Research Article

Wenfeng Liu*, Juanjuan Wu*, and Zhong Xi

Privacy protection methods of location services in big data

https://doi.org/10.1515/comp-2022-0250
received May 16, 2022; accepted July 22, 2022

Abstract: The rapid development of mobile communication technology not only brings convenience and fun to our life, but also brings a series of problems such as privacy disclosure. Therefore, it is very necessary to study the privacy protection method based on location service to strengthen the security of location privacy. The purpose of this work is to improve the security of location privacy and prevent the disclosure of user privacy by studying the characteristics of location services and privacy protection methods. This article first describes the characteristics of the important location privacy protection law, and then studies the structural characteristics and operation process of the location privacy protection law. This work evaluates the advantages and disadvantages of different methods, and finally compares the performance of several privacy protection algorithms through experimental analysis. Through the research of hiding space method, two-level cache method based on user grid, differential privacy protection method and experimental analysis of the algorithm, an effective privacy protection algorithm can be obtained. It can better protect the location privacy of users. For example, dual-active in the hidden space algorithm has the best privacy protection performance. Compared with other algorithms, the success rate of generating hidden space is increased by more than 10%, and the time of generating hidden space is shortened by about a quarter. The algorithm It has certain practical value and significance for use in the privacy protection of users.

Keywords: location service, privacy protection, hidden space, third party anonymous server

1 Introduction

With the progress of time and the further application of information network in life, location-based services are more and more widely used in user groups. For example, mobile communication, Internet, spatial positioning, and location information all need to provide corresponding location-based services. In recent years, mobile technologies such as geographic mapping, satellite navigation, and communication network are developing rapidly. Its services, including personal positioning, travel navigation, and big data push, are also more accurate and efficient, which undoubtedly improves people's quality of life to a new level. Through the location-based service function, users can keep abreast of information from all over the world and understand what is happening in a certain place. It can also share its location information, find friends around at any time, and enhance the friendship between friends. However, when people share location, their personal privacy will inevitably be exposed, and it may be used by illegal people to produce unpredictable risks. Therefore, how to create a safe and efficient privacy protection scheme so that users can protect their privacy while enjoying services is the research direction of many scholars.

Modern information technology is constantly changing our lives, and the use of the Internet has also changed from "fixed terminals" such as computers to "mobile terminals" such as smartphones and tablets. At present, location-based services provide users with convenient functions such as navigation, location sharing, advertising push, and nearby search. Relevant research shows that more than 90% of users will use the location function to browse location information. However, when users want to enjoy location-based services, they must transmit their real-time location to the service provider. This will greatly increase the possibility of personal

* **Corresponding author: Wenfeng Liu,** Student Affairs Office, Hunan College of Foreign Studies, Changsha, 410203, Hunan, China, e-mail: 472093717@qq.com
* **Corresponding author: Juanjuan Wu,** School of Humanities and Arts, Hunan Institute of Transportation Engineering, Hengyang, 421000, Hunan, China, e-mail: xmlw1230@163.com
Zhong Xi: College of Western Languages, Hunan College of Foreign Languages, Changsha, 410203, Hunan, China, e-mail: 510242981@qq.com

privacy disclosure and being used by others. Through the research on location services, users can reduce the leakage of their location information to achieve the purpose of protecting personal privacy. At the same time, the privacy protection method based on location service can effectively reduce information crime, which is of great significance to maintain social stability and promote social development.

This article can not only provide new ideas for protecting user privacy, but also provide a broader research direction for the application of location services in big data. Several privacy protection algorithms can be obtained through this process. (1) Through the analysis and comparison of anonymous space method, third-party anonymous server method, and other algorithms, we can find the advantages and disadvantages. (2) This article puts forward a theoretical standard to measure the quality of a X region by two indicators: privacy protection ability and location service quality assurance.

2 Relevant work

Many scholars have paid attention to the research on the privacy protection methods of location services in big data. Ruan O proposed a new location sharing protocol. He reduced communication and computing costs and protected users' identity privacy by correctly using symmetric and asymmetric encryption technologies. He analyzed and studied the establishment of virtual identity and built a model to prevent the location server from inferring the user's activity trajectory by updating the virtual identity in time [1]. Xue et al. proposed a location privacy protection scheme based on K-anonymity, where GeoHash encoding model and Voronoi diagram are used as grid area operation models. He used the client server to create user model to protect the location data [2] of the user on the client side and the server side, respectively. Lee et al. introduced an anonymous algorithm for location privacy of location-based service (LBS) models. He divided each algorithm into location K anonymity and location differential privacy, and compared and analyzed each algorithm [3]. Wu et al. proposed a location privacy protection system for LBS, which protects the query range related to the location query sequence by constructing "coverage." The theoretical analysis and experimental evaluation of his article have proved the effectiveness of the system, which can effectively improve the security performance of user location [4]. Li et al. proposed a user privacy location protection method based on Markov model. He used the transition probability matrix between states of Markov model to predict the occurrence state and development trend of events, so as to predict the user's location and hide and encrypt the location data [5]. These studies provided technical and data support for location privacy protection, but they are troublesome to operate, and need a lot of data and accurate models to ensure the accuracy of the results.

Trusted third-party (TTP) system can improve service response efficiency, and hiding technology can hide location information. Ashraf has studied three different privacy protection schemes based on LBS system. It includes untrusted third party, TTP, and mobile peer-to-peer (P2P) location server. In the LBS system based on TTP, the privacy of users is related to personal identity, location information, and time information. In order to realize privacy protection under these problems, he studied and analyzed the existing advanced mechanisms [6]. Reddy and Balaraju studied the cloud storage based on third party anonymous servers. They concluded that information technology enables a large number of access data to be updated and transferred to the centralized large-scale data center, transfers the service management trust to the cloud environment, and comprehensively processes the location information of the cloud environment [7]. Zhang et al. proposed a k-based anonymous virtual location privacy protection method based on geographic semantics. Experimental results show that this method can ensure the physical dispersion and semantic diversity of location, and ensure the accuracy of location and the efficient performance of privacy protection [8]. Zhou et al. designed a privacy protection scheme for remote medicine diagnostic system based on WBAN. In telemedicine system, this anonymity technology is conducive to hiding the identity of patients and protecting the privacy of patients. Performance analysis shows that this scheme can effectively reduce the communication cost, and the privacy security performance is good [9]. These methods improve the security of user privacy to a certain extent. However, they also have some technical disadvantages, which is the need to be continuously optimized and upgraded.

3 Privacy protection method based on location service

3.1 Location privacy protection method of hidden space

Hiding space protection is a method to protect users' location privacy. The entire use process is divided into

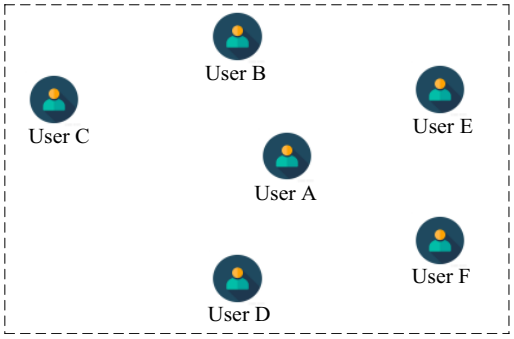

Figure 1: Hidden space of User A.

two parts, namely, access permission setting and access control decision. By creating a fuzzy area according to the user's real location, this work uses the fuzzy area to replace the real location, so as to hide the user and protect privacy [10]. If the area contains k users, the area can basically complete the K anonymous protection of user locations. As shown in Figure 1, hidden space technology mainly uses TTP architecture and mobile point-to-point architecture.

