a quadtree QT with location sets L. Then the location data is processed. According to the Hilbert curve technology mentioned above, the users' two-dimensional geographic locations are mapped into one-dimensional space, and the semantic information of the location elements is retained to obtain the corresponding complete tree data structure, which improves the efficiency of searching the target point in the future. The quadtree after Hilbert curve mapping is shown in Figure 5.

Grid division of geographic areas is one of the effective methods to describe the location of interest points. As shown in Figure 4, the *DPL-Hc scheme* starts from the area containing the users' locations and divides the geographic space into 4 grids at a time and then iterates to obtain 4^{h-1} grids (h is the division height) until a certain granularity is met. It is divided into some atomic regions that can no longer be divided, and the size of the atomic region is determined by the number of users' locations C which the region can accommodate. In Figure 4, the area containing the users' locations is divided into 8*8 grids, and each interest point uniquely belongs to a grid. In this paper, we set $h = 4, C = 1$. The *DPL-Hc* scheme uses the quadtree structure to generate users' location index, as shown in Algorithm 1. For the convenience of description, the symbol definitions involved in this algorithm are shown in Table 1.

The algorithm divides the geographic area into four parts according to the set of interest points P and the number of layers of quadtree and indexes $0 < j \leq 2$ the users' locations. If the number of divided layers creates four new subnodes for the node of the layer, as shown in steps 1–6, and if it creates four new subnodes for the nodes of other segmentation layers, as shown in step 10, then, for each interest point belonging to Z_p, if the interest point is stored in the region of the i-th child node of the quadtree node Z, these interest points are moved to their respective child nodes, as shown in steps 11–13. Finally, confirm which subnode the point p belongs to, and then recursively call *Qua dTreeInsert* to insert p into node D, as shown in steps 16 and 17. The above statements (1–17) are executed

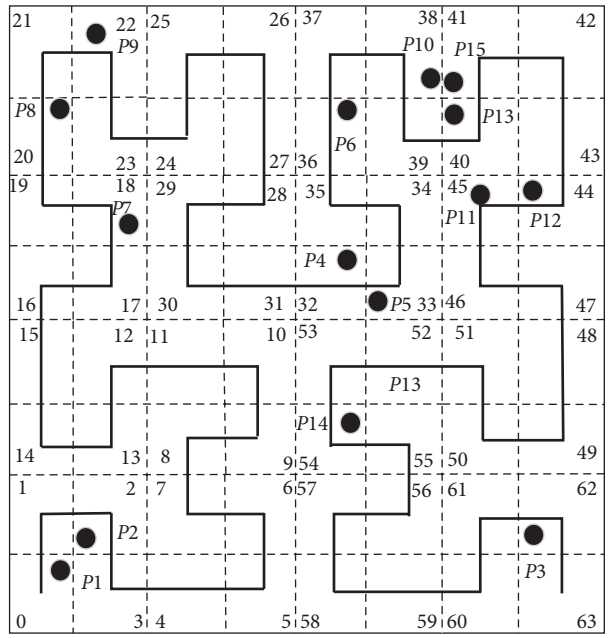

Figure 4: Location partition based on the Hilbert curve.

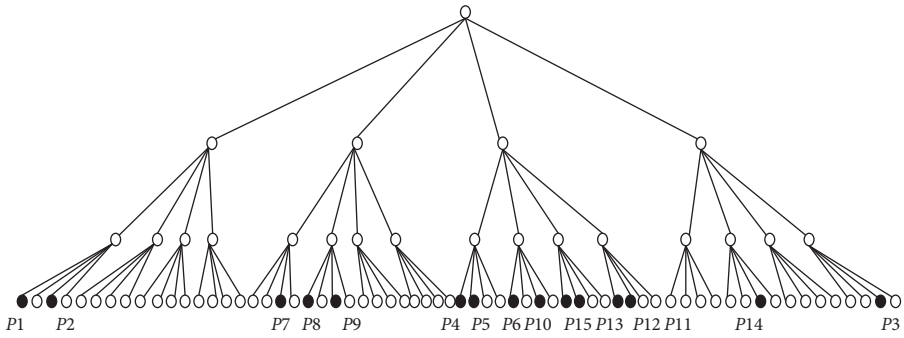

Figure 5: Quadtree representation of the Hilbert curve partition region.

circularly until all interest points are inserted into node D, and finally the quadtree QT with location set is output.

The DPL-Hc scheme traverses and searches the tree formed after the location processing, obtains the location data L and nonlocation data B of each marker point, and applies differential privacy technology to separately add noise to location data and over 70% of nonlocation data.

