Journal of Network Intelligence
Taiwan Ubiquitous Information
©2023 ISSN 2414-8105 (Online)
Volume **8**, Number **4**, November **2023**

Location Privacy Desensitization Algorithm Based on Multi-attribute Decision Model

Zihao Shen

School of Computer Science and Technology
Henan Polytechnic University
Jiaozuo; 454000, China
hpuxxfzyjs@qq.com

Xin Yang

School of Computer Science and Technology
Henan Polytechnic University
Jiaozuo; 454000, China
1521437793@qq.com

Hui Wang*

School of Software
Henan Polytechnic University
Jiaozuo; 454000, China
wanghui_jsj@foxmail.com

Peiqian Liu

School of Software
Henan Polytechnic University
Jiaozuo; 454000, China
2362695089@qq.com

Kun Liu

School of Software
Henan Polytechnic University
Jiaozuo; 454000, China

*Corresponding author: Hui Wang

Received December 12, 2022, revised March 9, 2023, accepted May 24, 2023.

Abstract. *Traditional methods of building secure anonymity sets using location privacy desensitization techniques commonly suffer from low dummy location differentiation and vulnerable to background knowledge attacks. To address this problem, this paper proposes a multi-attribute decision model-based location privacy desensitization (MDMLPD) algorithm. Firstly, the construction of the candidate dummy location set is obtained by historical query probability judgment; secondly, data preprocessing is performed on the defined five attribute values and attribute weights are determined using hierarchical analysis; finally, comprehensive decision making is performed to select the optimal dummy location and construct a secure anonymous set to achieve location privacy desensitization. The experimental results show that the MDMLPD algorithm reduces the time overhead by about 9.8% compared with the MMDS algorithm, and the probability of being identified by the semantic attack is reduced by 15.3% and 26.2% compared with the MMDS algorithm and the K-DLS algorithm, respectively, what is feasible and efficient to satisfy user privacy requirements.*

Keywords: Location Privacy, Multi-attribute Decision, Desensitization, Dummy Location, Privacy Protection

1. **Introduction.** In today's era of big data, mobile communications are gaining popularity and sensing devices are constantly being introduced. With the vigorous development of location-aware technology, the geographic locations of people and transactions exist in the form of data, so privacy protection becomes necessary [1-2]. After collecting the location data of moving objects directly or indirectly, users can make location-related queries [3-4]. Location-Based Services (LBS) integrate the location information of mobile devices with other information to provide users with value-added services [5-7]. However, Location Service Providers (LSP) often collect users' sensitive location information without the user's knowledge, and there is a risk of privacy information leakage. In the incident on July 21, 2022, when Cyberspace Administration of China imposed a fine of 8.026 billion yuan from Didi, Didi was accused of 16 illegal facts. These included the illegal collection of 153 million pieces of data on taxi addresses such as home, school and company, and the excessive collection of data on the precise location (latitude and longitude) of passengers when evaluating the chauffeur service and in the background of the app. The company also collected 167 million pieces of data on the precise location (latitude and longitude) of the app when the app was running. This collected information may be used for various illegal activities, which in turn may cause damage to the user's reputation.

To prevent the leakage of location privacy information, many scholars in this field at home and abroad have proposed various location privacy desensitization techniques, which can be broadly classified into three major categories: location generalization, location perturbation, and location encryption. Location generalization [8] is implemented using spatial anonymity technology, by using this technology, the real location of mobile users will be hidden in an anonymous spatial area. However, the size of the anonymity area can become a bottleneck, limiting the development of this technology. When the generated anonymous area is too large, not only the time overhead of the solution will increase, but also the query accuracy will be greatly reduced; if the generated anonymous area is too small, the quality of privacy protection will be degraded, and it will be easy for attackers to see through. Location perturbation [9-10] is to replace the real location of the user with a location with a certain offset or a dummy location, so that the attacker or the server cannot obtain the real location of the user. However, for some attackers, they have certain background knowledge, and some dummy user-generated locations can be easily ruled out by reasoning. In this case, how to generate dummy locations with high privacy protection has become a research hotspot. Location encryption [11-13] uses

various cryptographic techniques to encrypt the location data of mobile users, and the location information of users cannot be obtained without the relevant keys. Although the use of encryption algorithms has a high degree of privacy protection to a certain extent, most of the algorithms are too complicated and require a large amount of computation, resulting in a significant increase in operating overhead.

1.1. **Related Work.** Among the many schemes of location privacy protection, most scholars have devoted themselves to researching how to generate dummy locations with a high degree of discrimination from the real location and good privacy protection effect. The basic idea of dummy location technology is to add the user's real location to the dummy location and send it to the LBS server to confuse the authenticity of the user's location so that it cannot distinguish the user's real location, thereby realizing user location privacy desensitization. Kido et al. [14] proposed the use of dummy locations to achieve location privacy protection, whose main idea is to send the user's real location to the LSP together with an anonymous set composed of many dummy locations. This approach does not require a third-party anonymous server to join and avoids the risk of privacy leakage due to trust issues of the anonymous server or attacks on the anonymous server. However, it does not consider factors such as query probability and is vulnerable to edge information attacks with background knowledge.