3.1.1 TTP architecture

TTP is also known as trusted anonymous location server. The third party here refers to an object other than the two interrelated subjects, which is called the third party. The third party, which can be related to two subjects, or it can be content independent of the two subjects, is an intermediate medium between users and location service providers. It mainly includes three processes: collecting and recording the user's location information service, using the hidden space generation algorithm to generalize the user's real location into a fuzzy area and replace the user's real location, and sending a request to the location service provider and getting the corresponding feedback and results [11,12]. The specific TTP system framework and operation process are shown in Figure 2.

The advantage of this architecture is to reduce the computing power and power requirements of user equipment, because users can complete the computing task of equipment hiding space through a third party without passing through their own equipment. However, the shortcomings of this system are also very obvious. In this system, the TTP is a very key factor, so it has high requirements for the server. When many users send service requests to suppliers at the same time, the TTP server may run overload and reduce the processing speed of information. At the same time, users must trust the third party unconditionally. If the third party is invaded by malicious attackers, there will be a significant risk of exposing location and privacy disclosure.

3.1.2 Mobile point-to-point architecture

Based on mobile P2P architecture, it does not rely on third-party servers and has the same cooperative relationship between mobile users [13]. In P2P mode, individual components are called peers. A peer can act as a

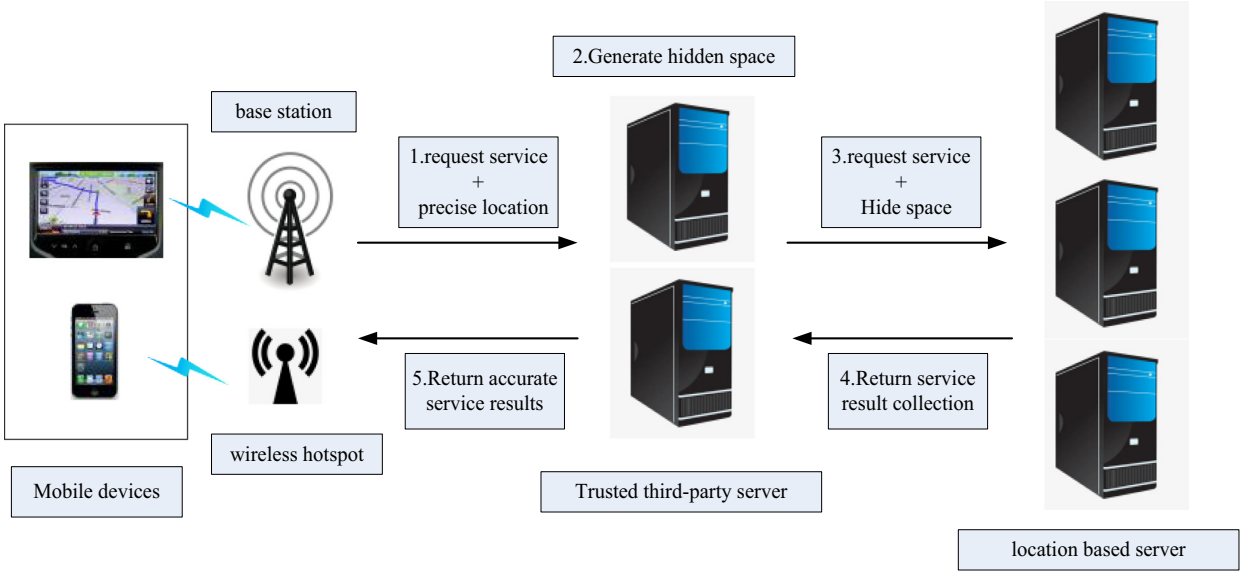

Figure 2: TTP architecture framework.

client, requesting services from other peers, and can also act as a server, providing services to other peers. A peer can act as a client or server or both, and can dynamically change its role over time. Users automatically organize mobile P2P networks through peers. Mobile P2P is a system for users to share location information. P2P technology can provide a platform on which users spontaneously form a P2P network to share location information with each other. When the user sends the location information service request, the point-to-point technology will generate a hidden space for replacing the real location according to the location information of other users in the platform and send it to the service provider [14]. Figure 3 shows the architecture based on mobile point-to-point.

The advantage of the system is that it does not need a TTP, and it also solves the risk of third-party leaking location privacy. But the disadvantage is that the user's equipment needs higher computing power, and the network traffic will increase accordingly.

3.1.3 Structure of privacy protection algorithm in hidden space

Five other users exist in the hidden space generated by user A, so a sufficient number can be generated for anonymous protection. When the scope of hiding space is larger, more users will be included in it, and the hiding degree of user location will be higher, but this will also improve the cost of server and network and reduce the quality of service [15].

The privacy protection method of hidden space is divided into three steps: collecting information on candidate locations, calculating hidden space, and processing service requests. When collecting candidate location information, users must collect enough location information, including the location information of surrounding neighbors and the location information of long-distance users transmitted by surrounding neighbors. The former is called single-hop candidate location, and the latter is called multi-hop candidate location. In a traditional wireless LAN, each client accesses the network through a wireless link linked to an access point (AP). If users want to communicate with each other, they must first access a fixed AP, which is called a single-hop network. In a wireless network, any unlimited device point can act as both AP and router, each node in the network can send and receive signals, and each node can communicate directly with one or more peer nodes, this kind of network is called multi-hop. Then, after collecting the information, it uses various algorithms to create hidden space. Finally, users replace the real geographical location with hidden space. It sends a service request to the location

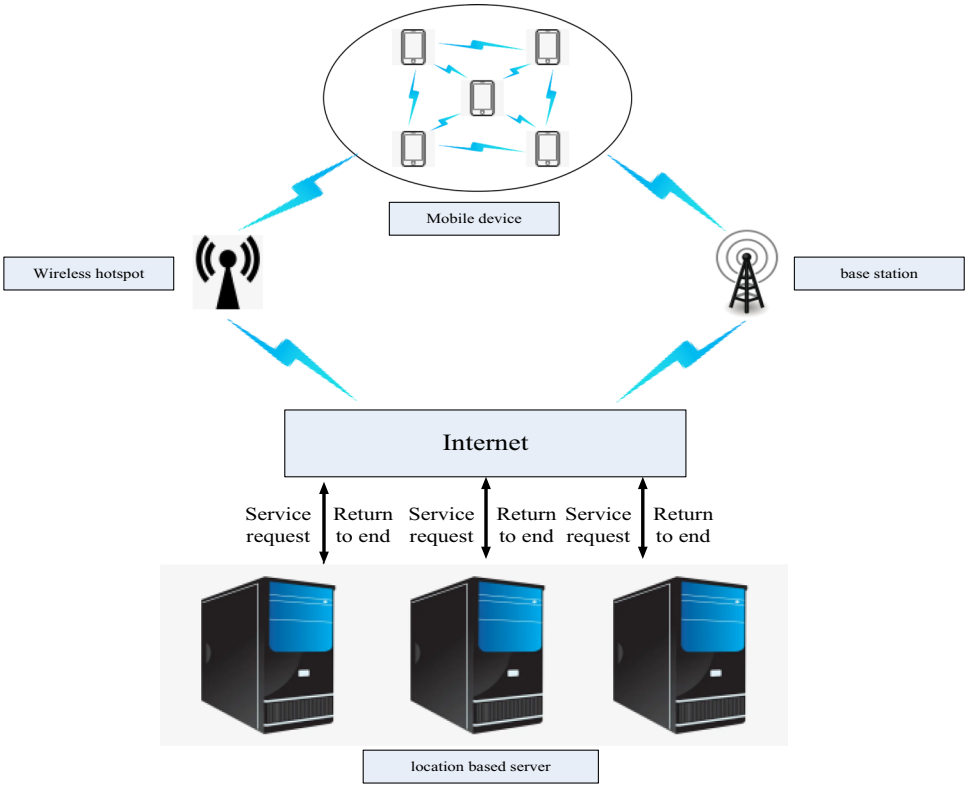

Figure 3: Architecture based on mobile point-to-point.

service provider based on the hidden space to protect its location privacy.

