Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 4696455, 12 pages
https://doi.org/10.1155/2021/4696455

Research Article

Differential Privacy Location Protection Method Based on the Markov Model

Hongtao Li,[1] **Yue Wang**,[1] **Feng Guo**,[2] **Jie Wang**,[1] **Bo Wang**,[1] **and Chuankun Wu**[2]

[1]*College of Mathematics and Computer Science, Shanxi Normal University, 041000 Linfen, China*
[2]*School of Information Science and Engineering, Linyi University, Linyi 276002, China*

Correspondence should be addressed to Yue Wang; 951727247@qq.com

Received 4 May 2021; Accepted 20 June 2021; Published 1 July 2021

Academic Editor: Ximeng Liu

Location-based services (LBS) have become an important research area with the rapid development of mobile Internet technology, GPS positioning technology, and the widespread application of smart phones and social networks. LBS can provide convenience and flexibility for the users' daily life, but at the same time, it also brings security risks to the users' privacy. Untrusted or malicious LBS servers can collect users' location data through various ways and disclose it to the third party, thus causing users' privacy leakage. In this paper, a differential privacy location protection method based on the Markov model for user's location privacy is proposed. Firstly, the transition probability matrix between states of the n-order Markov model is used to predict the occurrence state and development trend of events; thereby, the user's location is predicted, and then a location prediction algorithm based on the Markov model (LPAM) is proposed. Secondly, a location protection algorithm based on differential privacy (LPADP) is proposed, in which location privacy tree (LPT) is constructed according to the location data and the difficulty of retrieval, the two nodes with the largest predicted value of LPT are allocated with a reasonable privacy budget, and Laplace noise is added to protect location privacy. Theoretical analysis and experimental results show that the proposed method not only meets the requirements of differential privacy and protects location privacy effectively but also has high data availability and low time complexity.

1. Introduction

In recent years, the rapid development of mobile Internet technology, Internet of things technology, and GPS positioning technology has promoted the rapid development of various smart devices and social networks, making location-based services (LBS) widely applied in people's lives [1–4]. Users can send their identity, location, interests, and other information to the LBS server through the LBS application, so as to query and obtain the required information, such as the nearest shopping center, supermarket, and restaurant. The LBS service provider can also predict the next location of the user according to the current location of the user and provide the user with relevant information of the area before the user enters the next area. For example, in the aspect of traffic, vehicle positioning and prediction can enable users to get a faster and more convenient path. However, while users enjoy the convenience brought by LBS service, it will also lead to the risk of sensitive information leakage. When users query information from the LBS server, they need to send personal identity, location, interests, and other information to the LBS server. If this information is leaked by untrusted or malicious LBS servers, the attackers can not only link the user's identity with location and interests but can also infer more user's private information. Therefore, location privacy protection in LBS is becoming more and more important and has been attached great importance to by relevant fields.

At present, domestic and foreign researchers have conducted a large number of studies on location privacy protection and proposed a variety of solutions to the privacy protection problems in LBS. The dominating location privacy protection technologies include cryptography, k-anonymity, and differential privacy.

Cryptography was proposed by Diffie and Hellman with the idea of public key cryptography in 1976 [5]. The main

idea of location privacy protection technology using cryptography is to encrypt the user's query information. Because the users' query information is not visible to the server, the attacker cannot infer the true data of the user even after obtaining the encrypted data. Although cryptography can effectively protect the privacy of users, it costs a lot in computing and communication and suffers insufficient data availability.

k-anonymity was proposed by Samarati and Sweeney in 1998 [6], which can ensure that each individual record stored in the publication dataset cannot be distinguished from other $k-1$ individuals for sensitive attributes. k-anonymity mechanism requires that the same quasiidentifier must have at least k records; so, the attackers cannot link the records through the quasiidentifier. Although k-anonymity technology can prevent identity disclosure, it cannot prevent attribute disclosure nor can it resist homogeneous attacks and background knowledge attacks.

Differential privacy was proposed by Dwork et al. in 2006 [7], which can protect the privacy information effectively even if the attacker gets the user' background knowledge. Differential privacy has a rigorous statistical model that facilitates the use of mathematical tools and quantitative analysis and proof.

At present, location privacy protection faces great challenges. In this paper, a differential privacy location protection method based on the Markov model is proposed. The main contributions of this paper as follows:

(1) In this paper, the Markov model is used to predict the location information, and the probability transfer matrix between the states of the N-order Markov model is used to predict the state of occurrence of events and their development trend, so as to predict the user's location. Then, a location prediction algorithm based on the Markov model (LPAM) is proposed

(2) A location protection algorithm based on differential privacy (LPADP) is proposed, in which location privacy tree (LPT) is constructed according to the location data and the difficulty of retrieval, the two nodes with the largest predicted value of LPT are allocated a reasonable privacy budget, and Laplace noise is added to protect location privacy

(3) A comprehensive theoretical and experimental analysis has been done between the proposed method and the related works. Results show that our method meets the requirements of differential privacy and protects user location privacy effectively

The rest of this paper is organized as follows: Section 2 introduces the related works; Section 3 introduces the definition, transition probability matrix, system model, and attack model; Section 4 introduces the LPAM algorithm and LPADP proposed in this paper; Section 5 conducts experiments on data availability, privacy protection degree, and algorithm run-time of algorithm proposed in this paper; Section 6 is the conclusion of this paper.

2. Preliminaries

2.1. Definitions

Definition 1. (Markov model) [8, 9]. Let E be the discrete state space of random sequence $\{X(n), n = 0, 1, 2, \cdots\}$. If for any m nonnegative integers $n_1, n_2, \cdots, n_m (0 \le n_1 < n_2 < \cdots < n_m)$ and any natural number k, and any $i_1, i_2, \cdots, i_m$, $j \in E$ satisfies the following conditions:

$$P\{X(n_m + k) \mid X(n_1), X(n_2), \cdots, X(n_m)\} = P\{X(n_m + k) \mid X(n_m)\}. \tag{1}$$

Then, $\{X(n), n = 0, 1, 2, \cdots\}$ is called the one-order Markov model. This equation shows that the state of the next moment only depends on the present moment and has nothing to do with the past moment. This property is the Markov model with no aftereffect.

The n-order Markov model means that the state of the next moment is not only related to the present moment but also related to the past moment; so, the prediction is more comprehensive and effective.

Definition 2. (Neighboring dataset). Let the data set D and D' have the same attribute structure, and the symmetric difference between the D and D' is recorded as $D \Delta D'$, $|D \Delta D'|$ represents the number of symmetry differences. If $|D \Delta D'| = 1$, then D and $D'D'$ are called neighboring dataset (also known as brothers data sets).

Definition 3. (Differential privacy) [10, 11]. There is a random algorithm M and all possible outputs of M are SM. For any two neighboring datasets D and D', if algorithm M satisfies the following conditions:

$$\Pr[\mathrm{M(D)} \in \mathrm{SM}] \le \mathrm{e}^{\varepsilon} \times \Pr\left[\mathrm{M}\left(\mathrm{D}'\right) \in \mathrm{SM}\right], \tag{2}$$

then algorithm M provides ε-differential privacy protection, where parameter ε is called privacy protection budget. The larger the ε is, the higher the data availability is, and the lower the degree of privacy protection is; on the contrary, the lower the data availability is, the higher the degree of privacy protection is.