Considering the attacker to mine the user's historical request information, Niu et al. [15] proposed the DLS algorithm, which calculates the historical query probability based on the historical query information of the location unit on the map, followed by using the entropy measure of security, and finally selects the dummy location that meets its requirements to construct the anonymous set. The time complexity of the algorithm is high because of the need to compute the entropy value of the anonymity set extensively. It is difficult to achieve good privacy protection when used on some resource-constrained communication devices. Sun et al. [16] proposed the DLP algorithm based on DLS by analyzing their proposed attack algorithm ADLS. The algorithm makes a trade-off between time complexity and user privacy requirements and has better privacy protection to some extent, but it ignores location units with zero probability of historical queries such as mountains and lakes. Yang et al. [17] proposed the K-DLS algorithm, which improves the distribution of dummy locations while taking into account the zero probability of historical query of location units, and generates an anonymous set with a high location entropy value, which improves the security of location privacy protection. However, when attackers use semantic attacks and other means, K-DLS is clearly insufficient to protect users' location privacy by considering only the historical query probability of location units.

For attackers to grasp the semantic information of mobile users' location points, the scheme proposed in [18] considers the semantic information of user's visiting locations and satisfies the user's privacy protection requirements. However, it does not consider the impact of location distribution on dummy location sets, and the expected privacy protection effect is often difficult to achieve when subjected to location homogeneity attacks. Wang et al. [19] proposed the MMDS algorithm, which calculates the semantic difference between locations by means of a location semantic tree, while considering the query probability and geographic distribution of the dummy locations, and generates dummy locations based on these three dimensions. However, in areas with fewer POI categories, the semantic tree has fewer child nodes, and the algorithm only filters dummy locations based on location semantics, resulting in poor privacy protection. Yang et al. [20] designed a virtual location selection algorithm based on location semantic diversity, physical dispersion, and query probability, which can protect user privacy from single-point attacks, considering

location semantic diversity, physical dispersion and query probability. However, the algorithm uses a cryptographic algorithm in the process of generating anonymous sets, and its algorithmic process is too complex and computationally intensive, which leads to a significant increase in running overhead.

For the same location point, different users have different levels of sensitivity to it [21]. It is worth mentioning that the information of sensitive location is mostly associated with the user's identity. If an attacker locates a sensitive location point of a user, it is highly likely to infer the identity of that user. Yin et al. [22] proposed a location data security protection method based on location sensitivity classification. The method classifies the location sensitivity level and allocates a privacy budget based on it, which effectively protects the user's location privacy. Liu et al. [8] proposed an RSABPP algorithm based on the concept of random anonymous regions, each region has different sensitivity values, and it will be difficult for the adversary to identify the real location by inferring the sensitivity of the anonymous user.

1.2. **Motivation and contribution.** Existing location privacy protection methods have limitations with respect to background knowledge attacks and semantic attacks, and they do not take into account exhaustive factors and pose a high risk of privacy leakage. And individual methods suffer from excessive time overhead. To address these problems, the MDMLPD algorithm is proposed to better achieve location privacy desensitization. The main contributions of this paper are as follows:

(1) The regularity of trajectory location points of individual and group users is analyzed, several attribute parameters such as semantic sensitivity level, location universality and semantic similarity are defined, and MDMLPD is proposed.

(2) Analytic Hierarchy Process (AHP) is used to analyze the relationship between attributes and determine the weights reasonableness. Using the multi-attribute decision-making method, the best dummy location is selected through comprehensive decision-making, and a more indistinguishable secure anonymity set is constructed.

The MDMLPD algorithm proposed comprehensively considers the background knowledge of the user's real location and other information, which is adaptable in different scenarios. At the same time, the algorithm effectively reduces the risk of the user's real location being identified when the attacker has background knowledge such as user behavior rules and location semantics.

2. Preliminary Knowledge.

2.1. Related Definition.

Definition 1. *Privacy requirements: The personalized privacy requirements of mobile users are expressed as* $req(k, q_1, q_2)$*, where* k *is the degree of anonymity;* q_1 *and* q_2 *are the minimum and maximum distances between the dummy location and the real location, respectively, set by the user in advance.*

Definition 2. *Side information: Different location points on the map have side information, and the attacker has side information to help infer the real location of the mobile user [23]. In this paper, side information refers to information such as historical query probability and POI type of location units. Clearly, mobile users have such information, which helps them choose the best dummy location.*

Definition 3. *Historical query probability* p_i*: The map is divided into grids, different grids represent different location units, and the historical query probability is represented by the probability of users visiting the location unit in the past. After the attributes of the*

location unit are clarified, the query probability of the location unit is calculated, and the Equation is as follows:

$$p_i = \frac{N_i}{\sum_{i=1}^{n} N_i} \tag{1}$$

Among them, N_i represents the historical query times of the location unit i after grid division, and $\sum_{i=1}^{n} N_i$ represents the sum of the historical query times of all location units.