Figure 4 shows the operation diagram of the privacy protection law of hidden space. When the user X is ready to send the location service request to the supplier, the user needs to request the surrounding neighbors first to get their location information, then the neighbors x_1, x_2, and x_3 send their accurate location information to X. If the location information received by X is not enough, it cannot meet the K anonymous privacy protection standard, and X continues to collect information from next neighbors. When the users x_4 and x_5 receive the request of X via x_3, x_4, and x_5, they transmit their location information to X through x_3. After the user X meets the quantity requirement, then a hidden space is created based on these real location information. Then, X randomly selects a user in the hiding space as an agent, sends a service request to the location service provider through the agent, and hides his real location information. At this time, the result set returned by the service provider is based on the hidden space location information. The result set contains not only the information needed by X, but also some useless information, such as spatial environment data, initial location information, or other information that has no practical use. User X can choose the best protection result of privacy from these information.

3.1.4 False trajectory method

The so-called false location refers to hiding the user's real location information (a, b). This method of sending false location (a, b) to obtain the corresponding location information service is a very simple but reliable privacy protection method. The false track method generates some false track data on each track the user runs, and replaces the true track with the false track, so as to hide the real track and protect privacy [16]. For example, it records the user's original trajectory data as shown in Table 1, uploads the location information of users A_1, A_2, A_3 based on s_1, s_2, s_3 time to the server, and obtains three different trajectories.

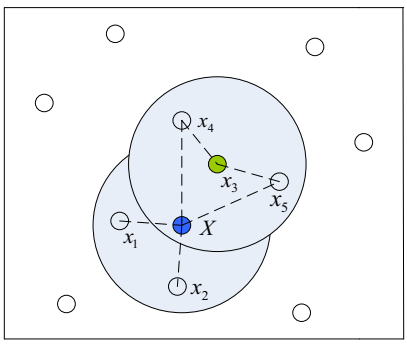

Figure 4: Operation of privacy protection method in hidden space.

Table 1: Original track data

Users	s_1	s_2	s_3
A_1	(2,4)	(2,6)	(5,8)
A_2	(1,3)	(4,5)	(6,6)
A_3	(1,5)	(5,7)	(7,9)

Table 2 shows the results after the original trajectory data are photographed by the false trajectory data. In this Table, A_1, A_2, A_3 is the original trajectory data and A_1', A_2', A_3' is the false trajectory data. Therefore, each real trajectory has only half the probability of being recognized. In general, the more false trajectories generated, the lower the possibility of identifying the real trajectories [17].

Under continuous query, connecting the user's location information at different time points is the user's real motion trajectory. If each point only randomly generates a false position, the attacker can easily trace the real motion trajectory through this information [18,19]. Therefore, the false trajectory should be similar to the real trajectory as much as possible, so that the attacker cannot distinguish the false trajectory from the real trajectory. It is best to generate it according to the requirements of the nearest neighbor principle, the similarity principle, and the memory principle.

The false trajectory generated in Figure 5a meets the requirements, which can make it difficult for the attacker to distinguish the true from the false. Figure 5b does not meet the requirements because the starting point s_1 of the false track l_2 is out of range, and Figure 5c does not meet the requirements because the direction of the false track l_2 deviates greatly from the direction of the original track.

Table 2: Original trajectory data and false trajectory data

Users	s_1	s_2	s_3
A_1	(2,4)	(2,6)	(5,8)
A_2	(1,3)	(4,5)	(6,6)
A_3	(1,5)	(5,7)	(7,9)
A_1'	(3,4)	(4,7)	(3,9)
A_2'	(1,5)	(3,3)	(2,1)
A_3'	(4,5)	(2,7)	(3,3)

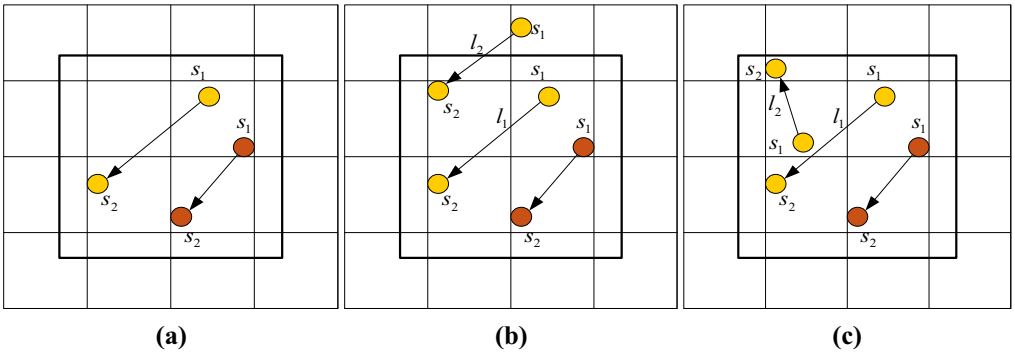

Figure 5: Examples of generated false position (red dot is the real position and yellow dot is the false position). (a) Meet the requirements and (b and c) do not meet the requirements.

3.2 Two-level cache privacy protection method based on user grid

3.2.1 Definition

It divides the map into fixed grid area size as the smallest anonymous area, and stores each fixed grid area object in a quadtree [20]. The process of this method is as follows: it divides the geographical space into different tree structures according to different levels (level of complexity), and takes the fixed grid area space as the smallest subspace. It then recursively divides the geospatial into four subspaces with the same size, and continues to divide until each subspace reaches the size of a fixed grid area.

Figure 6a is the division of fixed grid areas. The position of area 1 can be represented by (x_0, y_0), the position of area 2 can be represented by (x_1, y_0), and so on. The position of each grid can be positioned. Figure 6b is a quadtree storage structure with 2 layers and $U(2)$ operation complexity. It has 16 leaf nodes and 5 flying leaf nodes.

3.2.2 Algorithm design

The user takes a grid area instead of its real location, and the area of the grid should be smaller than the minimum anonymous area $V_{\min}$ set by the user. The algorithm flow chart is shown in Figure 7.

When a user sends a location service request, first compare the size of the user's grid area D_{area} and the minimum anonymous area $V_{\min}$. If $D_{\text{area}} > V_{\min}$ is invalid, it is directly replaced by $V_{\min}$. If $D_{\text{area}} < V_{\min}$, it is the area D_{area} set by the user. Then, two percentages are randomly generated to determine the proportion between the user's real position and the grid area. Finally, the percentage and the user's real location are used to determine the location of the grid area and generate the user's grid area.