Definition 4. (Sensitivity). Let d be a positive integer, D is a a set of data sets, and $f : D \to R^d$ is a function. The function sensitivity represented by Δf has the following definition: $\Delta \mathrm{f} = \max \|f(D) - f(D')\|_1$,where $\|.\|_1$ is the Manhattan distance.

Definition 5. (Laplace mechanism) [12, 13]. Given dataset D, there is a function $f : D \longrightarrow R^d$, the sensitivity is Δf, and then the random algorithm $M(D) = f(D) + Y$ provides ε-differential privacy protection, where $Y \sim Lap(\Delta f/\varepsilon)$ is the random noise and obeys the Laplace distribution with the scale parameter $\Delta f/\varepsilon$.

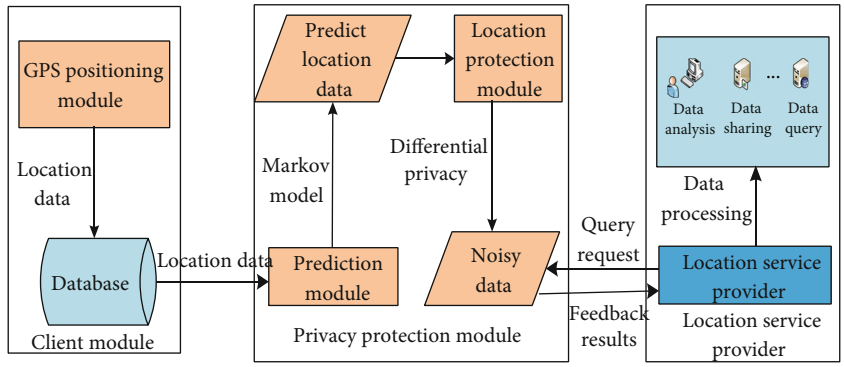


Figure 1: System structure.

2.2. Transition Probability. In this paper, the n-order Markov model is used to predict the location. The basic method of Markov prediction is to use the transition probability matrix between states to predict the occurrence and development probability of events.

The transition probability is derived by using the one-order Markov model [14, 15].

$$P = \frac{N_{ij}}{\sum_{j=1}^{n} N_{ij}}, \tag{3}$$

where N_{ij} is the number of times that location i turns to location j, and P is called one-step transition probability.

The recurrence relation can be obtained by C-K equation.

$$P(n) = PP(n-1) = P(n-1)P, \tag{4}$$

$$P(n) = P^n, \tag{5}$$

where P^n is called the n-step transition probability matrix of the Markov model.

2.3. System Structure and Threat Model. The system structure of this paper is shown in Figure 1, which is mainly composed of client module, privacy protection module, and location service provider module. The client module acquires the user's location information mainly through the GPS positioning module and stores the location data in the database. The privacy protection module is composed of prediction module and location protection module. The prediction module predicts users' location by the Markov model, while the location protection module protects users' location by differential privacy. Location service providers can respond to users' query requests, feedback the query results to users, and use the feedback results for data analysis, data sharing and data query, and other services.

In this paper, a differential privacy location protection method based on the Markov model is proposed to solve the problem of users' location privacy disclosure. The user's location is acquired through the GPS positioning module and stored in the database. In the prediction module, the n-order Markov model is used to predict the user's location, and the LPAM algorithm is proposed. In the location protection module, differential privacy technology is used to protect location data, and LPADP algorithm is proposed. Location service providers can respond to users' query requests, feedback the query results to users, and use the feedback results for data analysis, data sharing, and data query and other services.

Almost all LBS providers collect users' personal data, such as identity, location, and interests. Many LBS providers provide different security guarantees, such as Google, Twitter, and Youtube. Once these LBS providers are attacked, users' privacy information will be leaked. The threat model of this paper is shown in Figure 2. The users' location data is acquired through the smart mobile devices equipped with positioning technology, such as mobile phones, portable computers, and cars, and the obtained location data is uploaded to the database. Then, the location data is transferred to the LBS servers for further intelligent data processing, which allows users to get convenient services from the LBS providers, such as in the aspect of traffic, vehicle positioning, and prediction that can enable users to get a faster and more convenient path; in terms of travel, location positioning and prediction can help users obtain nearby scenic spots and accommodations with better evaluations. The intelligent data processing of LBS servers mainly includes two parts: location prediction and location protection. The attackers can obtain the user's personal data by attacking the user's smart terminals, LBS servers, or location service providers, which will result in the users' privacy being breached.

3. Differential Privacy Location Protection Method Based on the Markov Model

To solve the problem of users' location privacy leakage, a differential privacy location protection method based on the Markov model is proposed in this paper. Firstly, the transition probability matrix between states of the n-order Markov model is used to predict the location information, and LPAM algorithm is proposed. Secondly, LPT is constructed according to the characteristics of location data and the difficulty of retrieval. Finally, the LPADP algorithm is proposed to protect users' location information by using differential privacy technology.

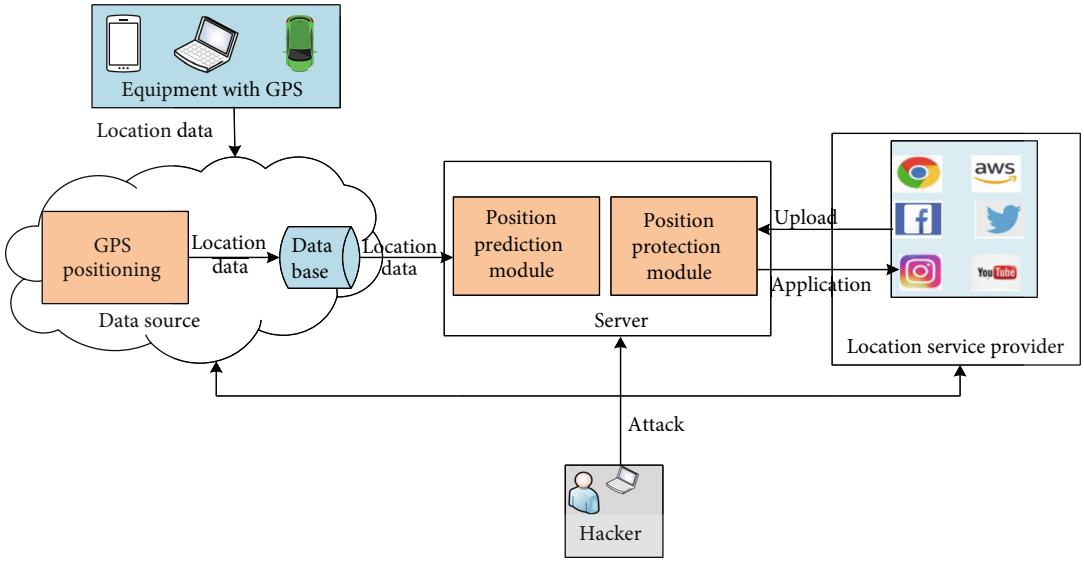


Figure 2: Threat model.

3.1. Location Prediction Algorithm Based on the Markov Model. Location prediction enriches and expands LBS, which is of great significance to LBS. The location prediction methods can be mainly divided into three categories: the location prediction method based on linear or nonlinear mathematical model [16], the location prediction method based on frequent track pattern mining [17], and the location prediction method based on the Markov model.