The *DPL-Hc* scheme needs to combine the users' location information with more than 70% of nonlocation information to protect the users' privacy information. Adding noise to the coordinates of an interest point separately provides a higher degree of privacy protection than adding noise to the point alone. Formally let $l_1, \ldots, l_k$ denote k interest points, $l_t = (x_t, y_t)$ is the user's real interest point, and $l_p(x_p, y_p)$ is the disturbed interest point. For any two interest points l_i and l_j, their probability of generating a perturbed interest point $l_p(x_p, y_p)$ should satisfy the following equation:

$$P\left(\frac{x_p}{x_i}\right) \le e^{\varepsilon} P\left(\frac{x_p}{x_j}\right),$$
$$P\left(\frac{y_p}{y_i}\right) \le e^{\varepsilon} P\left(\frac{y_p}{y_j}\right). \tag{6}$$

For more than 70% of the nonlocation information, given the privacy budgetε, noise is added to the nonlocation information collected by traversal to meet the requirements of differential privacy. For the attribute set $B = \{B_1, \ldots, B_K\}$ of nonlocation data D, its continuous-valued attribute is marked as $A_j^B = \{B_j \in A \cap (B = B_j)\}$, and its noncontinuous-valued attribute is marked as $C_j^B = \{B_j \in C \cap (B = B_j)\}$. Different mechanisms of noise are added to the nonlocation information of different attributes. The Laplace noise is added to the continuous value to be disturbed, and the exponential noise is added to the noncontinuous value to be

Input: quad-tree root node Q, interest point set P, quad-tree sub node-set Z, the number of levels of quad-tree j
Output: quad-tree index QT with location set
(1) If $0 < j \leq 2$
(2) $j \leftarrow 1$;
(3) While $j(z) < j$ do
(4) For layer j do
(5) $Z_C \leftarrow \{Z(0), Z(1), Z(2), Z(3)\}$;//Create four new child nodes for the node on layer j
(6) $j++$;
(7) End For
(8) End While
(9) Else
(10) $Z_C \leftarrow \{Z(0), Z(1), Z(2), Z(3)\}$;//Create four new child nodes for the node on this layer
(11) For each $E \in Z_p$ //For each interest point belonging to Z_p
(12) If $E_L \in Z(i)_R$ //If the interest point is stored in the region of the i-th child node of the quad-tree node Z
(13) $Z(i)_P \leftarrow Z(i)_R \cup E$;//Move the interest points to their child nodes
(14) End If
(15) End For
(16) get $D \in Z_C$ which meets $p_L \in D_R$ //Mark the child node of interest point Pas D
(17) $Quad TreeInsert(p, D)$; //Recursively call $Quad TreeInsert$ to insert the interest point p into node D
(18) End If
(19) Return Q

Algorithm 1: Spatial quadtree generation.

Table 1: Symbol definition.

Symbol	Definition
Z_R	The region corresponding to node Z of quadtree
Z_P	The set of interest points stored in node Z of quadtree
Z_C	The set of children of node Z in quadtree
$Z(i)$	The i-th child of the quadtree node Z
p_L	The position coordinates of the interest point p

disturbed. The exponential mechanism outputs discrete values with a probability proportional to $e^{(\varepsilon u(D,r)/2S(u))}$. In the *DPL-Hc* scheme, the anonymous data processing procedure is shown in Algorithm 2.

The input parameters of the algorithm are users' location data set L, given the privacy budget ε, users' nonlocation data set D, nonlocation data attribute set B, continuous attribute data set A_j^B, discrete attribute data set C_j^B, and tree height h. The processing objects are location data set L and nonlocation data set D. The algorithm first allocates the privacy budget in step 2; then the location data and nonlocation data are classified. For any element L_i of the location data set L, Laplace noise is added to its abscissa and ordinate, respectively, for differential perturbation in steps 3–5; next, the nonlocation data set is divided according to attributes in step 8. If the nonlocation data belongs to continuous-valued attribute, Laplace noise is added for differential perturbation in steps 9 and 10; if the nonlocation data belongs to discrete-valued attribute, exponential noise is added for differential perturbation in steps 11 and 12. Execute the above statements (3–12) circularly until anonymous processing is performed for all location data and more than 70% of the nonlocation data. Finally, the anonymous data set $T(T_{L_i}, T_{D_i})$ is output in step 16.