3.3 Calculation of privacy protection

3.3.1 Differential privacy protection

Let (a_w, b_w) be the central position coordinates of grid, grid_1, and (a_{wn}, b_{wn}) be the position coordinates of the nth

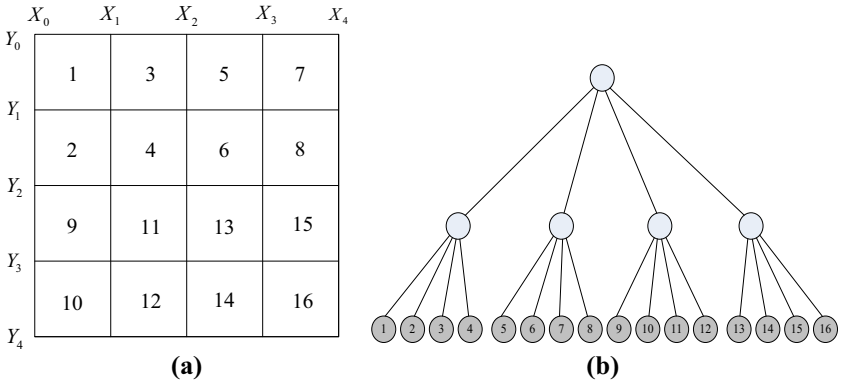

Figure 6: Quadtree structure of fixed grid area. (a) Partition of fixed grid region and (b) storage structure of quadtree.

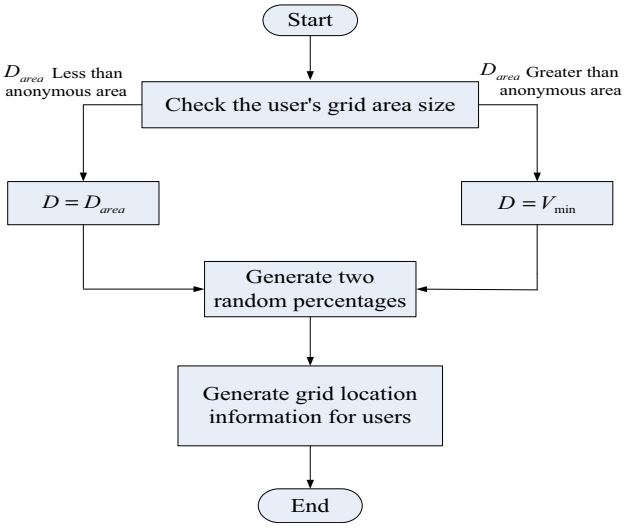

Figure 7: Algorithm flow for generating user's grid area.

grid, grid_n. Point = (a_q, b_q) is the centroid in the grid, and the formula of centroid is:

$$\text{Point} = (a_q, b_q) = \left[\frac{\sum_{i=1}^{n}\text{grid}_i \cdot a}{n}, \frac{\sum_{i=1}^{n}\text{grid}_i \cdot b}{n}\right]. \quad (1)$$

Let R be a random algorithm, G_R be all output sets of R, K and K' be two adjacent datasets, $|k \Delta K'| \le 1$, C_R is any subset of G_R. If algorithm R satisfies formula (1), algorithm R satisfies ε- differential privacy.

$$Ur[R(K) \in C_R] \le e^{\varepsilon} \cdot Ur[R(K') \in C_R]. \quad (2)$$

In formula (1), the datasets K and K' are used as the input of algorithm r to obtain $R(k)$ and $R(k')$. The probability Ur is the risk of privacy disclosure, and the degree of privacy protection can be expressed by the privacy budget coefficient ε. ε is negatively correlated with the privacy protection. If ε is smaller, the degree of privacy protection is higher. If $\varepsilon = 0$, it means that the effect of privacy protection is the best. The value of ε will affect the search range and offset distance of the algorithm, and the offset distance is related to the quality of service of the location and the degree of privacy protection of the location. The larger the offset distance, the lower the service quality of the location and the lower the privacy protection degree. On the contrary, if the offset distance is smaller, the service quality of the location is higher, and the privacy protection degree is also higher. Therefore, it is very important to select an appropriate value of ε.

Laplace technology completes ε-differential privacy protection by adding random noise that obeys the Laplace distribution. Laplace is used to add noise to the data. Let c be a continuous random variable, and the probability density function is obtained as follows:

$$G(a) = \frac{1}{2\mu}\exp\left(-\frac{|c|}{\mu}\right). \quad (3)$$

Noise parameter μ has a positive correlation with the degree of privacy protection and a negative correlation with noise sensitivity ω. Therefore, the larger the noise parameter ε, the smaller the ε and the higher the degree of privacy protection. The relationship of differential privacy coefficient ε is obtained:

$$\varepsilon = \frac{c}{\mu}. \quad (4)$$

After the anonymous area is generated, Laplace noise reduction is carried out, and the centroid position in the anonymous area is set as $v_t = (a_m, b_n)$ and $v_s = (a_s, b_s)$ disturbance positions. Therefore, v_s meets the following requirements:

$$Ur(a_m \to a_s) \le e^{\varepsilon} Ur(a_n \to a_s), \quad (5)$$

$$Ur(b_m \to b_s) \le e^{\varepsilon} Ur(b_n \to b_s). \quad (6)$$

$\varepsilon \ge 0$, $m, n \in \{1, \ldots, i\}$; $Ur(a_m \to a_s)$ is the probability of a_s corresponding to a_m; $Ur(b_m \to b_s)$ is the probability of b_m for b_s.

By adding a difference disturbance to the abscissa and ordinate coordinates of the centroid points in the anonymous area, and then combining them into disturbing points, adding the Laplace distribution with scale parameter μ to the coordinates, and disturbing the coordinates a_m and a_n in the center position $v_t = (a_m, b_n)$, the following is obtained:

$$Ur(a_m \to a_s) = \frac{1}{2\mu} e^{-\frac{|a_m - a_s|}{\mu}}, \quad (7)$$

$$Ur(b_m \to b_s) = \frac{1}{2\mu} e^{-\frac{|b_m - b_s|}{\mu}}, \quad (8)$$

The noise number −psign(rnd)In(1−2|rnd|) is added to each coordinate, and r_{nd} is a uniform random value between −0.5 and 0.5. Set μ as $(\max a_x - \min a_x)/\varepsilon$ and $(\max b_x - \min b_x)/\varepsilon$ to generate a_s and b_s, respectively, so that the disturbance position is $v_s = (a_s, b_s)$.

V represents a position component, which is quantified by applying the triangular infinitive formula:

$$|v_n - v_s| \le |v_n - v_m| + |v_m - v_s|. \quad (9)$$

Lifting both sides to a power function to obtain:

$$e^{-\frac{|v_n - v_s|}{\mu}} \ge e^{-\frac{|v_n - v_m| + |v_m - v_s|}{\mu}}. \quad (10)$$

Conversion:

$$e^{-\frac{|v_n-v_s|}{\mu}} \geq \frac{e^{-\frac{|v_m-v_s|}{\mu}}}{e^{-\frac{|v_n-v_m|}{\mu}}}. \tag{11}$$

Sorting:

$$e^{-\frac{|v_n-v_s|}{\mu}} \geq e^{-\frac{|v_n-v_m|}{\mu}} e^{-\frac{|v_n-v_s|}{\mu}}. \tag{12}$$

Multiplying $1/2\mu$ with two variables to obtain:

$$\frac{1}{2\mu} e^{-\frac{|v_n-v_s|}{\mu}} \geq \frac{1}{2\mu} e^{-\frac{|v_n-v_m|}{\mu}} e^{-\frac{|v_n-v_s|}{\mu}}. \tag{13}$$