The location prediction method based on the linear or nonlinear mathematical model is to establish a mathematical model according to the current running speed and time to simulate the trajectory of moving objects, thereby predicting the location. The location prediction method based on frequent trajectory pattern mining is to find the frequent trajectory pattern from the user's historical trajectory and then match the current query trajectory with the frequent trajectory pattern to predict the location. The location prediction method based on the Markov model uses the transition probability matrix between states to predict the state of the event and its development trend, so as to predict the user's location.

The n-order Markov model is used to predict the user's next location in this paper. The basic method of Markov prediction is to predict the occurrence and development trend of events by using the transfer probability matrix between states. The Markov model has the advantages of low time complexity and high prediction accuracy, which not only avoids the problem that the user's moving speed and direction are affected by the road network in the first method but also avoids the problem that the query time is too long in the second method, which affects the prediction efficiency and the redundant noise affects the trajectory prediction accuracy.

Location prediction is fundamentally determined by the current location and historical location. Obviously, the historical location that is closer to the current location has the greatest impact on the next location. Therefore, this paper obtains the predicted value of each location based on the Markov model weighting method.

$$X(t) = a_1 S(t-1)P + a_2 S(t-2)P^2 + \cdots + a_n S(t-n)P^n. \quad (6)$$

In equation (6), t is the time of the next location, and $t-1$ is the time of the current location. $X(t)$ is a $1 \times n$ matrix that represents the predicted value of each location. $S(i), 1 \le i \le n$, is a $1 \times$ n matrix, the value of column i is 1, and the rest is 0. P is an $n \times n$ probability transition matrix. a_1, a_2,… a_k are weights, representing the influence degree of the 1,2,3,...n locations on the next location decision. Based on the Markov model, this paper proposes a location prediction algorithm. The specific content of the algorithm is as follows:

Analysis shows that Algorithm 1 is a location prediction algorithm based on the Markov model, which contains four modules. First, step 1 to step 5, the one-step transition probability matrix M_1 is obtained according to equation (3). The one-step transition probability matrix indicates that the next predicted position is only related to the current position; secondly, step 6 to step 8, the n-step transition probability matrix is obtained according to equation (5). The n-step transition probability matrix indicates that the predicted next position is related to all historical positions and is comprehensive; thirdly, step 9 to step 12, the predicted value of each position is calculated according to equation (6). Because the closer the historical position has the greater influence on the next position, the weight a is set for each historical position; finally, step 14 outputs the predicted probabilities of all positions.

3.2. Location Protection Algorithm Based on Differential Privacy. In the location protection module, this paper proposes the LPADP algorithm. The basic principle is as follows: Firstly, LPT is constructed for all locations predicted by the LPAM algorithm; secondly, the two nodes with the largest prediction value on LPT are protected by adding Laplacian noise. The algorithm is as follows:

The analysis shows that Algorithm 2 is a location protection algorithm based on differential privacy, which contains three modules. The main purpose of Algorithm 2 is to add

Input: $N = \{N_{ij}\}_{n*n}$; // degree transition matrix
$S = \{S_i\}$; // location at time k-1
now; // current location
$X = \{X_i\}_{1*n}$; // estimate each location
$a = \{a_i\}_{1*n}$; // weight array
Output: result // output all predicted locations
1. FOREACH $N_i \in N$
2. sum = cumulate $N_{ij} \in N_i$; //cumulate is an accumulation process
3. FOREACH $N_{ij} \in N_i$
4. $P_{ij} = N_{ij}/sum$;
5. $M_1 = P$ // one step transition probability matrix P is obtained
6. FOR i=2 to n Do
7. M_i = matrixMul(M_{i-1}, P); // calculate n-step transition probability matrix
8. ENDFOR
9. setZero(X); // clear X and calculate the estimate
10. FOR i=1 to n Do
11. Tepmatrix = matrixMul(S_i, M_i, a_i); // multiplication of weight and matrix
12. X = matrixAdd(X, Tepmatrix); // calculate X
13. ENDFOR
14. result = put($X_1, X_2, \cdots, X_n$); // output all predicted locations
15. RETURN result;

Algorithm 1: Location prediction based on the Markov model (LPAM).

Input: $X = (X_i)_{1*n}$; // location from Algorithm 2
Output: The two locations with the largest prediction probability are protected by adding noise
1. Constructing LPT;
2. void fun(int $*X$, int $*X_1$, int $*X_2$) // select the two nodes with the largest prediction probability on LPT(X_1,X_2)
3. {
4. int i;
5. $*$max = $X[0]$;
6. for(i=1; $i < strlen(X)$; i++)
7. if($*$max < $*(X + i)$) $*X_1 = *(X + i)$; // select the nodes with the highest prediction probability X_1
8. $*X_2 = X[0]$;
9. for(i=1; $i < strlen(X)$; i++)
10. {
11. if($*X_2 < *(X + i)$ && $*X_2 < *X_1$)
12. $*X_2 = *(X + i)$; // select the next largest value node X_2
13. }
14. result = put(X_1, X_2);//output X_1, X_2
15. }
16. $\varepsilon = \varepsilon_1 + \varepsilon_2$; //$\varepsilon_1 < \varepsilon_2$
17. $X_i' = X_i + Lap(\varepsilon_i)$ // Laplacian noise is added to the two location nodes with the largest prediction value

Algorithm 2: Location protection based on Differential privacy (LPADP).

Laplacian noise to the two position nodes with the largest predicted value for protection. Firstly, the first step is to construct LPT for all positions predicted by Algorithm 1, which LPT is constructed according to the location data and the difficulty of retrieval; secondly, the function of the second step to the fifteenth step is to traverse all positions on the LPT to obtain the node with the largest predicted value and the node with the second largest predicted value. There are two loop functions in the second step to the fifteenth step. Among them, the first position node on the LPT is defaulted to the maximum value node, and then the nodes on the LPT are traversed in turn, the function of the first loop is to obtain the node X_1 with the largest predicted value (that is, the fourth to seventh steps), and the function of the second loop is to obtain the node X_2 with the second largest predicted value (that is, the eighth to thirteenth steps); finally, the sixteenth to seventeenth steps are to protect the two locations X_1 and X_2. The sixteenth step is to allocate a reasonable privacy budget to the two locations, and the seventeenth step is to add Laplace noise to X_1 and X_2 according to the privacy budget, so as to protect the two positions.

3.3. Algorithm Analysis

3.3.1. Safety Analysis. The basic principle of differential privacy technology is as follows: when the user submits a query request to the data provider, if the user directly publishes the accurate query results, it may lead to privacy leakage, because the attacker may use the query result to deduct private information. In order to avoid this problem, the differential privacy technology requires a middleware to be extracted from the database, and a specially designed random algorithm is used to inject an appropriate amount of noise into the middleware to obtain a noisy middleware; then, a noisy query result is derived from the noisy middleware and returned to the user. In this way, even if the attacker can deduce the noisy middleware from the noisy result, it is impossible for him to infer the noiseless middleware accurately, let alone infer the original database, so as to achieve the purpose of protecting the user's privacy.

This paper uses differential privacy technology to protect the user's location privacy. The main reason is that differential privacy technology has three major advantages: (1) differential privacy strictly defines the background knowledge of the attacker: except for a certain record, the attacker knows all the information in the original data. Such an attacker is almost the most powerful. In this case, differential privacy can still effectively protect private information; (2) differential privacy has a rigorous statistical model, which greatly facilitates the use of mathematical tools and quantitative analysis and verification; and (3) differential privacy does not require special attack assumptions, does not care about the background knowledge of the attacker, and quantitatively analyzes the risk of privacy leakage.