4. Theoretical Analysis of Algorithm

4.1. Privacy Analysis. In this paper, the users' location points are represented by specific abscissa and ordinate. Given the users' location points set $l = (l_1, l_2, \ldots, l_k)$, one of which is the user's real location $l_t = (x_t, y_t)$, the Laplace noise is added to the users' location points to disturb the user's real location, and $l_p(x_p, y_p)$ is the interest point after the disturbance. For any two interest points $l_i(x_i, y_i)$ and $l_j(x_j, y_j)$ in the k interest points, according to the Laplace mechanism, there is

$$P\left(\frac{x_p}{x_i}\right) = \frac{1}{2b} e^{-\left(\left|x_i - x_p\right|/b\right)},$$
$$P\left(\frac{y_p}{y_i}\right) = \frac{1}{2b} e^{-\left(\left|y_i - y_p\right|/b\right)}. \tag{7}$$

Let b be $(\max_n x_n - \min_n x_n)/\varepsilon$ to produce x_p and $(\max_n y_n - \min_n y_n)/\varepsilon$ to produce y_p, where $\max_n x_n$ is the maximum in $x_1, \ldots, x_n$ and $\min_n x_n$ is the minimum in $x_1, \ldots, x_n$. Thus, the perturbed interest point $l_p(x_p, y_p)$ can be obtained.

For interest points l_i, l_j, and l_p, we can get the following results from triangle inequality:

$$\left|l_j - l_p\right| \leq \left|l_j - l_i\right| + \left|l_i - l_p\right|. \tag{8}$$

Rearrange formula (8), and divide both sides by b and raise them to a power exponent with base e; then multiply both sides by $(1/2b)$ to get

$$\frac{1}{2b} e^{-\frac{\left|l_i - l_p\right|}{b}} \leq \frac{1}{2b} e^{-\frac{\left|l_j - l_p\right|}{b}} e^{-\frac{\left|l_i - l_j\right|}{b}}. \tag{9}$$

Input: ε-privacy budget L-location data set D-non-location data set, $B = \{B_1, \ldots, B_K\}$-non-location data attribute collection, A_j^B-continuous attribute data set, C_j^B-discrete attribute data set, h-the height of the tree.
Output: anonymous data set T satisfying differential privacy protection.
(1) Begin Procedure $T(\varepsilon, L, D, A_j^B, C_j^B)$
(2) $\overline{\varepsilon} = (\varepsilon/h)$;//privacy budget allocation
(3) If the data belongs to location data L
(4) For any element L_i in set L
(5) $L_{x_p} = \text{LapNoise}_\varepsilon[Q(L_{x_i})], L_{y_p} = \text{LapNoise}_\varepsilon[Q(L_{y_i})]$; //$Q$ is the user's query function, which has global sensitivity and adds Laplace noise to the location data
(6) End For
(7) Else If the data belongs to nonlocation data D
(8) For any element P in set D, and the element satisfies $D_i \in \{D \cap B_j \in \{B_1, \ldots, B_k\}\}$ //Attribute partition of nonlocation data
(9) If B_j is a continuous-valued attribute, then $D_i \in A_j^B$
(10) $T_{D_i} = LapNoise_\varepsilon[Q(D_i)]$; //Adding Laplace noise to nonlocation data with the continuous-valued attribute
(11) Else If B_j is a discrete-valued attribute, then $D_i \in A_j^B$
(12) $T_{D_i} = ExpMech_\varepsilon[Q(D_i)]$; //Laplace noise is added to nonlocation data with the discrete-valued attribute
(13) End If
(14) End For
(15) End If
(16) Return $T(T_{L_i}, T_{D_i})$ //Output anonymous data set
(17) End Procedure

Algorithm 2: Anonymous data processing procedure.

Table 2: Experimental data set.

Data set	Attribute set	
	Location data	Nonlocation data
Geolife (38494)	Latitude Longitude Country, city, street	Mode of transportation Mode of life Social information search behavior
Diversification data set Div400 (43418)	Latitude Longitude Country, city, street	Sharing information Text image Personal data folder
Amazon Access Samples (60000)	User location information	User account number User transaction information User occupation User work unit Basic family information

From equations (8) and (9), we have the following equation:

$$P\left(\frac{l_p}{l_i}\right) \le P\left(\frac{l_p}{l_j}\right) e^{\frac{|l_j - l_i|}{b}}. \tag{10}$$

For coordinates x_i and y_i, equation (10) can be expressed as

$$P\left(\frac{x_p}{x_i}\right) \le e^{\frac{|x_j - x_i|}{b}} P\left(\frac{x_p}{x_j}\right), \tag{11}$$

$$P\left(\frac{y_p}{y_i}\right) \le e^{\frac{|y_j - y_i|}{b}} P\left(\frac{y_p}{y_j}\right). \tag{12}$$

By setting the exponential boundary of equations (11) and (12), we can get

$$\begin{aligned} P\left(\frac{x_p}{x_i}\right) &\le e^{\frac{|\max_n x_n - \min_n x_n|}{b}} P\left(\frac{x_p}{x_j}\right), \\ P\left(\frac{y_p}{y_i}\right) &\le e^{\frac{|\max_n y_n - \min_n y_n|}{b}} P\left(\frac{y_p}{y_j}\right), \end{aligned} \tag{13}$$

that is,

$$\begin{aligned} P\left(\frac{x_p}{x_i}\right) &\le e^{\varepsilon} P\left(\frac{x_p}{x_j}\right), \\ P\left(\frac{y_p}{y_i}\right) &\le e^{\varepsilon} P\left(\frac{y_p}{y_j}\right). \end{aligned} \tag{14}$$

It can be seen from the above that the anonymous data processing algorithm in the *DPL-Hc* scheme satisfies ε-differential privacy.