Combining formulas (6) and (7), formula (12) is modified as follows:

$$Ur(v_n - v_s) \leq e^{\frac{|v_n-v_m|}{\mu}} Ur(v_n - v_s). \tag{14}$$

Arranging each coordinate a_s, b_s to obtain

$$\begin{aligned} Ur(a_m \to a_s) &\leq e^{-\frac{|a_n-a_m|}{\mu}} Ur(a_n \to a_s) \\ &\leq e^{\frac{|\max a_x - \min a_x|}{\mu}} Ur(a_n \to a_s), \end{aligned} \tag{15}$$

$$\begin{aligned} Ur(b_m \to b_s) &\leq e^{-\frac{|b_n-b_m|}{\mu}} Ur(b_n \to b_s) \\ &\leq e^{\frac{|\max b_x - \min b_x|}{\mu}} Ur(b_n \to a_s). \end{aligned} \tag{16}$$

3.3.2 Algorithm privacy analysis

For a group of probability distributions r_1, r_2, ..., r_n, the relation of information entropy is:

$$F = -\sum r_k \log_2 r_k. \tag{17}$$

The location where users access at $c + 1$ is sensitive location, and privacy protection algorithm makes $s - 1$ candidate location in $c + 1$. The probability of the user accessing the s location at $c + 1$ is r_1, r_2, ..., r_j, and the probability of staying at the c location at $c + 1$ is r_0. Therefore, the information entropy of privacy protection in (c, c + 1) time is

$$F_{(c,c+1)} = -\sum_{k=1}^{j} r_k \log r_k - r_0 \log r_0. \tag{18}$$

When the user accesses all candidate locations with the same probability when $c = 1$, the index of information entropy is the largest, and the following is obtained:

$$\text{Max } F_{(c,c+1)} = -\log\left(\frac{1}{j+1}\right). \tag{19}$$

Getting

$$F_{\%} = \frac{F_{(c,c+1)}}{\text{Max } F_{(c,c+1)}} 100\%. \tag{20}$$

To sum up, $F_{\%}$ is positively correlated with the intensity of privacy protection. The greater the $F_{\%}$, the better the privacy protection effect.

4 Experiment and privacy protection based on location service

4.1 Privacy and security of third-party anonymous servers

4.1.1 Impact of human traffic on cache hit rate

This work sets 2,000 initial users and sends 200 user data randomly every 15 min. The change trend is shown in Figure 8.

Figure 8(a) shows that the cache utilization is positively correlated with the number of users. As the number of users increases, cache utilization also increases. It makes the client and anonymous server query service requests in the cache as much as possible, and the number of communications between users, service providers, and servers will be reduced accordingly. This greatly reduces the burden of anonymous servers and service providers, and improves the efficiency of service response and the degree of privacy protection [21]. From Figure 8 as a whole, the probability that the cache list provides services increases with the increase in K. The highest can reach about 0.9.

4.1.2 Cache hit rate analysis

As can be seen from Figure 9, the cache hit rate is positively correlated with the anonymous size k and the simulation time t. When the user's privacy changes, the cache hit rate of the third-party anonymous server is higher than 85%. Therefore, the third-party anonymous server can greatly improve the response speed and cache utilization, so as to enhance the security of the location and protect the user's privacy.

4.1.3 Safety performance analysis

1) In terms of users, when users request location services, they only need to establish contact with the third-party

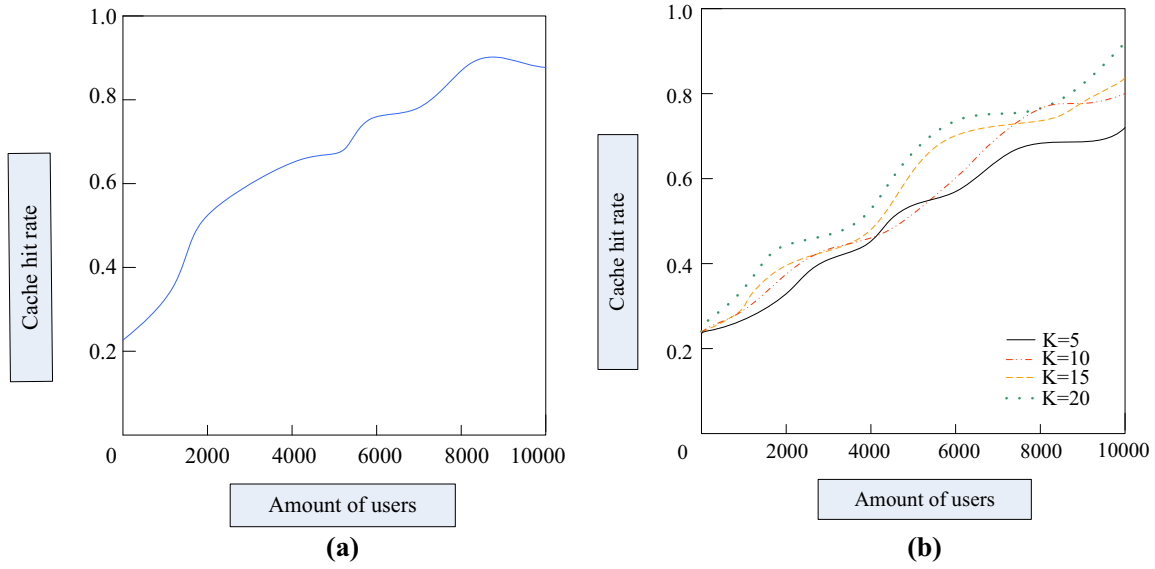

Figure 8: Change trend of cache usage times. (a) Change in cache usage with the number of users. (b) Change in cache usage with anonymous size *k*.

anonymous server, and do not send the real location information to the third-party anonymous server, and there is no user's location information in the result set sent by the server [22]. Therefore, the privacy protection performance at the user level is high.

2) In terms of third-party anonymous server, it is assumed that the attacker has invaded the third-party anonymous server and obtained all the information of the user grid area. Because the user's position in the grid area is randomly generated, the user's real position cannot be captured by the attacker [23]. In the traditional *K* anonymous method, when the user sends the location service request, the real location will be sent together. If the attacker invades the third-party anonymous server, the real location information of all users stored in the server will be captured by the attacker. Therefore, the privacy and security of users cannot be guaranteed.

4.2 Performance of differential privacy protection

This experiment simulates the user's location sequence based on the simulation dataset. In order to better fit the user's location information in real life, the trajectory

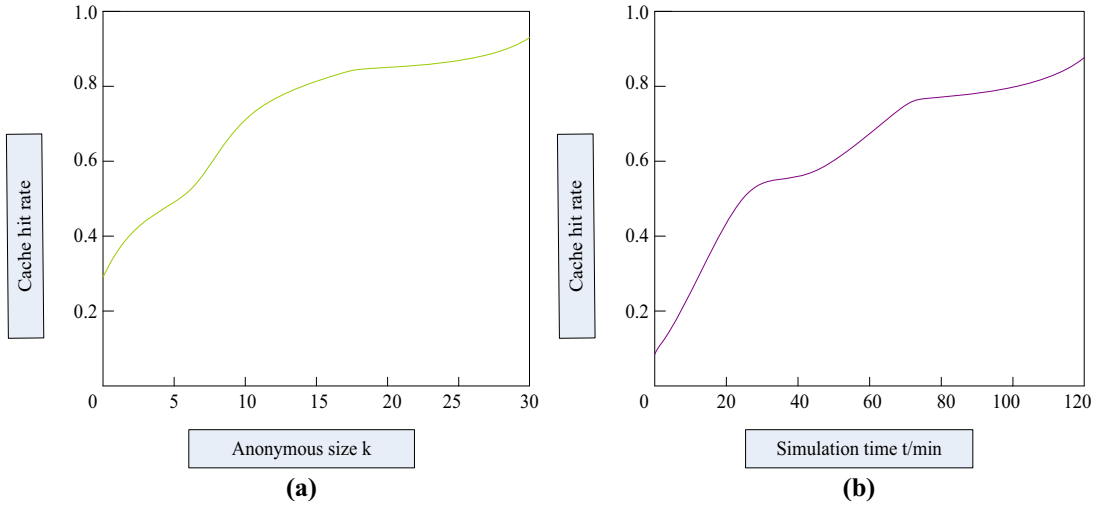

Figure 9: Change trend of cache hit rate. (a) Change in cache hit rate with anonymous size *k*. (b) Change in cache hit rate with simulation time *t*.