The main implementation mechanism of differential privacy technology is to add random noise to input or output to protect the privacy of users', such as Laplace mechanism, Gaussian mechanism, and Exponential mechanism. In this paper, Laplacian mechanism is used to protect the user's location by adding Laplacian noise.

Laplacian noise is essentially a group of random values satisfying the Laplacian distribution, and the basic principle is to add noise that obeys Lap (b) to the original data and statistical results, so that the query results after adding the noise meet the differential privacy constraint effect. Laplacian noise is added in the LPADP algorithm, which conforms to ε-differential privacy. The proof process is as follows:

It can be known from the probability density function of the laplace mechanism:

$$\begin{aligned}\frac{P_x(z)}{P_y(z)} &= \prod_{i=1}^{k} \frac{e^{\left(-\varepsilon\left|f(x)_i - z_i\right|/\Delta f\right)}}{e^{\left(-\varepsilon\left|f(y)_i - z_i\right|/\Delta f\right)}} = \prod_{i=1}^{k} e^{\frac{\varepsilon\left(\left|f(y)_i - z_i\right| - \left|f(x)_i - z_i\right|\right)}{\Delta f}} \\ &\le \prod_{i=1}^{k} e^{\frac{\varepsilon\left(\left|f(x)_i - f(y)_i\right|\right)}{\Delta f}} \le e^{\frac{\varepsilon\|f(x) - f(y)\|_1}{\Delta f}} = e^{\epsilon}.\end{aligned} \tag{7}$$

According to the definition of differential privacy, the LPADP algorithm proposed in this paper satisfies ε-differential privacy.

3.3.2. Complexity Analysis. Assuming that the location data table contains n pieces of records data. The privacy protection module in Figure 1 mainly contains two modules: prediction module and location protection module. So, the complexity of the algorithm in this paper mainly includes two aspects: the time complexity of the LPAM algorithm in the prediction module and the LPADP algorithm in the location protection module. The LPAM algorithm mainly uses the n-order Markov model to predict the position.

The realization of the LPAM algorithm mainly includes three parts: first, calculate the one-step transition probability matrix according to formula (3), and its time complexity is $O(n^2)$, reflected in the first to fifth steps of Algorithm 1; secondly, calculate the n-step transition probability matrix according to formula (5), and its time complexity is $O(n)$, reflected in the sixth to eighth steps of Algorithm 1; finally, calculate and output the predicted value of each position according to formula (6), and its time complexity is $O(n)$, reflected in the tenth to fourteenth steps of Algorithm 1.

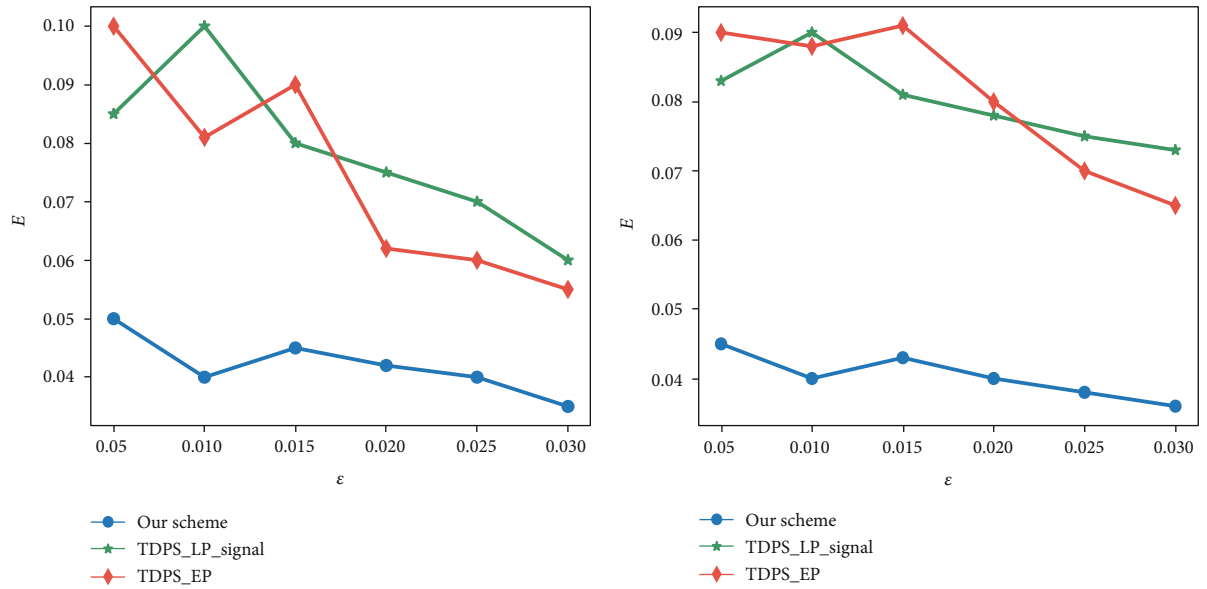
The LPADP algorithm mainly allocates a reasonable privacy budget to the two locations with the larger predicted value on the LPT and then adds Laplacian noise to protect the location privacy. The realization of the LPADP algorithm mainly includes three parts: first, construct LPT for all positions predicted by Algorithm 1, and its time complexity is $O(n)$, reflected in the first step of Algorithm 2; secondly, traverse all the position nodes on the LPT and then select the two nodes with the largest predicted value, and the time complexity is $O(n)$, reflected in the second to the fifteenth steps of Algorithm 2; finally, a reasonable privacy budget is allocated to the two nodes with the largest predicted value, Laplacian noise is added for protection, and the time complexity is O (1), reflected in the sixteenth to seventeenth steps of Algorithm 2. In general, the time complexity required in this article is

$$O\left(n^2\right) + O(n) + O(n) + O(n) + O(n) + O(1) \approx O\left(n^2\right). \tag{8}$$

4. Experimental Results and Analysis

4.1. Environment Configuration. In order to test the performance of the location privacy protection method proposed in this paper, the algorithm has been fully experimented in terms of data availability, privacy protection degree, and algorithm running time. The experiment is implemented using Python, and the data sets are Gowalla data set and Geolife data set [18, 19]. The experimental environment of this article is PyCharm. The hardware environment is 2.60GHz i7 CPU, 8.00RAM, Win10 system 64-bit.

4.2. Data Availability Analysis. For the same query function Q, the similarity of the output query results before and after the noise is added to the data that reflects the influence of the privacy protection algorithm on the availability of the data. Let $G(Q)$ be the query result of the data before adding noise, and $G'(Q)$ be the query result of the data after adding noise, and then the degree of approximation S_Q can be



(a) Gowalla (b) Geolife

Figure 3: The effect of ε on data availability.

defined as the first-order normal form distance between the two output results $S_Q = \|G(Q) - G'(Q)\|_1$.

For continuous query $Q_i \in \{Q_1, Q_2, \cdots, Q_k\}$, the availability of data published by location services is defined as

$$E = \frac{1}{Q_i} \sum \left(\frac{S_{Q_i}}{e^{\varepsilon/2}} \right). \tag{9}$$

Comparing the method proposed in this paper with TDPS_LP_Signal and TDPS_EP [20] in terms of data availability, the results are shown in Figure 3. The X-axis represents the value of ε, and the Y-axis represents the data availability. The E value of the three algorithms will decrease with the increase of ε, because the larger the ε, the smaller the noise addition and the better the data availability. When $\varepsilon > 0.015$, the data availability of the method proposed in this paper tends to be stable. Therefore, the method proposed in this paper has better data availability compared with TDPS_LP_Signal and TDPS_EP. The data availability of the TDPS_LP_Signal algorithm is between the algorithm proposed in this paper and TDPS_EP, and the data availability of TDPS_EP is relatively poor.