4.2. Safety Analysis. In location privacy protection method provided in this paper, the user will submit a query service to the *LBS* server in order to query the interest point closest to the current location; for example, query the movie theater closest to the user's location. Ideally, due to the disturbance, the attacker cannot identify any connection between the disturbed location and the user's real location. However, when the attacker knows the density of interest points on the map, the user's approximate location knowledge, and noise distribution, the attacker can infer the user's real location based on this information. In the process of anonymity, the Laplace distribution mechanism with scale parameter $b \geq 0$ is used to calculate the probability of the same disturbance location l_p from the location point set $l_1, \ldots, l_r, \ldots, l_k$ which is limited to a small constant factor e^{ε}. Given the perturbation point $l_p = (x_p, y_p)$, no matter which interest point is used to implement the perturbation, the probability of k interest points to produce the perturbation is the same (in the range of constant factor e^{ε}), and the attacker cannot use the above information to improve the probability of guessing the user's real location. Therefore, the *DPL-Hc* scheme can effectively protect the user's location privacy.

4.3. Algorithm Complexity Analysis. Firstly, the anonymous data processing algorithm in this paper uses a greedy method to recurse the quadtree from top to bottom, and the time complexity is $O(\log_2 n)$. Then the algorithm classifies the data information contained in each node, and the time complexity of adding noise to the location data set is $O(|L|)$. For the data with continuous attributes in nonlocation data set, the time complexity of adding Laplace noise is $O(|D_1|\log_2|D_1|)$; for the data with discrete attributes in nonlocation data set, the time complexity of adding exponential noise is $O(|D_1|\log_2|D_1|)$.

5. Experimental Results and Analysis

5.1. Experimental Setup. In order to study the feasibility of the algorithm proposed in this paper, the system hardware configuration adopted is an Intel*(R)* Core(*TM*) *i7* compatible *PC* with a main frequency of 3.4 *GHz*, 4 *GB* of memory, and more than 200 GB of free disk space; the software configuration platform is Windows 7 operating system, Microsoft *SQL* Server database system, and *C/S* structure operating mode. The experiment is based on three data sets: Geolife, Amazon Access Samples, and Diversification data set Div400. The source databases of the data sets are Geolife *GPS* Trajectory stores, *UCI* Machine Learning Repository, and UMass Trace Repository. The experimental data sets include the users' location information and nonlocation information. The data set size and attribute characteristics are shown in Table 2.

5.2. Experimental Results. Commonly used spatial indexing technology can improve the operational efficiency of spatial information databases. The *DPL-Hc* scheme adds differential privacy anonymity technology based on the quadtree spatial index technology and divides location space recursively into a tree structure of different levels. When spatial data objects are evenly distributed, they have higher spatial data insertion and query efficiency. The *KDCK-medoids* algorithm [14] adds differential privacy anonymity technology based on the *k-d* tree spatial index technology. The *k-d* tree is a structural form of multidimensional retrieval. It divides the location points in the k-dimensional space and makes branching decisions for the corresponding objects according to the discriminator of the layer in each layer. It has the same good performance as a binary tree for matching and finding exact points (the average search length is $1 + \log_2 n$). The *PR-CAN* algorithm [15] introduces r-tree spatial index technology to make local indexes meet the requirements of differential privacy. All leaf nodes with overlapping regions in the r-tree are redivided into disjoint regions. For leaf nodes with mutually exclusive regions, independent noise adding mechanism makes *PR*-tree meet the requirements of differential privacy. In this paper, the data availability, the degree of privacy protection, and the generation time of anonymous data sets are used to verify the effectiveness of the scheme.

5.2.1. Data Availability. In the simulation experiment, under the condition of a gradual increase of privacy budgetε, we test three data sets that meet the privacy requirements, compare the accuracy of anonymous data output by the algorithm, and analyze the data availability of the algorithm under different privacy budgetsε. Choosing different scale transformation parameters $(S(f)/\varepsilon)$, the amount of noise is proportional to the global sensitivity $S(f)$ and inversely proportional to the privacy budget ε. To verify the impact of the *DPL-Hc* scheme, the *KDCK-medoids* algorithm, and the *PR-CAN* algorithm on the data availability under the requirements of differential privacy protection, the performance of the algorithm under different data sets and different privacy budget requirements is tested in this paper. The precision of analyzing the publishable data can reflect the availability of the algorithm to process the data set under the condition of meeting the query requirements. The published data precision of each experimental data set is shown in Figures 6–8.