Table 3: Experimental results under different c values, number of real elements, and number of users

Real elements and fake elements	User	$c = 1$		$c = 3$		$c = 5$	
		NDCG	MRE	NDCG	MRE	NDCG	MRE
$e = 20$	$u = 2 \times 10^4$	0.97	0.4	0.91	0.09	0.21	1.02
	$u = 4 \times 10^4$	0.99	0.3	0.95	0.05	0.48	0.44
$f = 10$	$u = 2 \times 10^5$	1	0.1	0.98	0.02	0.77	0.78
	$u = 4 \times 10^5$	1	0.1	0.98	0.01	0.90	0.05
$e = 25$	$u = 2 \times 10^4$	0.98	0.6	0.79	0.23	0	8.72
	$u = 4 \times 10^4$	0.99	0.3	0.81	0.15	0.01	1.26
$f = 10$	$u = 2 \times 10^5$	1	0.1	0.97	0.08	0.05	0.47
	$u = 4 \times 10^5$	1	0.1	0.99	0.03	0.11	0.19

sequence contains interest points such as shopping malls, cinemas, and restaurants, rather than location coordinates.

Differential privacy algorithm has an essential problem. When the number of users is small, the result of data analysis is not ideal [24]. Differential privacy algorithm is a pattern that uses random feedback mechanism to randomly generate the sequence of each user. It is basically based on random feedback, and each user will randomly feedback the sequence pattern set. The sequence value domain will affect the accuracy of frequent mode results, that is, the random feedback will become scattered with the increase in the value domain of the sequence mode, and the performance of the final frequency estimation will also be reduced. This problem will become more serious as there are more and more data elements and longer sequence patterns. Suppose that there are $e + f$ real elements and false elements, c is the length of sequence patterns, then it is the number of sequence patterns R_{e+f}^{c}, and the index is directly proportional to e, f, and c. Normalized discounted cumulative gain (NDCG), normalized loss recovery, is a measure of the ranking quality of the sequence. The average relative error is used to estimate the accuracy of the sequence frequency. Mean relative error (MRE) is the average relative error between the real frequency and the estimated frequency, which is used to evaluate the accuracy of sequence mode frequency [25,26].

The performance of privacy protection results is analyzed by adjusting the number of real elements and users of the simulation dataset. Parameters $\varepsilon = 5$, $k = 25$, and sequence pattern length c are taken as $c = 1$, $c = 3$, and $c = 5$, respectively, to observe the experimental results of different factors. At this point, the algorithm P-LDPSPM degenerates into LDPSPM and stops the operation of pruning steps.

It can be seen from Table 3 that when the value of c becomes smaller, the performance of NDCG value and MRE value is better. When the value of c becomes larger, the performance of the algorithm decreases gradually. Therefore, even if the number of users increases, better performance can be maintained. For example, when the real element is 20 and $c = 5$ and when the number of users is increased from 20,000 to 400,000, the NDCG value increases from 0.21 to 0.90 and MRE decreases from 1.02 to 0.05. At the same time, if the number of users increases to a certain number, that is, when it reaches 4×10^5, the MRE index will gradually be in a stable state.

There are 30,000 user sequences, with 8 fixed sequences for each user, including 50 real positions and 10 false positions. Set the parameter $\varepsilon = 5$, of the dataset, and conduct the test on the simulated dataset and the real dataset, respectively, and change the value of parameter k at the same time. P-LDPSPM is an algorithm that includes pruning steps, which is the optimization and upgrading of LDPSPM. The test results are shown in Figure 10.

As can be seen from Figure 10, the k value and NDCG value are positively correlated. When the k value becomes larger, the probability of sorting failure in the original data decreases, thus the effect on NDCG decreases and the value of NDCG increases correspondingly. When the value of k increases, the value of average relative error also increases. That is, when more frequent sequences are taken into account, the relative error of the average value of the corresponding frequency estimation increases. On the contrary, when less frequent sequences are considered, the relative error decreases and the accuracy of frequency estimation is better. Because the mode of the original data is centralized and is of high frequency, it often appears in the random feedback of users, and the corresponding frequency estimation is closer to the real data. When the K value decreases, the P-LDPSPM with pruning step is better than LDPSPM, and its performance is improved by about 7%, and as the value of increases, the performance of both methods is slow.