Definition 4. *Geographical distance $d_{(loc_i,loc_j)}$: The Haversine Equation is used to calculate the geographic distance between two locations. The calculation of geographic distance in this paper is only used in the construction of anonymity sets, and the results obtained do not involve deeper applications. Using the Haversine Equation can reduce the time complexity and time overhead while ensuring sufficient accuracy. The Haversine Equation is as follows:*

$$hav\left(\frac{d}{r}\right) = hav(\varphi_1 - \varphi_2) + \cos(\varphi_1)\cos(\varphi_2)hav(\lambda_1 - \lambda_2) \tag{2}$$

Where $hav(\theta) = \sin^2(\frac{\theta}{2}) = \frac{\cos(1-\theta)}{2}$, R is the radius of the earth, λ_1 and λ_2 represent the longitude of loc_i and loc_j respectively, and φ_1 and φ_2 represent the latitude of loc_i and loc_j respectively.

Simplified from Equation (2), we can get:

$$\begin{aligned} d &= 2r \arcsin\left(\sqrt{hav(\varphi_1 - \varphi_2) + \cos(\varphi_1)\cos(\varphi_2)hav(\lambda_1 - \lambda_2)}\right) \\ &= 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_1 - \varphi_2}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_1 - \lambda_2}{2}\right)}\right) \end{aligned} \tag{3}$$

Definition 5. *POI score S_{loc_i}: In LBS applications, such as yelp!, the score of each POI is based on the scores and judgments of historical users of that POI, so the score of each POI influences the user's willingness to choose.*

Definition 6. *Location prevalence De_{loc_i}: If a single location appears more frequently in a particular user's track dataset, and rarely in other users' track datasets. Then for the attacker, the location has good category differentiation ability and it is very easy to infer the identity of the user based on the location. The location prevalence is defined according to the TF-IDF technology [24], and the location prevalence of loc_i can be expressed as:*

$$De_{loc_i} = \frac{t_s^{loc}}{T_S} \times log\frac{T_A}{t_a^{loc}} \tag{4}$$

Among them, t_s^{loc} represents the number of times the location appears in the current user trajectory data set, and T_S is the total number of times that all locations appear in the current user trajectory data set; where T_A is the number of all users in the current stay area, and t_a^{loc} represents each location in the current stay area.

Definition 7. *Semantic similarity $Sim_{(loc_i,loc_j)}$: A WordNet data dictionary measure of similarity between words is used to assess the degree of semantic similarity between locations. WordNet data dictionaries are more efficient and in line with modern semantic computing, covering a wide range of nouns, verbs, and adverbs, etc., which each form a hierarchical collection of synonyms representing a basic semantic concept. These sets are connected by various relations to form a semantic tree [25], and the semantic similarity refers to the distance between two words (loc_i, loc_j) at different locations in the semantic*

tree with the following:

$$Sim_{(loc_i,loc_j)} = \frac{\sum_{m\in1,\ldots,|Snum_i|}\max_{n\in1,\ldots,|Snum_j|}Sim(Snum_i,Snum_j)}{\mid Snum_i \mid + \mid Snum_j \mid} + \frac{\sum_{m\in1,\ldots,|Snum_j|}\max_{n\in1,\ldots,|Snum_i|}Sim(Snum_j,Snum_i)}{\mid Snum_i \mid + \mid Snum_j \mid} \tag{5}$$

Where $\mid Snum_i \mid$ *and* $\mid Snum_j \mid$ *denote the number of senses of* loc_i *and* loc_j*, and sense refers to the number of semantic meanings implied by a single word. The larger the value of semantic similarity, the higher the semantic similarity of the two words, and the smaller the value of semantic similarity, the lower the semantic similarity of the two words.*

Definition 8. *Location entropy: The location entropy is determined by the probability distribution of historical queries for each location in the secure anonymity set, and can be used to measure the uncertainty of the user's location. Its calculation Equation is:*

$$H = -\sum_{i=1}^{k} q_i log_2 q_i \tag{6}$$

where $q_i = \frac{p_i}{\sum_{i=1}^{k} p_i}$*, which represents the query probability of different locations in the secure anonymity set.*

2.2. **System Architecture.** The MDMLPD algorithm proposed is based on user mobile terminals with a distributed system architecture [26]. There are no third-party servers in this architecture, effectively avoiding problems such as single point of compromise [27]. The system architecture in this paper mainly consists of three entities, i.e., mobile terminal, communication base station, and LBS server, as shown in Figure 1.