Comparing the Markov model with the trajectory mining model and linear or nonlinear mathematical model in data availability, the results are shown in Figure 4. The X-axis represents the number of historical locations, and the Y-axis represents data availability. The E value of the three algorithms will decrease with the increase in the number of historical locations, because the increase in the number of historical locations, the more accurate the prediction and the better the data availability. The n-order Markov model is more accurate and comprehensive in location prediction; so, it has better data availability. The trajectory mining model is between the Markov model and linear or nonlinear mathematical model in terms of data availability, and the data availability of the linear or nonlinear mathematical model is poor.

4.3. Analysis of the Degree of Privacy Protection. Comparing the method proposed in this paper with TDPS_LP_Signal and TDPS_EP in terms of privacy protection degree, the result is shown in Figure 5. The X-axis represents the value of ε, and the Y-axis represents the degree of privacy protection. The degree of privacy protection of the three algorithms will decrease with the increase of ε, because the larger the ε, the smaller the noise addition and the worse the degree of privacy protection. The algorithm proposed in this paper uses differential privacy technology to protect the location and has better security. The degree of privacy protection of the TDPS_LP_Signal algorithm is between the algorithm proposed in this paper and the TDPS_EP algorithm, and the degree of privacy protection of the TDPS_EP algorithm is relatively low.

Comparing the Markov model with the trajectory mining model and linear or nonlinear mathematical model in the degree of privacy protection, the result is shown in Figure 6. The X-axis represents the number of historical locations, and the Y-axis represents the degree of privacy protection. The degree of privacy protection of the three algorithms will increase with the increase in the number of historical locations, because the increase in the number of historical locations, the more accurate the prediction and the better the degree of privacy protection. The n-order Markov model is more accurate and comprehensive in location prediction and has better security. The privacy protection degree of the trajectory mining mode algorithm is between the Markov model and the linear and nonlinear mathematical model. The privacy protection degree of the linear or nonlinear mathematical model algorithm is relatively low.

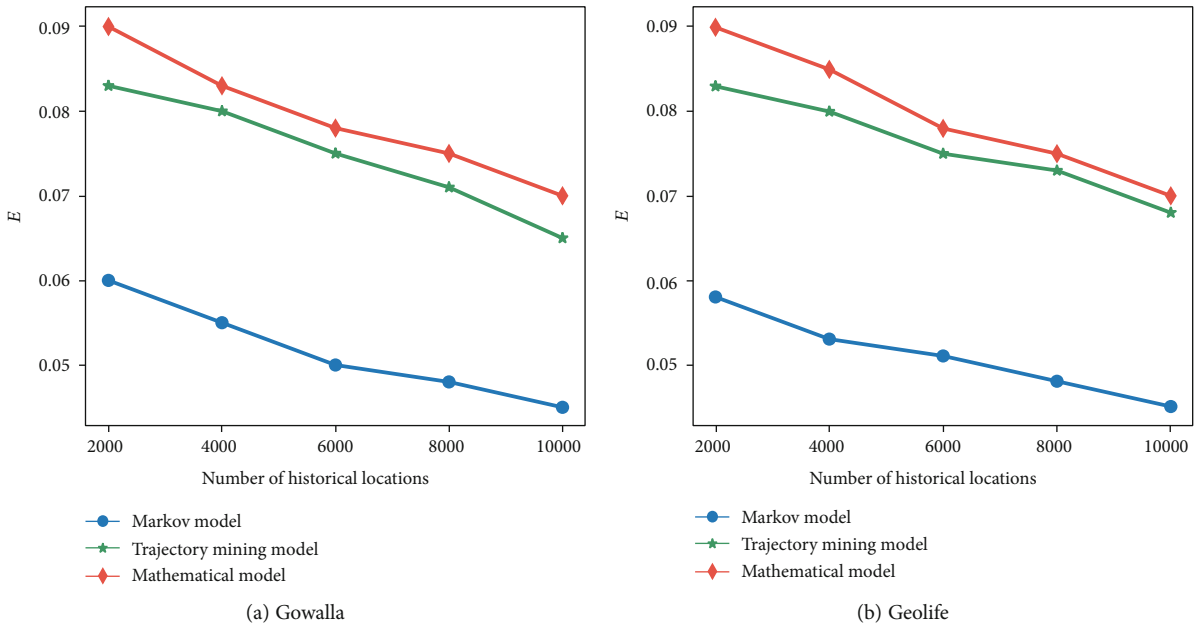


Figure 4: The impact of the number of historical locations on data availability.

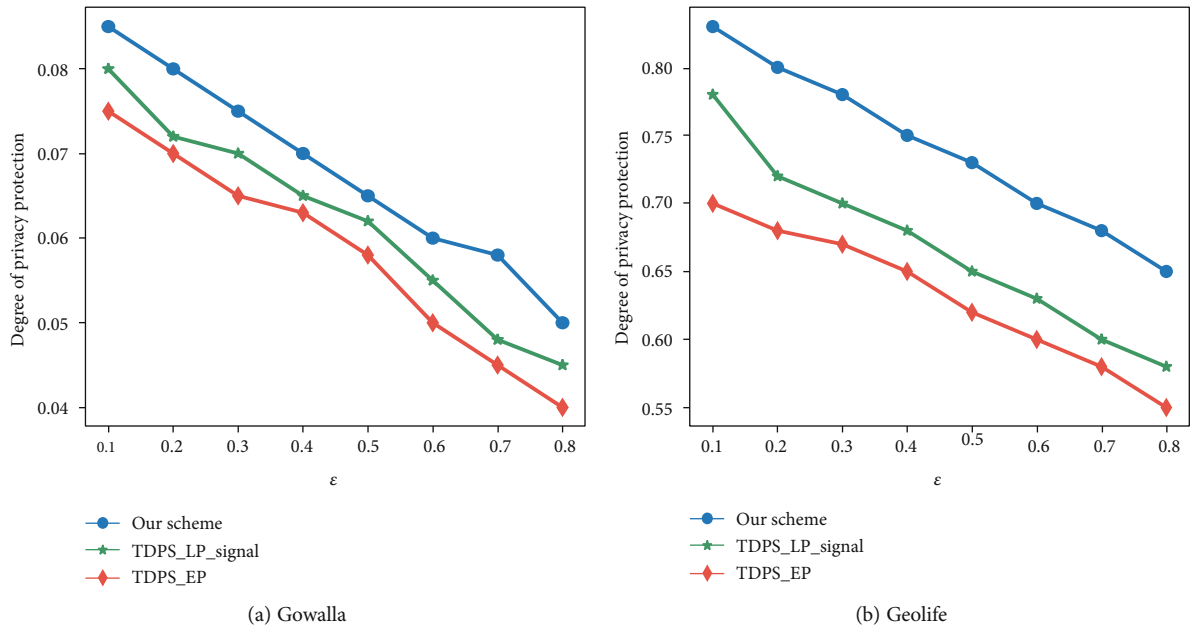
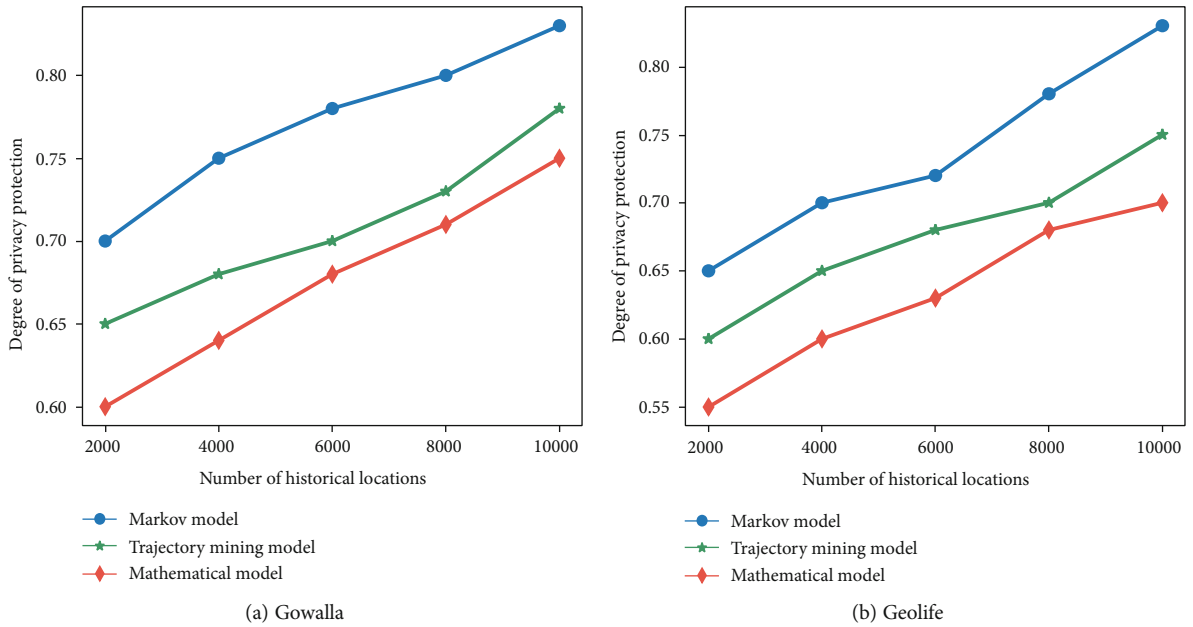