The privacy budget is inversely proportional to the degree of privacy protection. When the privacy budget is smaller, the degree of data protection is greater, and when the degree of privacy protection$\varepsilon = 0$, perfect privacy is achieved. As can be seen from Figures 6–8, with the increase of privacy protection budget, the protection degree of differential algorithm on published data decreases, so the data availability of the three algorithms has increased. The location space tree structure generated by the quadtree index in this paper has nothing to do with the nature of the experimental data set, so the precision of differential privacy publishing data based on the quadtree index is improved with the increase of privacy budget. Compared with the *KDCK-medoids* algorithm and the *PR-CAN* algorithm, the *DPL-Hc* scheme has higher data availability, while maintaining certain algorithm stability. However, with the increase of privacy budget, the precision of the *KDCK-medoids* algorithm and the *PR-CAN* algorithm is lower than that of the *DPL-Hc* scheme; that is, they have poor data availability and algorithm stability.

Table 3: Generation time of Geolife anonymous data set.

ε	*DPL-Hc*	*KDCK-medoids*	*PR-CAN*
0.1	2.50	3.85	4.89
0.2	2.12	3.02	4.25
0.3	1.83	2.53	4.01
0.4	1.52	2.03	3.56
0.5	1.03	1.75	3.03
0.6	0.81	1.01	2.89
0.7	0.52	0.43	2.65
0.8	0.35	0.15	2.43
0.9	0.27	0.11	2.21
1.0	0.15	0.05	1.98

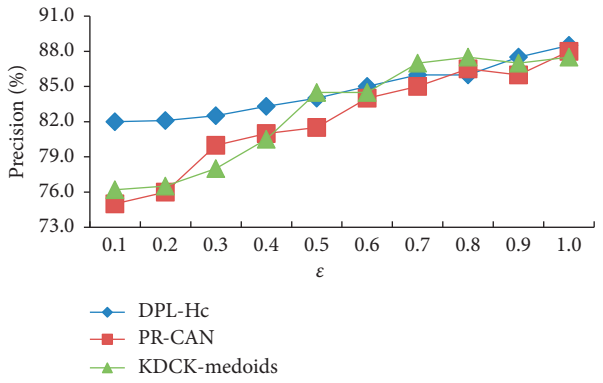

Figure 6: Data availability of Geolife data set.

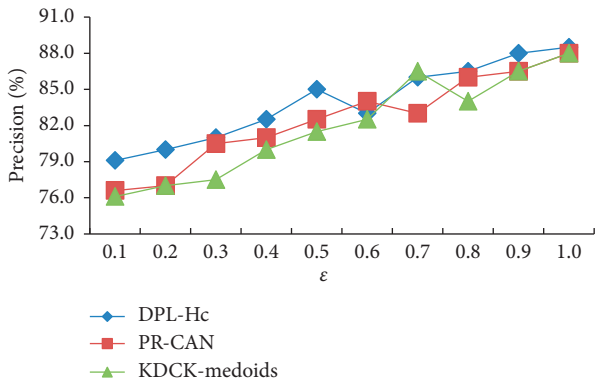

Figure 7: Data availability of Amazon Access Samples data set.

5.2.2. The Degree of Privacy Protection

(1) The Relationship between the Degree of Privacy Protection and the Laplace Transform Scale Parameter b. To research the privacy protection performance of the algorithm and find the best balance between data availability and the degree of privacy protection, the experiment analyzes and compares the average privacy protection degree of each data set under different Laplace transform scale parameter $b = (S(f)/\varepsilon)$, and the larger the scale parameter b, the higher the degree of privacy protection. Comparing the *DPL-Hc* scheme with the *KDCK-medoids* algorithm and the *PR-CAN* algorithm, the results of the experimental data set being anonymously

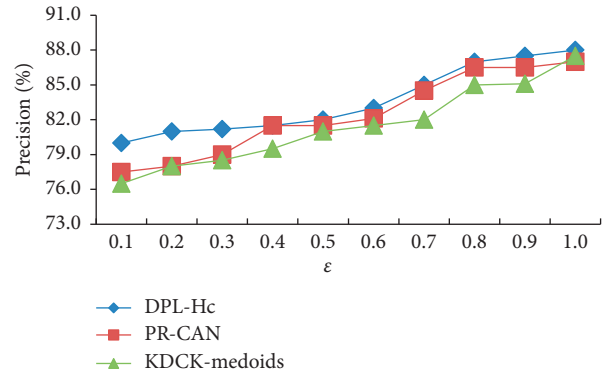

Figure 8: Data availability of Div400 data set.