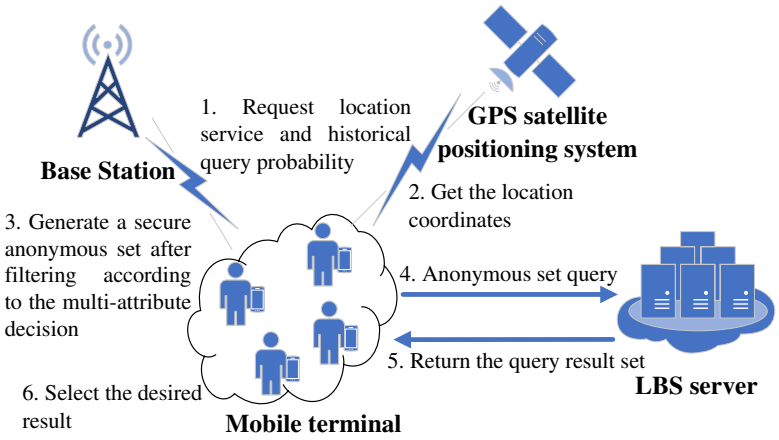

FIGURE 1. System structure

(1) Mobile terminal: the mobile terminal is used to execute the MDMLPD algorithm, followed by sending a secure anonymous set for querying, and finally selecting its desired result in the query result set. The coordinates of each location point are obtained by the mobile terminal device through the GPS satellite positioning system.

(2) Base station: Communication base stations provide network communication services for mobile terminal devices, and at the same time calculate and store the historical query probability of all locations within their coverage area. Nowadays, communication base

stations not only have wide signal coverage, but also their ability to calculate and store information is strong, so they can realize the required functions.

(3) LBS server: The LBS server is used to receive location service requests from users and return query results to provide location services to users.

3. Location Privacy Desensitization Algorithm Based on Multi-attribute Decision Model.

The selection process of dummy location has the characteristics of multi-dimensionality and multi-attributes, which belongs to the category of multi-attribute decision making. The selection of dummy location needs to be considered from various aspects, and finally an anonymous set with high indistinguishability and high location entropy is constructed to achieve location data desensitization. Hierarchical analysis in multi-attribute decision theory has been used in the field of information security for quantitative evaluation research of social media privacy security, construction of leaked data value assessment model, and other practical work with good results. Therefore, a multi-attribute decision model is used to study the location data desensitization, and hierarchical analysis is used to calculate the attribute weights of each evaluation index.

3.1. Construction and Solution of Multi-attribute Decision-making Model.

Multi-attribute decision making [28], also known as finite solution multi-objective decision making, i.e., combining multiple attributes to select the optimal alternative, occupies an important location in today's decision science. Mobile users submit the virtual locations of different queries and real information of mobile users to the LBS server and request for relevant service information. Here we use a multi-attribute decision making method based on the weighted arithmetic average operator (WAA operator) for the selection of dummy locations.

3.1.1. *Multi-attribute Decision-making Method Based on WAA Operator.* The WAA operator evaluates the merits of a solution by clustering the individual data in each row of the decision matrix according to their weight values. It is defined as follows:

Let WAA: $R \to R^n$. If $WAA_w(\alpha_1, \alpha_2, \ldots, \alpha_n) = \sum_{i=1}^{n} w_j \alpha_j$, where $w = (w_1, w_2, \ldots, w_n)$ is a weighted vector of $(\alpha_1, \alpha_2, \ldots, \alpha_n)$, $w_j \in [0, 1]$, $j \in N$, and $\sum_{i=1}^{n} w_j = 1$, then WAA is said to be a weighted arithmetic average operator.

After defining the WAA operator, the decision steps are given below:

Step1: Determine the decision objective of the proposed method in this paper and specify the set of scenarios and the set of indicator attributes. The decision objective of this paper is to select the best dummy location that meets the user privacy protection requirements to construct a secure anonymity set. Let CLS and U be the solution set and indicator attribute set, respectively. $CLS = (loc_1, loc_2, \ldots, loc_n)$ is the candidate dummy location set, and $w = (w_1, w_2, w_3, w_4, w_5)$ is the attribute weights given by the decision maker for the five attributes.

Step2: For any location loc_i, the corresponding attribute value a_{ij} is calculated according to the types of different attributes, thus constructing the decision matrix $A = (a_{ij})_{n \times m}$. The matrix A needs to be normalized because of the different magnitudes among the attributes. After normalizing the matrix A, we get $A' = (r_{ij})_{n \times m}$.

Step3: Determine the priority among the attributes defined in this paper and assign the weights corresponding to them using hierarchical analysis.

Step4: When the attribute values of each location point in the candidate dummy location set are known under each target, the loc_i composite attribute value $C_i = \sum_{j=1}^{n} w_j z_{ij}$

of the dummy location is obtained by calculating the data in row $i\,(i = 1, 2, \ldots, n)$ of the normalization matrix A' by the operator.

3.1.2. *Establishment and Normalization of Decision Matrix.* An important part of multi-attribute decision making to solve a specific problem is to build a decision matrix. In this paper, there are m alternative dummy locations for the problem of choosing dummy locations and 5 attributes that affect the security of anonymous sets. Using the set of dummy locations obtained after the initial screening of query probabilities, a decision matrix A is built based on the values of the dummy location loc_i for each attribute, as shown in Table 1.