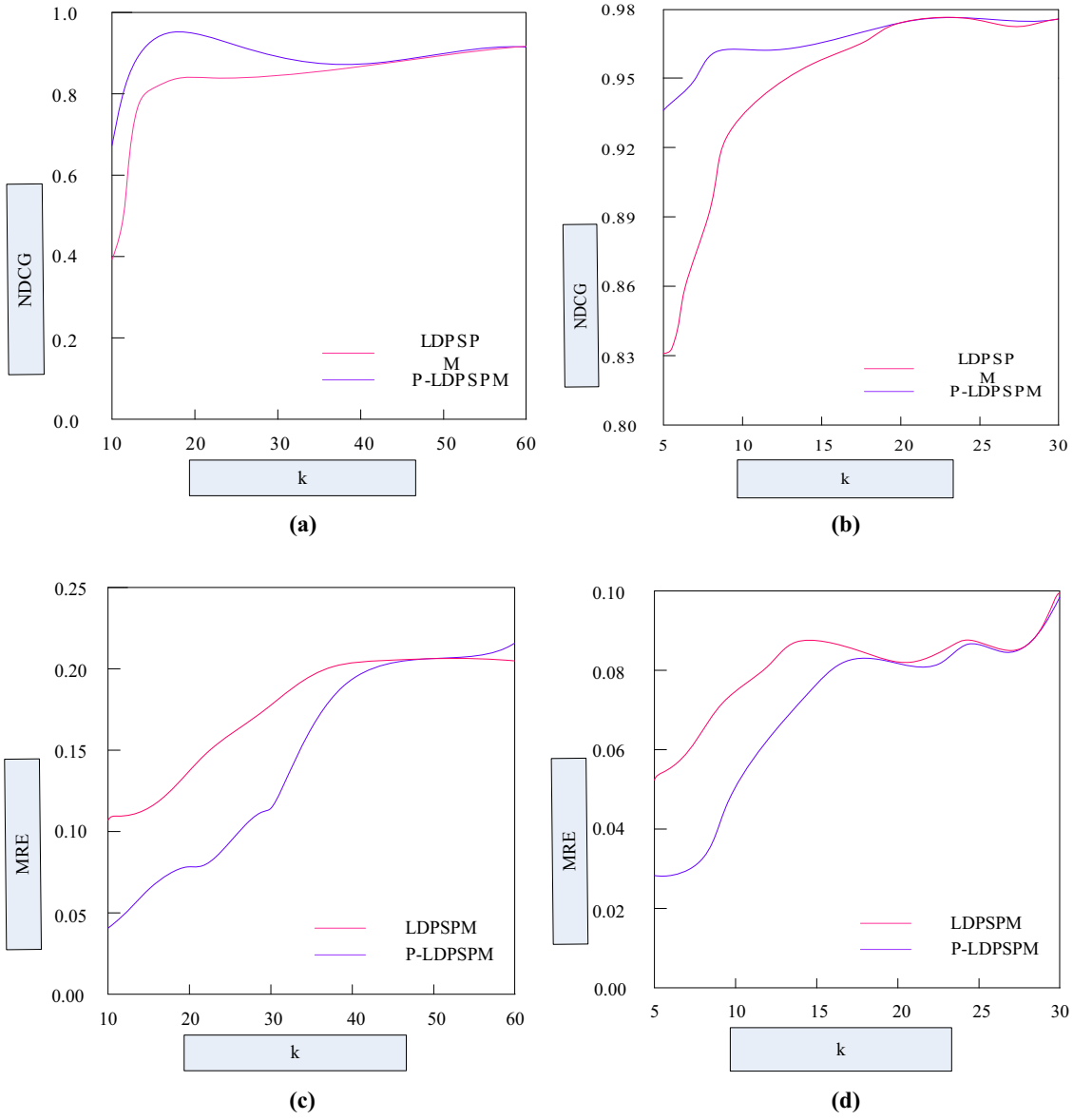

Figure 10: Impact of privacy parameters. (a) NDCG simulated dataset, $\varepsilon = 5$, (b) NDCG real dataset, $\varepsilon = 5$, (c) MRE simulation dataset, $\varepsilon = 5$, and (d) MRE real dataset, $\varepsilon = 5$.

Table 4: Test scenarios

Parameter Total number of users	Scenario 1 2,000	Scenario 2 4,000	Scenario 3 6,000	Scenario 4 8,000
k anonymous request	10–100			
Point-to-point transmission distance	400–500 m			
The maximum acceptable hop value	10			
Location information processing time	200 ms/bar			
The length of the location information	128 bytes			
The maximum validity period for location information	10 s			
Maximum wait time for service requests	5 s			

With the increase in K value, the proportion of candidate data elements generated by pruning steps increases, and the P-LDPSPM performance advantage of pruning step is no longer prominent, and may also have adverse effects on data availability. When the k value reaches a certain height, the whole dataset becomes a candidate set generated by pruning steps. At this time, the pruning step stops running, and the result of P-LDPSPM and LDPSPM is the same.

4.3 Performance of hidden space algorithm

Four experimental scenarios are generated by the simulator, as shown in Table 4.

It can be seen from Table 4 that the number of users has increased from 2,000 to 4,000. Considering the different needs of different users, K anonymous protection requirements are randomly selected from 10 to 100. Because different devices have different operation and processing capabilities, the point-to-point transmission distance is randomly specified at 400–500 m, and the processing time of location information is 200 ms each. In order to facilitate calculation, the average length of a message is set to 128 bytes. The maximum valid time of location information is 8 s, and the maximum waiting time of service request is 5 s. The test results are shown in Figure 11.

Figure 11(a) shows the success rate of generating hidden space by on-demand, active, and dual-active algorithms in different scenarios. The success rate of on-demand and active hiding space generation is the same, while the success rate of dual-active is higher, which is 13% higher than that of the other two algorithms. Dual-active can use candidate location information within a period of validity, so users have more location information for generating hidden space, which makes it easier for users to meet their K anonymous protection requirements. Comparing the four scenarios, it can be seen that the total number of users is positively correlated with the success rate of hidden space generation. When the number of users increases, the probability of generating hidden space is higher. The success rate of hidden space generation increases because the more the users, the easier it is to collect candidate location information, and so the higher the success probability.

Figure 11(b) shows the average time for the three algorithms to generate hidden space in different scenarios. As can be seen from the Figure, on-demand takes the longest time to generate hidden space and dual-active takes the shortest time, which is 25% shorter than the other two algorithms. In experimental scenario 4, the average generation time of on-demand, active, and dual active algorithms is 410, 150, and 60 ms, respectively. This is because even if the user does not send a location service request, the algorithm will continue to collect the user's candidate location information during the validity period. In addition, it can be seen that the time of generating hidden space is negatively correlated with the total number of users. When the number of users

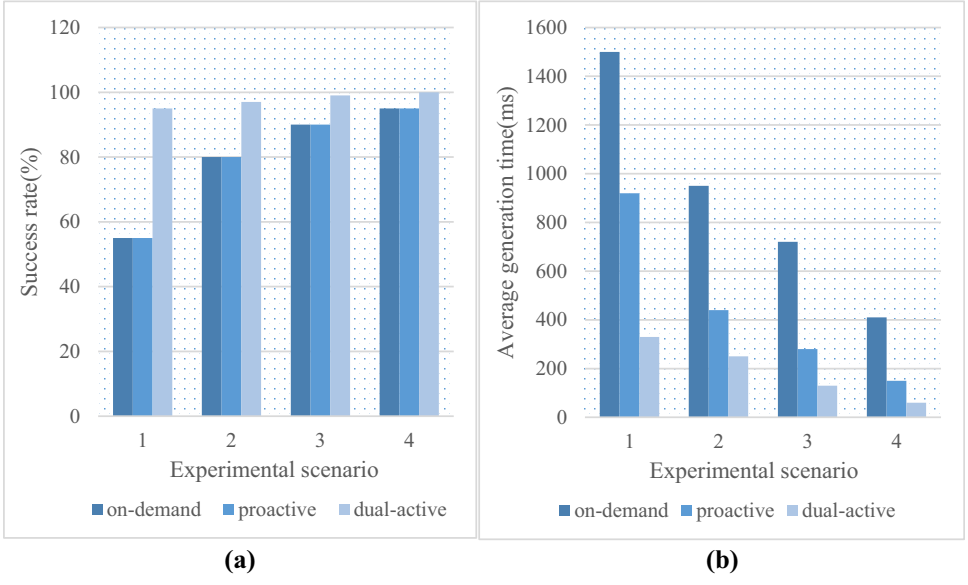

Figure 11: Success rate and average generation time of hidden space under different scenarios. (a) Success rate of generating hidden space in different scenes and (b) average time of generating hidden space in different scenes.

increases, the average time to generate hidden space decreases. This is because the denser the distribution of users, the easier it is to collect more candidate location information for generating hidden space.

It can be concluded that the privacy protection performance of dual-active algorithm is the best. By allowing users to share candidate location messages within the validity period, this algorithm can improve the success rate of generating hidden space, shorten the generation time, and protect location privacy.

5 Discussion

This article focuses on the privacy protection methods of location services in big data. It not only expounds and analyzes the construction and operation of location services and privacy protection algorithms, but is also a new attempt on privacy protection methods.

The analysis of this article shows that it is very necessary to protect users' location privacy in the era of big data. This work uses different algorithms to analyze, study, and compare the methods of user location privacy protection, which can ensure the security of user privacy. It reduces the possibility of adverse consequences caused by the use of private information by criminals, maintains social stability, and promotes social development. This work mainly uses the hidden space method, two-level cache method based on user grid, differential privacy protection method, and false trajectory method. Finally, this work analyzes and compares the performance of privacy protection algorithms, and comes to the conclusion that dual-active algorithm has better privacy protection performance. The performance of LDPSPM algorithm can be improved.