protected by differential privacy are shown in Figures 9–11. From the experimental results, we can see that, with the transformation of Laplace scale parameter b, the privacy protection degree of *DPL-Hc* scheme can basically reach more than 80%, and, with the improvement of anonymity requirements, the execution efficiency of the algorithm will not be greatly reduced. However, the privacy protection degree of the *KDCK-medoids* algorithm and the *PR-CAN* algorithm is less than 80% when the scale parameter b is relatively small. It can also be seen from the figure that when the scale parameter b is determined, the privacy protection degree of the *DPL-Hc* scheme is higher than those of the *KDCK-medoids* algorithm and *PR-CAN* algorithm, and, with the increase of scale parameter b, the privacy protection performance of *DPL-Hc* scheme is more stable.

(2) The Relationship between the Degree of Privacy Protection and Anonymized Nonlocation Data Ratio k. To protect users' location privacy better, the *DPL-Hc* scheme takes into account the inference of the user's location privacy by nonlocation information and performs differential anonymous processing on more than 70% of nonlocation data. The greater the proportion k of anonymized nonlocation data, the higher the user's location privacy protection. Assuming that the sensitivity of each data in the users' nonlocation data set is equal, Figure 12 shows the influence of the *DPL-Hc* scheme on the degree of location privacy protection under different anonymized nonlocation data ratio k and data set requirements.

Through the experimental comparison, we can see that, with the increase of the proportion k of anonymous nonlocation data, the degree of privacy protection of each data set is improved. It can also be seen from the experiment that when the anonymized nonlocation data ratio k is fixed, the privacy protection degree of each data set is not much different; that is, the *DPL-Hc* scheme has good algorithm stability on the premise of ensuring the privacy protection degree. However, from the perspective of the data availability, the anonymous nonlocation data ratio k will have a certain impact on the data availability, so the nonlocation data ratio k should not be too high, and k is set to 75% in this paper.

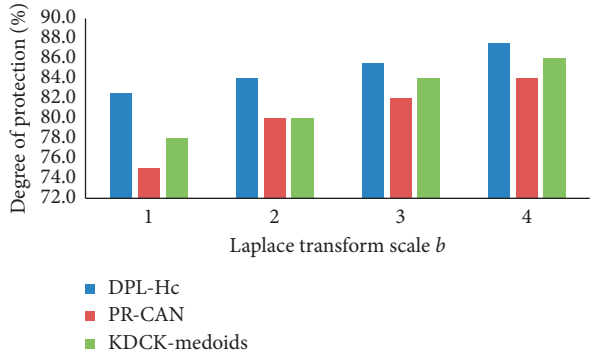

Figure 9: Privacy protection degree of Geolife data set.

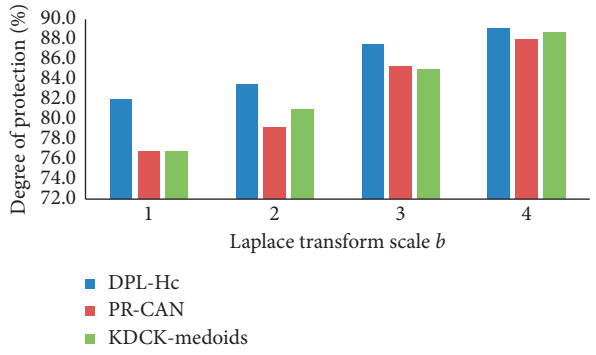

Figure 10: Privacy protection degree of Amazon Access Samples data set.

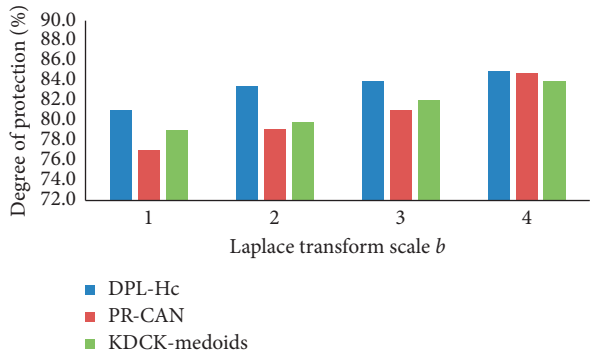

Figure 11: Privacy protection degree of Div400 data set.

5.2.3. The Generation Time of Anonymous Data Sets. Considering the choice of adding differential privacy anonymity methods to different spatial index trees, the average times taken by the *DPL-Hc* scheme, the *KDCK-medoids* algorithm, and the *PR-CAN* algorithm to generate anonymous data sets are shown in Table 3–5. Figures 13–15 show the comparison results of the generation time of three methods under different data sets and different privacy budgetsε, where $\varepsilon \in [0.1, 1.0]$.