TABLE 1. Decision matrix

	S_{loc_i}	$d_{(loc_i,loc_j)}$	Lv_{loc_i}	$Sim_{(loc_i,loc_j)}$	De_{loc_i}
loc_1	S_{loc_1}	$d_{(loc_1,loc_j)}$	Lv_{loc_1}	$Sim_{(loc_1,loc_j)}$	De_{loc_1}
loc_2	S_{loc_2}	$d_{(loc_2,loc_j)}$	Lv_{loc_2}	$Sim_{(loc_2,loc_j)}$	De_{loc_2}
loc_3	S_{loc_3}	$d_{(loc_3,loc_j)}$	Lv_{loc_3}	$Sim_{(loc_3,loc_j)}$	De_{loc_3}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
loc_n	S_{loc_n}	$d_{(loc_n,loc_j)}$	Lv_{loc_n}	$Sim_{(loc_n,loc_j)}$	De_{loc_n}

The premise of multi-attribute decision making is that the decision needs to be based on multiple attribute values, which types of attributes are generally benefit, cost, fixed, deviation, and interval, etc. The calculation method for each type is shown below:

1) Benefit type: The larger the attribute value, the better;

$$r_{ij} = \frac{a_{ij} - \min_i a_{ij}}{\max_i a_{ij} - \min_i a_{ij}} \tag{7}$$

2) Cost type: The smaller the attribute value, the better;

$$r_{ij} = \frac{\max_i a_{ij} - a_{ij}}{\max_i a_{ij} - \min_i a_{ij}} \tag{8}$$

3) Fixed type: The smaller the difference between the attribute value and the set fixed value α, the better;

$$r_{ij} = 1 - \frac{a_{ij} - \alpha}{\max_i \mid a_{ij} - \alpha \mid} \tag{9}$$

4) Deviation type: The larger the difference between the attribute value and the set fixed value β, the better;

$$r_{ij} = \mid a_{ij} - \beta \mid - \frac{\min_i \mid a_{ij} - \beta \mid}{\max_i \mid a_{ij} - \beta \mid - \min_i \mid a_{ij} - \beta \mid} \tag{10}$$

5) Interval type: The closer the attribute value is to the set interval $[q_1, q_2]$, the better;

$$r_{ij} = \begin{cases} 1 - \frac{\max(q_1 - a_{ij}, a_{ij} - q_2)}{\max(q_1 - \min_i a_{ij}, \max_i a_{ij} - q_2)}, & a_{ij} \notin [q_1, q_2] \\ 1 \qquad\qquad\qquad\qquad\quad , & a_{ij} \in [q_1, q_2] \end{cases} \tag{11}$$

According to the needs of location privacy protection, we classify the types of various attributes. The higher the POI score of the selected dummy location, the more realistic it is, and the higher score can achieve better protection, so we categorize the POI score as the benefit type. The geographic distances are classified as interval type based on the

minimum and maximum distances of the dummy location from the real location set by the user in the privacy requirements. To resist semantic attacks based on information such as location semantics, the semantics of the selected dummy location and the real location should be as different as possible, i.e., the lower the semantic similarity the more confusing it is, so we categorize the semantic span as cost-based. For location points with higher semantic sensitivity level, users clearly do not want to expose to others, while for location points with lower level users are insensitive, so the semantic sensitivity level is categorized as cost type. For each location point, the more it appears in the mass data set, the lower the location prevalence, and vice versa the higher the location prevalence, we should choose the relatively prevalent location to construct the safe anonymous set, so the location prevalence is categorized as cost type.

The decision matrix A is normalized after determining the type of each attribute, and the normalized decision matrix is shown in Table 2.

Table 2. Normalized decision matrix

	S_{loc_i}	$d_{(loc_i,loc_j)}$	Lv_{loc_i}	$Sim_{(loc_i,loc_j)}$	De_{loc_i}
loc_1	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}
loc_2	r_{21}	r_{22}	r_{23}	r_{24}	r_{25}
loc_3	r_{31}	r_{32}	r_{33}	r_{34}	r_{35}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
loc_n	r_{n1}	r_{n2}	r_{n3}	r_{n4}	r_{n5}

3.1.3. *AHP to Calculate Attribute Weight.* Analytic Hierarchy Process [29] constructs a pairwise comparison matrix by analyzing the relationship between each two attributes, decomposes the problem into different levels according to the goal to be achieved, and then determines the weight of each attribute reasonably through comparison and calculation. In the construction of the AHP-based location privacy desensitization system in this paper, there are four main steps as follows:

Step1: AHP first deals with the problem in layers. It decomposes the problem into different constituent elements according to the nature of the problem and the goal to be achieved, and forms a multi-level analysis structure model according to the subordination and mutual connection between the elements. In this paper, five factors affecting the security of anonymous sets, namely, semantic similarity, location prevalence, semantic sensitivity level, geographic distance, and POI score, are used as the main basis for selecting dummy locations. Figure 2 shows the hierarchical structure model of various attributes mentioned in this paper. The figure establishes three hierarchical structures, namely, the target layer, the criterion layer, and the factor layer.