Figure 5: The effect of ε on the degree of privacy protection.

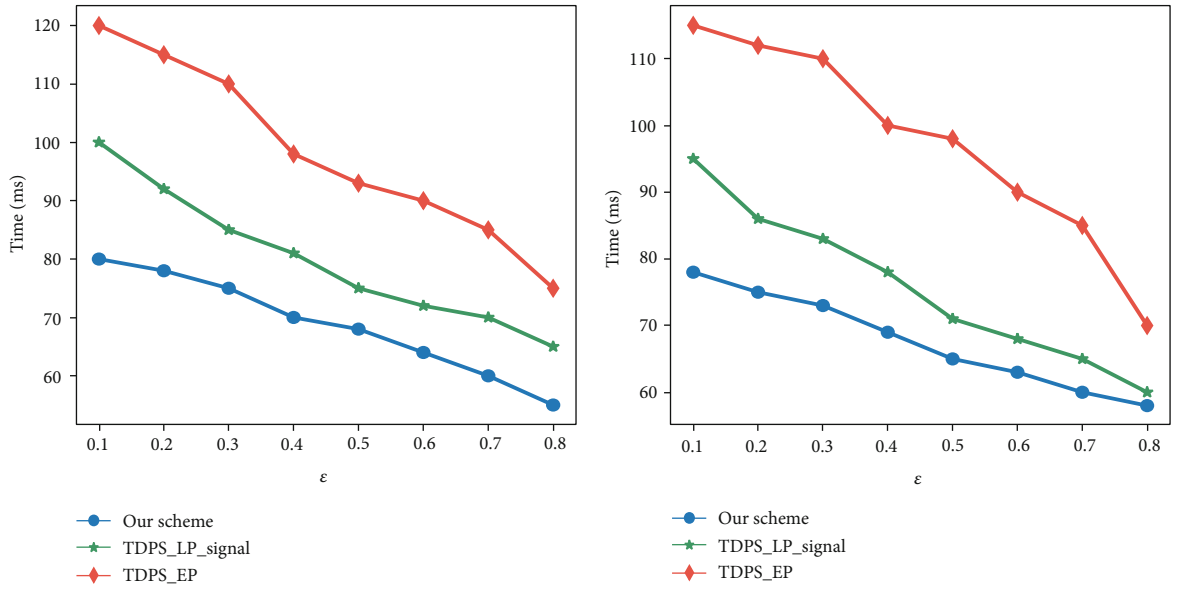
4.4. Analysis of Algorithm Running Time. Comparing the method proposed in this paper with TDPS_LP_Signal and TDPS_EP in terms of algorithm running time, the result is shown in Figure 7. The X-axis represents the value of ε, and the Y-axis represents the running time of the algorithm. The running time of the three algorithms will decrease with the increase of ε, because the larger the ε, the smaller the noise addition and the shorter the running time. The method proposed in this paper only protects the two locations with the largest predicted value and has less algorithm running time. The running time of the TDPS_LP_Signal algorithm is between the algorithm proposed in this paper and the TDPS_EP algorithm, and the TDPS_EP algorithm requires relatively more time.

Comparing the Markov model with the trajectory mining model and linear or nonlinear mathematical model in terms of algorithm running time, the results are shown in Figure 8. The X-axis represents the number of historical locations, and the Y-axis represents the running time of the algorithm. The running time of the three algorithms will increase as the number of historical locations increases, because the number of historical locations increases, the prediction time increases, thereby increasing the running time. Because the Markov model has the advantage of low time complexity, it



(a) Gowalla (b) Geolife

Figure 6: The influence of the number of historical locations on the degree of privacy protection.



(a) Gowalla (b) Geolife

Figure 7: The effect of ε on the running time of the algorithm.

has less algorithm running time. The running time of the trajectory mining pattern algorithm is between the Markov model and linear and nonlinear mathematical model, and linear and nonlinear mathematical model algorithm takes more time.

5. Related Work

As LBS has privacy that becomes the focus of research, more and more scholars have paid close attention to LBS privacy protection methods. At present, the main methods of location privacy protection include cryptography, k-anonymity, and differential privacy.