As can be seen from Figures 13–15, with the increase of privacy budget ε, that is, the reduction of privacy protection degree, the *DPL-Hc* scheme takes much less time to generate anonymous data sets than the *PR-CAN* algorithm, and the *DPL-Hc* scheme is more efficient to generate anonymous data sets. Because *DPL-Hc* scheme uses the spatial index technology of quadtree, this technology has nothing to do with data in the process of constructing quadtree and avoids unnecessary overhead in the process of constructing quadtree. When $\varepsilon \geq 0.7$, the time of generating anonymous data set by the *DPL-Hc* scheme is slightly higher than that by the *KDCK-medoids* algorithm. When $\varepsilon \leq 0.6$, the time of generating anonymous data set by the *DPL-Hc* scheme is lower than that by the *KDCK-medoids* algorithm.

Through the comparison, we can also see that, with the decrease of ε, the time for the three algorithms to generate anonymous data sets is increasing, but the *DPL-Hc* scheme is obviously smaller than the *KDCK-medoids* algorithm and the *PR-CAN* algorithm. In other words, with the increase ofε, the *DPL-Hc* scheme takes less time, has more obvious advantages, and is more practical.

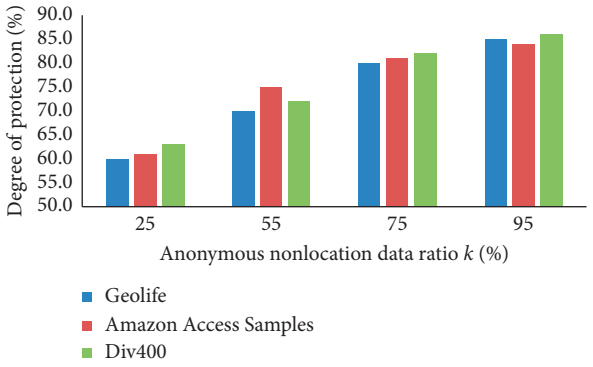

Figure 12: The degree of privacy protection.

Table 4: Generation time of Amazon Access Samples anonymous data set.

ε	*DPL-Hc*	*KDCK-medoids*	*PR-CAN*
0.1	2.30	3.02	4.25
0.2	171	2.76	3.23
0.3	1.65	2.32	3.10
0.4	1.42	1.52	3.02
0.5	1.21	1.44	3.00
0.6	0.81	1.12	2.52
0.7	0.63	0.21	2.33
0.8	0.43	0.15	2.12
0.9	0.23	0.11	2.09
1.0	0.15	0.05	1.78

Table 5: Generation time of Div400 anonymous data set.

ε	*DPL-Hc* (s)	*KDCK-medoids* (s)	*PR-CAN* (s)
0.1	2.20	3.25	4.36
0.2	2.00	3.02	4.25
0.3	1.75	2.72	3.50
0.4	1.50	1.99	2.95
0.5	1.0	1.76	2.89
0.6	0.87	1.43	2.43
0.7	0.55	0.15	2.21
0.8	0.39	0.11	2.03
0.9	0.25	0.07	1.92
1.0	0.15	0.04	1.63

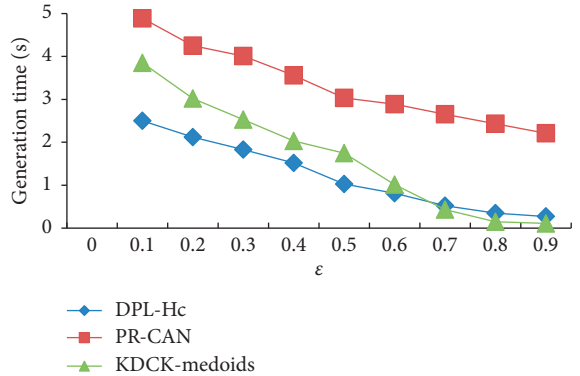

Figure 13: Generation time of Geolife anonymous data set.

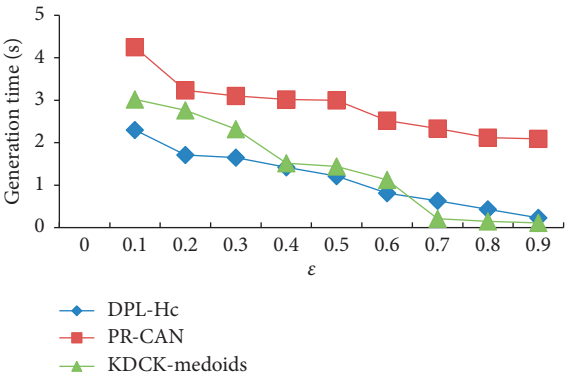

Figure 14: Generation time of Amazon Access Samples anonymous data set.