Step2: AHP requires users to compare the relative importance of each two indicators, and then construct a pairwise comparison matrix according to the quantified value of their importance. Before constructing the pairwise comparison matrix, we define the priority of each attribute according to its impact on the security of the anonymous set: semantic similarity > location prevalence > semantic sensitivity level > geographic distance > POI score. In this paper, we use the 1-9 scale to evaluate the relative importance between the two attributes, as shown in Table 3.

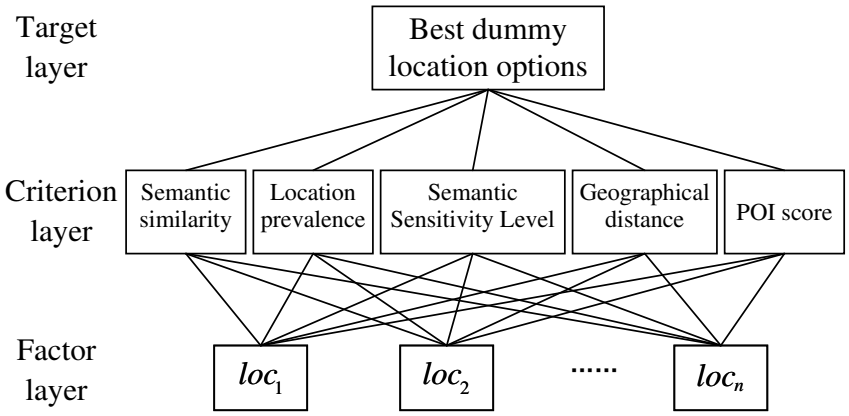

FIGURE 2. Hierarchy diagram

TABLE 3. Index relative importance level table

Quantized value	Meaning
1	The two factors are of equal importance.
3	The former is slightly more important than the latter.
5	The former is obviously more important than the latter.
7	The former is strongly more important than the latter.
9	The former is essential than the latter.
2, 4, 6, 8	The median value of the above adjacent judgments.
$a_{ij} = 1/a_{ji}$	Reciprocal

A pairwise comparison matrix is constructed for the 5 attributes according to Table 3 as follows:

$$B = \begin{pmatrix} 1 & 1/7 & 3 & 1/6 & 1/4 \\ 7 & 1 & 8 & 3 & 5 \\ 1/3 & 1/8 & 1 & 1/7 & 1/5 \\ 6 & 1/3 & 7 & 1 & 3 \\ 4 & 1/5 & 5 & 1/3 & 1 \end{pmatrix}$$

Step3: To ensure the scientificity and reliability of the final result, the consistency check of the matrix is required. The standard of consistency test is when $CR < 0.1$, indicating that the pairwise comparison matrix has passed the consistency test. CR is calculated as

$$CR = \frac{CI}{RI} \tag{12}$$

The calculation of the CI index is shown in Equation (13), and RI is the average random consistency value corresponding to n.

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{13}$$

where λ_{max} is the largest eigenroot of the pairwise comparison matrix, and n is the order of the matrix. The calculation of λ_{max} is shown

$$\lambda_{max} = \sum_{i=1}^{n} \frac{[AW]_i}{nw_i} \tag{14}$$

The maximum characteristic root λ_{max}=5.3364 of the pairwise comparison matrix can be calculated, and then CI=0.0841 can be obtained. According to Table 4, RI=1.12 when n=5. From Equation (12), CR=0.0751 $<$ 0.1 can be obtained, indicating that the judgment matrix has passed the consistency test.

TABLE 4. Value of Stochastic Consistency Indicator RI

n	1	2	3	4	5	6	7	8
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41

Step4: After the matrix passes the consistency check, the attribute weights from w_1 to w_5 calculated according to the pairwise comparison matrix are 0.0584, 0.5013, 0.0341, 0.2692, and 0.1370. Among them, w_1 to w_5 correspond to geographic distance, semantic similarity, POI score, location universality, and semantic sensitivity level in turn.

3.2. **Algorithm Description.** In this paper, we propose the MDMLPD algorithm, which fully considers the influence of several dimensions on the security of dummy locations, and finally constructs a set of optimal secure anonymity set SC containing the real locations of users and of size k. The main flow of the algorithm is shown in Figure 3.