Cryptography is a privacy protection method based on encryption and signature, which realizes privacy protection by encrypting users' information [21–23]. Liang et al. proposed a privacy protection method based on POI query inthe road network environment by combining Hilbert curve with anonymous technology, which effectively avoided inference attack against location information [24].While it is known that unconditionally secure position-based cryptography is impossible both in the classical and the quantum setting, it

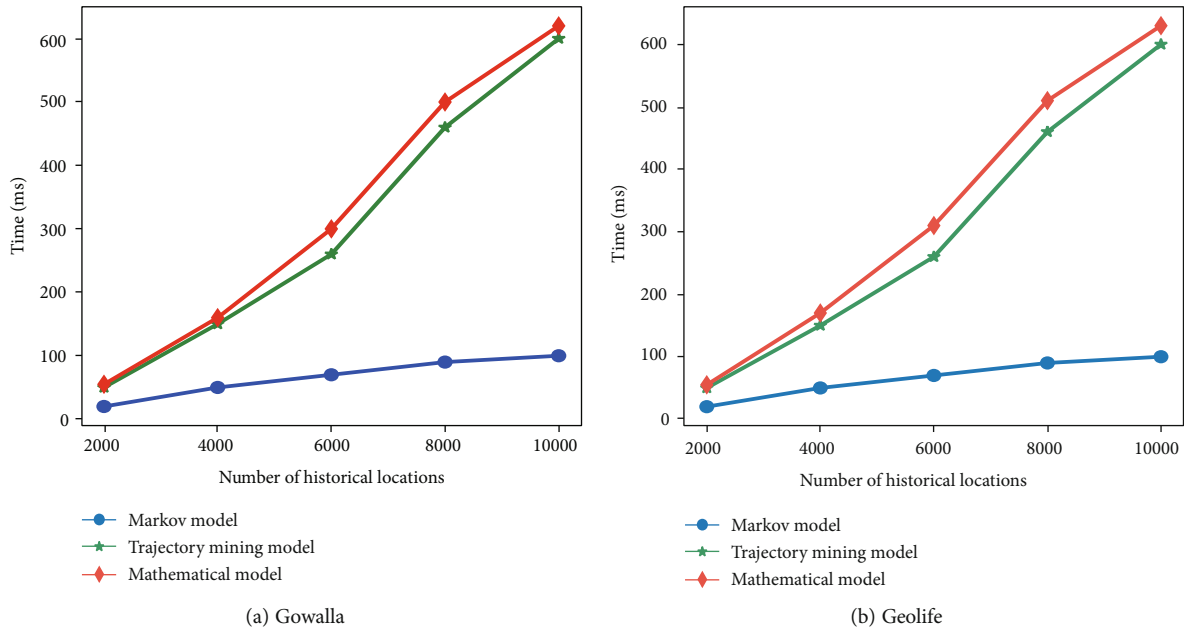


FIGURE 8: The effect of the number of historical locations on the running time of the algorithm.

has been shown that some quantum protocols for position verification are secure against attackers which share a quantum state of bounded dimension. Bluhm et al. considered the security of the qubit routing protocol. The protocol has the advantage that an honest prover only has to manipulate a single qubit and a classical string of length $2n$ and shows that the protocol is secure if each of the attackers holds at most $n/2 - 3$ qubits [25]. However, cryptography is difficult to implement because of huge computing and communication costs.

k-anonymity requires that the same quasiidentifier must have at least k records, and each individual record cannot be distinguished from other k-1 individuals for sensitive attributes; so, the attackers cannot link the records through the quasiidentifier [26–28]. In reference [29], the user's real location was replaced by the anonymous users' area; so, the attacker could not identify the user's real location. In reference [30], the users used historical information to process real information anonymously, so as to protect users' location privacy. In reference [31], the users cooperated with each other, shared part of the location information, and formed an anonymous space to achieve the effect of k-anonymity. Mingyan et al. [32] proposed a location anonymity algorithm based on the mobile P2P structure, which avoided the risk of information leakage caused by single point failure. Xingyou et al. [33] selected the location anonymous set in the grid that published the request according to the real service request data and sent the location anonymous set to the server instead of the user's real location. Although k-anonymity technology can prevent the disclosure of identity, it cannot resist homogeneous attacks and background knowledge attacks.

Differential privacy can protect privacy effectively and has a rigorous statistical model [34–36]. Zhiqiang et al. [37] proposed a location data acquisition scheme based on local differential privacy, which used the random response mechanism to obtain location data, and the data collector used direct statistics and expectation maximum method to analyze the location data to ensure that the normal analysis can be carried out. In order to solve the problem of privacy leakage in crowdsourcing, Zheng et al. [38] proposed a crowdsourcing location data acquisition scheme that satisfied the localized differential privacy. In this scheme, the road network space was divided into Voronoi diagram, and a method of spatial range query on disturbed data set was designed. Fuzzy C-means clustering algorithm is one of the typical clustering algorithms in data mining applications. However, due to the sensitive information in the dataset, there is a risk of user privacy being leaked during the clustering process. Zhang et al. [39] aimed at the problem that the algorithm accuracy is reduced by randomly initializing the membership matrix of fuzzy C-means; in this paper, the maximum distance method is firstly used to determine the initial center point. Then, the Gaussian value of the cluster center point is used to calculate the privacy budget allocation ratio. Additionally, Laplace noise is added to complete differential privacy protection. Wei et al. [40] proposed a differential privacy-based location protection (DPLP) scheme, and DPLP splits the exact locations of both workers and tasks into noisy multilevel grids by using adaptive three-level grid decomposition (ATGD) algorithm and DP-based adaptive complete pyramid grid (DPACPG) algorithm, respectively, thereby considering the grid granularity and location privacy. Furthermore, DPLP adopts an optimal greedy algorithm to calculate a geocast region around the task grid, which achieves the trade-off between acceptance rate and system overhead, which protects the location privacy of both workers and tasks, and achieves task allocation with high data utility.

In view of the problem of location privacy protection, this paper uses differential privacy technology to protect the location privacy of users'. Differential privacy can not only resist the background knowledge attack and homogeneous

attacks but also can effectively protect the user's privacy when adding or deleting a record without affecting the query result.

6. Conclusions

The continuous use of LBS will expose the user's location information, which results in the disclosure of user's privacy. In order to solve issues of user privacy disclosure in LBS, a differential privacy location protection method based on the Markov model is proposed in this paper. Experiments show that this method can protect location privacy effectively and has high data availability and low time complexity. In the future research, the research mainly focuses on two aspects. On the one hand, the location prediction of the Markov model does not consider the situation of new users; so, the future research direction is to predict the location of new users and protect the predicted location information. On the other hand, the Markov model predicts and protects the position, which realizes the direct protection of the position, but ignores the spatiotemporal correlation between the predicted positions. Therefore, the future research direction is to protect the position indirectly according to the spatiotemporal correlation between the predicted positions.

Data Availability

All data, models, and codes generated or used during the study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Key Research and Development Project of Shandong Province under grant no. 2019JZZY010134, the Natural Science Foundation of Shanxi Province under grant no. 201901D111280, and the Scientific and Technological Innovation Project in Colleges and Universities of Shanxi Province under grant no. 2019L0459.

References

[1] H. Huang, G. Gartner, J. M. Krisp, M. Raubal, and N. van de Weghe, "Location based services: ongoing evolution and research agenda," *Journal of Location Based Services*, vol. 12, no. 2, pp. 63–93, 2018.

[2] M. Yu, G. Fan, H. Yu, and L. Chen, "Location-based and time-aware service recommendation in mobile edge computing," *International Journal of Parallel Programming*, vol. 2, 2021.

[3] A. Aloui, O. Kazar, S. Bourekkache, and F. Omary, "An efficient approach for privacy-preserving of the client's location and query in M-business supplying LBS services," *International Journal of Wireless Information Networks*, vol. 27, no. 3, pp. 433–454, 2020.

[4] A. Sz and C. B. Xin, "Multiple-user closest keyword-set querying in road networks," *Information Sciences*, vol. 509, pp. 133–149, 2020.

[5] Q. Zeng, X. Han, and Y. M. Cao, "Integrated public key encryption and public key encryption with keyword search," *Computer and Modernization*, vol. 284, no. 4, pp. 107–111, 2019.