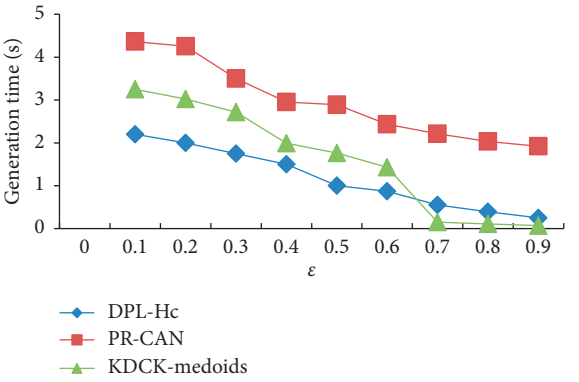

Figure 15: Generation time of Div400 anonymous data set.

Through the above experimental comparison, it is found that the *DPL-Hc* scheme has higher data availability and shorter generation time of anonymous data set on the premise of satisfying privacy protection as far as possible. The *DPL-Hc* scheme can protect the user's location privacy and improve the quality of location services effectively.

6. Conclusion

Aiming at balancing the degree of privacy protection and the quality of services in the *LBS* system, a differential privacy location protection scheme based on Hilbert curve on the basis of the existing differential privacy model is proposed in this paper. The scheme no longer relies on *TTP* and adds Laplace noise to the users' location in one-dimensional space mapped by the Hilbert curve. It can prevent the attacks of adversaries with background information and has strong privacy protection strength. It can effectively solve the balance problem between the degree of privacy protection and the quality of services. Experimental results show that the *DPL-Hc* scheme has obvious advantages in data availability, the degree of privacy protection, and the generation time of anonymous data sets.

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 work was supported by the National Natural Science Foundation of China (grant no. 61702316) and Shanxi Provincial Natural Science Foundation (Grant nos. 201801D221177 and 201901D111280).

References

[1] L. Wang and X. F. Meng, "Location privacy preservation in big data era: a survey," *Journal of Software*, vol. 25, no. 4, pp. 693–712, 2014.

[2] R. Dewri, "Local differential perturbations: location privacy under approximate knowledge attackers," *IEEE Transactions on Mobile Computing*, vol. 12, no. 12, pp. 2360–2372, 2013.

[3] L. Zhang, Y. Liu, and R.-C. Wang, "Location publishing technology based on differential privacy-preserving for big data services," *Journal on Communications*, vol. 37, no. 9, pp. 46–54, 2016.

[4] Z. Jialei, Z. Bocheng, F. Baogang et al., "An improvement of track privacy protection method based on k-anonymity technology," *Intelligent Computer and Applications*, vol. 14, 2019.

[5] Y. B. Zhang, Q. Y. Zhang, Z. Y. Li et al., "A *k*-anonymous location privacy protection method of dummy based on geographical semantics," *International Journal of Network Security*, vol. 21, no. 6, pp. 937–946, 2019.

[6] T. Wang, Y. Liu, X. Jin et al., "Research on *k*-anonymity-based privacy protection in crowd sensing," *Journal on Communications*, vol. 39, no. 1, pp. 170–178, 2018.

[7] B. Zhou, J. Pei, and W. S. Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data," *ACM SIGKDD Explorations Newsletter*, vol. 10, 2008.

[8] J. Han, J. Yu, Y. U. Hui-Qun et al., "Individuation privacy preservation oriented to sensitive values," *Acta Electronica Sinica*, vol. 38, no. 7, pp. 1723–1728, 2010.

[9] S. Zhang, F. Tian, and Z. Wu, "An adaptive trajectory data publishing algorithm based on differential privacy," *Journal of Shaanxi Normal University(Natural Science Edition)*, vol. 46, no. 5, pp. 9–15, 2018.

[10] P. Xiong, Z. H. U. Tian-Qing, and X.-F. Wang, "A survey on differential privacy and applications," *Chinese Journal of Computers*, vol. 37, no. 1, pp. 101–122, 2014.

[11] Y. Xiao and L. Xiong, "Protecting location with dynamic differential privacy under temporal correlations," 2014.

[12] 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*, Toronto, ON, USA, April 2014.

[13] L. Ou, Z. Qin, Y. Liu et al., "Multi-user location correlation protection with differential privacy," in *Proceedings of the IEEE International Conference on Parallel & Distributed Systems*, Shenzhen, China, December 2017.

[14] M. A. Yin-Fang and L. Zhang, "*KDCK-medoids* dynamic clustering algorithm based on differential privacy," *Computer Science*, vol. 15, 2016.

[15] Q. Yu, *Research on Efficient Index Technology Based on Data Privacy Protection on Cloud platform*, Elsevier, Amsterdam, Netherlands, 2014.