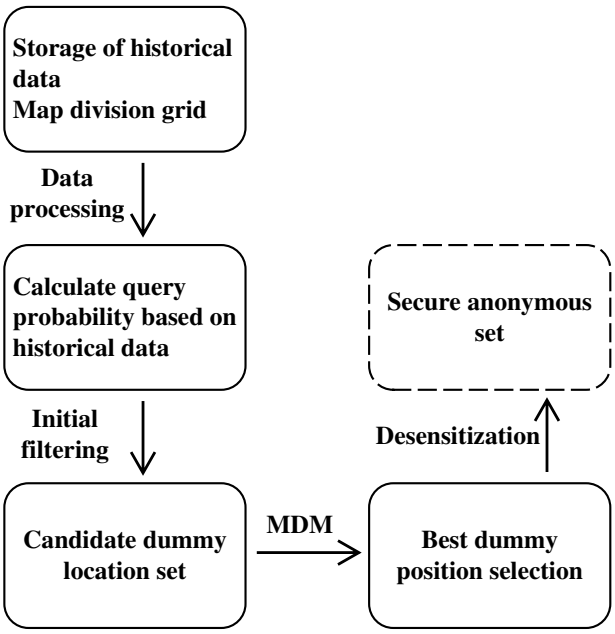

FIGURE 3. Algorithm flow

The pseudo-code of the algorithm in this paper is shown in Algorithm 1. First the user sends a location request and reads the map information (lines 1-2). Then the query probability of each location point in the map is calculated (lines 3-6). For each location in the set of candidate dummy locations, the attribute values of each indicator between them and the real location are calculated (lines 10-12), then the decision matrix is built and

Algorithm1 : MDMLPD

Input : real location loc_{real}, map information MAP, user privacy requirements $req(k, q_1, q_2)$

Output : secure anonymity set SC

1. Send(Q);
2. $n \leftarrow| MAP |$;
3. Receive $req(k, q_1, q_2)$;
4. For $n = 1$ to n do
5. $p_i \leftarrow \frac{N_i}{\sum_{i=1}^{n} N_i}$;
6. End for
7. If $(p_i) \approx (p_{real})$ then
8. Insert loc_{real} into CLS;
9. For each loc_i in CLS do /*Perform dummy location selection*/
10. $Sim_{(loc_1,loc_{real})}$,$Sim_{(loc_2,loc_{real})}$,...,$Sim_{(loc_n,loc_{real})} \leftarrow Sim_{(loc_i,loc_{real})}$;
11. De_{loc_1},De_{loc_2},...,$De_{loc_n} \leftarrow$ TF-IDF;
12. $d_{(loc_i,loc_j)} \leftarrow$ Haversine.set(loc_i);
13. Decisionmatrix.set(A); /*Create a decision matrix*/
14. MaxMinNormalization(x,Max,Min); /*Normalization process*/
15. μ_1,μ_2,...,$\mu_n \leftarrow$ AHP;
16. $C_i = \sum_{j=1}^{n} w_j z_{ij}$
17. QuickSort (CLS,C_1,C_n); /*Sequence*/
18. If $C_1 < C_n$ then
19. q=Partition(CLS,C_1,C_n);
20. QuickSort (CLS,C_1,$q-1$);
21. QuickSort (CLS,$q+1$,C_n);
22. End if
23. End for
24. End if
25. $[C_n] \rightarrow [min]$
26. If $n > C_n$ then /*Output anonymous set*/
27. $DLS \leftarrow C = \{loc_1, loc_2, loc_3, \ldots, loc_{k-1}\}$
28. $m = 1 + (k - 1)$
29. If $C_1 \geq C_n$ then
30. Return SC;
31. Else
32. Apply to generate dummy $D = k - m$, Add the generated dummy to SC;
33. Return SC;
34. End if
35. End if

normalized (lines 13-14), followed by determining the attribute weights by AHP, and then calculating the composite attribute values (lines 15-16). The combined attribute values of the candidate locations are sorted in descending order using fast sorting (lines 18-22). At this point, we have a collection of locations sorted in descending order by composite attribute value (line 25). If the number of locations in the candidate location set is greater than k, the first $k-1$ locations are taken to build and return the anonymous set (lines 29-30). If the number of locations is less than k, then $k-m$ locations are randomly generated together with the candidate location set to build and return the anonymous set (lines 31-33).

4. Experiments and Analysis.

4.1. Experimental Environment.

This paper uses the real dataset GeoLife [30] for experimental simulation. The dataset contains a large amount of trajectory data of different users, including not only the trajectories of daily routines such as commuting to get off work, and shopping, but also the trajectories of outdoor activities in different fields such as catering, medical treatment, fitness and mountaineering. Most of the trajectories in the dataset were created in Beijing, China, which has been fully covered by communication base stations and has a large number of mobile users. Each user's track dataset contains the latitude and longitude of different location points and uses these location point information as the user's history information to calculate the historical query probability.

The data set chosen for the experiment is the location geographic information in a rectangular area of 10km×10km in the central city of Beijing, and the sample space is uniformly divided into 100100 location units. The obtained sample trajectory points are used as historical data, based on which the historical query probability of each location cell after grid division is calculated. In addition, q_2 and q_2 in the user requirements are set to 300 and 1500. The value range of the main parameter k in the experiment is $k \geq 2$.

4.2. Analysis of Results.

First, the experiments analyze the anonymization success rate using MDMLPD algorithm. Then, the experiments evaluate MDMLPD algorithm in four aspects: anonymity set generation time, semantic diversity, attack algorithm recognition probability, and location entropy. It is compared with DLP [12], MMDS [15], and K-DLS [13], which also use dummy location generation techniques.

4.2.1. *Anonymous Success Rate.* Figure.4 shows the variation of the anonymity success rate relative to the number of locations in the map and the anonymity degree k.