6 Conclusion

Through the analysis of this work, the following conclusions are drawn: (1) the third-party anonymous server greatly reduces the burden of servers and service providers, and improves the efficiency of service response and the degree of privacy protection. When the user's privacy changes, the cache hit rate of the third-party anonymous server is higher than 85%, so the security of the location is effectively guaranteed. (2) The advantage of mobile point-to-point system is that it does not need to pass through a TTP, which can better avoid the risk of location exposure and privacy disclosure caused by the third party. However, the disadvantage is that the requirements for users' equipment are higher, and the cost of network traffic will increase accordingly. (3) The hiding space method has better privacy protection performance, and the larger the scope of the hiding space, the higher the hiding degree of the user's location. However, the cost of servers and communications will increase accordingly, and the quality of service will decrease. (4) In the differential privacy protection method, when the value of K is small, the privacy protection performance of P-LDPSPM with pruning step is better than that of LDPSPM, and the performance is improved by more than 7%. (5) This work has made some contributions to the research on privacy protection methods of location services in big data, but there are also deficiencies. The experimental level of this work is limited, and all interference factors are not considered when verifying the privacy protection performance of different algorithms, which needs to be improved. The attack on location information is also an important part of the research on location privacy protection. This work does not introduce it comprehensively and specifically, which is the direction of further research in the future.

Acknowledgement: Teaching reform projects of Hunan Province in 2021: Research and Practice of online and offline mixed teaching mode of College English course under the background of "new infrastructure" of higher Education (HNJG-2021-1283).

Conflict of interest: The authors declare that they have no conflicts of interest to report regarding the present study.

Data availability statement: Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

[1] O. Ruan, L. Zhang, and Y. Zhang, "Location-sharing protocol for privacy protection in mobile online social networks," *EURASIP J. Wirel. Commun. Netw.*, vol. 2021, no. 1, pp. 1–14, 2021.

[2] X. Xue, H. Li, Z. Li, and J. Xiong, "Location privacy protection scheme for LBS in IoT ," *Wirel. Commun. Mob. Comput.*, vol. 2021, no. 5, pp. 1–18, 2021.

[3] M. R. Lee, W. W. Lee, and B. Jang, "Anonymization algorithms for location privacy protection: A survey," *JP J. Heat. Mass. Transf.*, vol. 15, no. 2, pp. 291–298, 2018.

[4] Z. Wu, R. Wang, Q. Li, X. Lian, G. Xu, E. Chen, et al., "A location privacy-preserving system based on query range cover-up or location-based services," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5244–5254, 2020.
[5] H. Li, Y. Wang, F. Guo, J. Wang, and C. Wu, "Differential privacy location protection method based on the Markov model," *Wirel. Commun. Mob. Comput.*, vol. 2021, no. 4, pp. 1–12, 2021.
[6] M. U. Ashraf, K. Jambi, R. Qayyum, H. Ejaz, and I. Ilyas, "IDP: A privacy provisioning framework for TIP attributes in trusted third party-based location-based services systems," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 7, pp. 604–617, 2020.
[7] K. S. Reddy and M. Balaraju, "Comparative study on trustee of third party auditor to provide integrity and security in cloud computing," *Mater. Today Proc.*, vol. 5, no. 1, pp. 57–564, 2018.
[8] Y. B. Zhang, Q. Y. Zhang, Z. Y. Li, Y. Yan, and M. Y. Zhang, "A *k*-anonymous location privacy protection method of dummy based on geographical semantics," *Int. J. Netw. Secur.*, vol. 21, no. 6, pp. 937–946, 2019.
[9] Y. Zhou, Z. Huang, J. Chen, and Q. Chen, "Privacy protection based on anonymous consultation for telemedicine diagnosis system," *Boletin Tecnico/Technical Bull.*, vol. 55, no. 17, pp. 443–452, 2017.
[10] A. M. Olteanu, M. Humbert, K. Huguenin, and J. P. Hubaux, "The (co-)location sharing game," *Proc. Priv. Enhancing Technol.*, vol. 2019, no. 2, p. 25, 2019.
[11] Y. Huo, C. Meng, R. Li, and T. Jing, "An overview of privacy preserving schemes for industrial Internet of Things," *China Commun.*, vol. 17, no. 10, pp. 1–18, 2020.
[12] B. Khalfoun, S. B. Mokhtar, S. Bouchenak, and V. Nitu, "EDEN: Enforcing location privacy through re-identification risk assessment: A federated learning approach," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 5, no. 2, pp. 1–25, 2021.
[13] C. Chen, Y. Luo, Q. Yu, and G. Hu, "TPPG: Privacy-preserving trajectory data publication based on 3D-grid partition," *Intell. Data Anal.*, vol. 23, no. 3, pp. 03–533, 2019.
[14] S. A. Kumar and M. S. Anbarasi, "Noble authentication protocol with privacy preservation policy for public auditing on shared data," *J. Comput. Theor. Nanosci.*, vol. 16, no. 8, pp. 3252–3258, 2019.
[15] A. Shukla, D. Singh, M. Sajwan, M. Kumar, D. Kumari, and A. Kumar, "SLP-RRFPR: a source location privacy protection scheme based on random ring and limited hop fake packet routing for wireless sensor networks," *Multimed. Tools Appl.*, vol. 81, no. 8, pp. 11145–11185, 2022.
[16] L. Zheng, L. Zhang, M. Cui, N. Cao, J. Ding, L. Yalemshet, et al., "The research of mobile location privacy protection access control method based on game theory," *Wirel. Commun. Mob. Comput.*, vol. 2018, no. 1, pp. 1–9, 2018.
[17] Z. Liu, Z. Xuan, Y. Dong, and B. Zhang, "Trajectory rotation privacy protection algorithm based on k anonymity," *J. Comput. Commun.*, vol. 6, no. 2, pp. 36–47, 2018.
[18] J. Chen, H. Ma, D. Zhao, and D. S. L. Wei, "Participant density-independent location privacy protection for data aggregation in mobile crowd-sensing," *Wirel. Personal. Commun.*, vol. 98, no. 1, pp. 699–723, 2018.
[19] S. C. Soma, T. Hashem, M. A. Cheema, and S. Samrose, "Trip planning queries with location privacy in spatial databases," *World Wide Web*, vol. 20, no. 2, pp. 205–236, 2017.
[20] Y. Dong and D. Pi, "Novel privacy-preserving algorithm based on frequent path for trajectory data publishing," *Knowl. Syst*, vol. 148, no. MAY 15, pp. 5–65, 2018.
[21] T. Murakami, H. Hino, and J. Sakuma, "Toward distribution estimation under local differential privacy with small samples," *Proc. Priv. Enhancing Technol.*, vol. 2018, no. 3, pp. 84–104, 2018.
[22] M. V. Ahluwalia, A. Gangopadhyay, Z. Chen, and Y. Yesha, "Target-based, privacy preserving, and incremental association rule mining," *IEEE Trans. Serv. Comput.*, vol. 10, no. 4, pp. 633–645, 2017.
[23] S. Li, H. Shen, Y. Sang, and H. Tian, "An efficient method for privacy-preserving trajectory data publishing based on data partitioning," *J. Supercomput.*, vol. 76, no. 1, pp. 1–25, 2020.
[24] M. Nanni, G. Andrienko, C. Boldrini, F. Bonchi, and A. Vespignani, "Give more data, awareness and control to individual citizens, and they will help COVID-19 containment," *Trans. Data Priv.*, vol. 13, no. 1, pp. 61–66, 2020.
[25] H. Shen, G. Bai, and M. Yang, "Protecting trajectory privacy: A user-centric analysis," *J. Netw. Comput. Appl.*, vol. 82, no. MAR, pp. 128–139, 2017.
[26] A. A. Muthana and M. M. Saeed, "Analysis of user identity privacy in LTE and proposed solution," *Int. J. Comput. Netw. Inf. Secur.*, vol. 9, no. 1, pp. 4–63, 2017.