[6] P. Samarati and L. Sweeney, "Generalizing data to provide anonymity when disclosing information," in *Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems - PODS '98*, pp. 188–202, Washington, 1998.

[7] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography. TCC 2006*, S. Halevi and T. Rabin, Eds., vol. 3876 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2006.

[8] E. Akinola and S. S. Daodu, "Location prediction in the Long term evolution network using ST-RNN and Markov model," *International Journal of Computer Applications*, vol. 176, no. 30, pp. 14–17, 2020.

[9] A. Rahimifar, "Predicting the energy consumption in software defined wireless sensor networks: a probabilistic Markov model approach," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020.

[10] C. Xia, J. Hua, W. Tong, and S. Zhong, "Distributed *K*-Means clustering guaranteeing local differential privacy," *Computers & Security*, vol. 90, p. 101699, 2020.

[11] Z.-q. Gao, X.-l. Cui, S. Zhou, and C. Yuan, "Local differential privacy protection and its application," *Journal of Computer Engineering and Science*, vol. 40, no. 6, pp. 1029–1036, 2018.

[12] J. Sharma, D. Kim, A. Lee, and D. Seo, "On differential privacy-based framework for enhancing user data privacy in mobile edge computing environment," *IEEE Access*, vol. 9, pp. 38107–38118, 2021.

[13] S. Chen, A. Fu, S. Yu, H. Ke, and M. Su, "DP-QIC: a differential privacy scheme based on quasi-identifier classification for big data publication," *Soft Computing*, vol. 25, pp. 7325–7339, 2021.

[14] M. N. Cakir, M. Saleemi, and K. H. Zimmermann, "On the theory of stochastic automata," 2021, https://arxiv.org/abs/2103.14423.

[15] A. Niessl, A. Allignol, C. Mueller, and J. Beyersmann, "Estimating state occupation and transition probabilities in non-Markov multi-state models subject to both random left-truncation and right-censoring," 2020, https://arxiv.org/abs/2004.06514.

[16] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 383–398, 2019.

[17] X. Pan, Q. Zhao, and P. Zhao, "Frequent trajectory of pattern mining with spatio-temporal attribute and relationship label," *Computer Engineering and Applications*, vol. 55, no. 10, pp. 83–89, 2019.

[18] K. Cao, Q. Sun, H. Liu, Y. Liu, G. Meng, and J. Guo, "Social space keyword query based on semantic trajectory," *Neurocomputing*, vol. 428, pp. 340–351, 2020.

[19] H. Luo, H. Zhang, S. Long, and Y. Lin, "Enhancing frequent location privacy-preserving strategy based on geo-Indistinguishability," *Multimedia Tools and Applications*, vol. 80, no. 14, pp. 21823–21841, 2021.

[20] H. Kang, S. Zhang, and Q. Jia, "A method for time-series location data publication based on differential privacy," *Wuhan*

University Journal of Natural Sciences, vol. 24, no. 2, pp. 107–115, 2019.

[21] G. Xu, H. Li, Y. Dai, K. Yang, and X. Lin, "Enabling efficient and geometric range query with access control over encrypted spatial data," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 4, pp. 870–885, 2019.

[22] M. Etemad, A. Küpçü, C. Papamanthou, and D. Evans, "Effficient dynamic searchable encryption with forward privacy," *Proceedings on Privacy Enhancing Technologies*, vol. 2018, no. 1, pp. 5–20, 2018.

[23] C. Cui, F. Li, T. Li, J. Yu, R. Ge, and H. Liu, "Research on direct anonymous attestation mechanism in enterprise information management," *Enterprise Information Systems*, vol. 15, no. 4, pp. 513–529, 2021.

[24] H. C. Liang, B. Wang, N. N. Cui, K. Yang, and X. C. Yang, "Privacy preserving method for point-of-interest query on road network," *Journal of Software*, vol. 29, no. 3, pp. 703–720, 2018.

[25] A. Bluhm, M. Christandl, and F. Speelman, "Position-based cryptography:single-qubit protocol secure against multi-qubit attacks," 2021, https://arxiv.org/abs/2104.06301.

[26] A. T. Truong, "Privacy preserving spatio-Temporal databases based on k-anonymity," *Science & Technology Development Journal - Engineering and Technology*, vol. 3, no. SI1, pp. - SI82–SI94, 2020.

[27] B. S. Kumar, T. Daniya, N. Sathya et al., "Investigation on privacy preserving using K-anonymity techniques," in *2020 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2020.

[28] T. K. Esmeel, M. M. Hasan, M. N. Kabir, and A. Firdaus, "Balancing data utility versus information loss in data-privacy protection using k-anonymity," in *2020 IEEE 8th Conference on Systems, Process and Control (ICSPC)*, Melaka, Malaysia, 2020.

[29] S. Zhang, X. Li, Z. Tan, T. Peng, and G. Wang, "A caching and spatial K-anonymity driven privacy enhancement scheme in continuous location-based services," *Future Generation Computer Systems*, vol. 94, pp. 40–50, 2019.

[30] Z. Wu, G. Li, S. Shen, X. Lian, E. Chen, and G. Xu, "Constructing dummy query sequences to protect location privacy and query privacy in location-based services," *World Wide Web*, vol. 24, pp. 25–49, 2020.

[31] D.-d. Wu and L. Xin, "Location anonymous algorithm based on user collaboration under distributed structure," *Computer Science*, vol. 46, no. 4, pp. 158–163, 2019.

[32] M.-y. Xu, Z. Hua, J. Xinsheng, and S. Juan, "Distribution-perceptive-based spatial cloaking algorithm for location privacy in mobile peer-to-peer environments," *Journal of Software*, vol. 29, no. 7, pp. 1852–1862, 2018.

[33] X.-Y. Xia, Z. -H. Bai, J. Li, and R. -Y. Yu, "A location cloaking algorithm based on dummy and Stackelberg game," *Chinese Journal of Computers*, vol. 42, no. 10, pp. 2216–2232, 2019.

[34] X. Zhao, D. Pi, and J. Chen, "Novel trajectory privacy-preserving method based on prefix tree using differential privacy," *Knowledge-Based Systems*, vol. 198, p. 105940, 2020.

[35] X. Niu, H. Huang, and Y. Li, "A real-time data collection mechanism with trajectory privacy in mobile crowd-sensing," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2114–2118, 2020.

[36] F. O. Olowononi, D. B. Rawat, and C. Liu, "Federated learning with differential privacy for resilient vehicular cyber physical systems," in *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*, Las Vegas, NV, USA, 2021.

[37] G. A. Zhiqiang, C. U. Xiaolong, D. U. Bo, Z. H. Sha, Y. U. Chen, and L. I. Ai, "Collection scheme of location data based on local differential privacy," *Journal of Tsinghua University: Science and Technology*, vol. 59, no. 1, pp. 23–27, 2019.

[38] Z. Huo, K. Zhang, P. He, and Y. Wu, "Crowdsourcing location data collection for local differential privacy," *Journal of Computer Applications*, vol. 39, no. 3, pp. 763–768, 2019.

[39] Y. Zhang and J. Han, "Differential privacy fuzzy C-means clustering algorithm based on gaussian kernel function," *PLoS One*, vol. 16, no. 3, article e0248737, 2021.

[40] J. Wei, Y. Lin, X. Yao, and J. Zhang, "Differential privacy-based location protection in spatial crowdsourcing," *IEEE Transactions on Services Computing*, 2019.