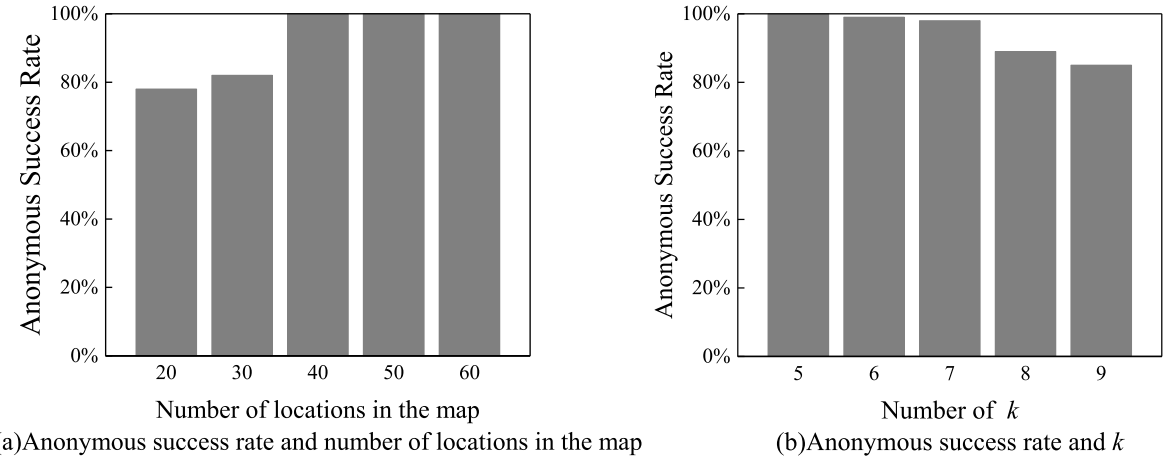

FIGURE 4. Change in anonymity success rate

The results in Figure 4(a) show that the higher the number of locations in the map, the more favorable the anonymity execution and the higher the anonymity success rate. This is because the higher the number of locations, the higher the number of types of location semantics, which is more conducive to constructing pseudo-location sets that satisfy privacy requirements, and the anonymization success rate is improved. In Figure 4(b), the anonymity success rate decreases as the anonymity degree k of privacy needs increases, because the number of locations that satisfy the user's privacy needs is insufficient.

4.2.2. *Anonymous Set Generation Time.* An important metric for constructing a secure anonymous set is the generation time of the anonymous set, and the time overhead of the algorithm in this paper is mainly focused on the computation of multiple metric values. Figure 4 shows the average time of generating anonymous sets for the algorithm in this paper compared with DLP, MMDS, and K-DLS.

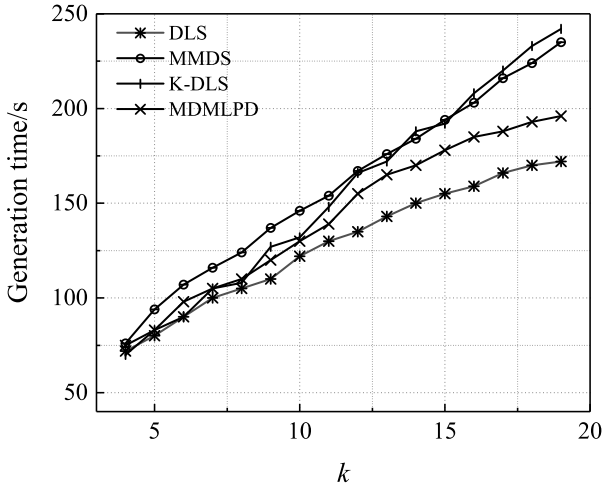

FIGURE 5. Comparison of anonymity set generation time

From Figure 5, it can be seen that the time required to construct the anonymous set increases with increasing k for all four algorithms. When k is small, the generation time of several algorithms is almost the same. With the increase of k, the generation time of the dummy location of DLP is the least, followed by MDMLPD, and the time required by the other two algorithms is higher than that of the algorithm in this paper. Since DLP does not consider semantic information when constructing anonymous sets, it has the least running time. The algorithm in this paper, on the other hand, considers the query probability of each location unit and performs the decision process of semantic similarity, semantic sensitivity level, and other attributes, which is more practical for users and requires a certain amount of time, so its anonymous set generation time is slightly higher than that of DLP.

4.2.3. *Comparison of Semantic Diversity.* The θ-security value proposed in the literature [19] is used to evaluate the semantic diversity of the anonymity set submitted by the algorithm. A larger θ-security value indicates the richer semantic information belonging to each dummy location in the constructed anonymity set and the harder it is for the attacker to determine the real location of the user. Figure 5 shows the semantic diversity of the algorithm in this paper compared with DLP, MMDS, and K-DLS.

As can be seen in Figure 6, the θ-value of MMDS changes minimally and stays high as k increases. The θ-value of the algorithm in this paper can also be maintained at a

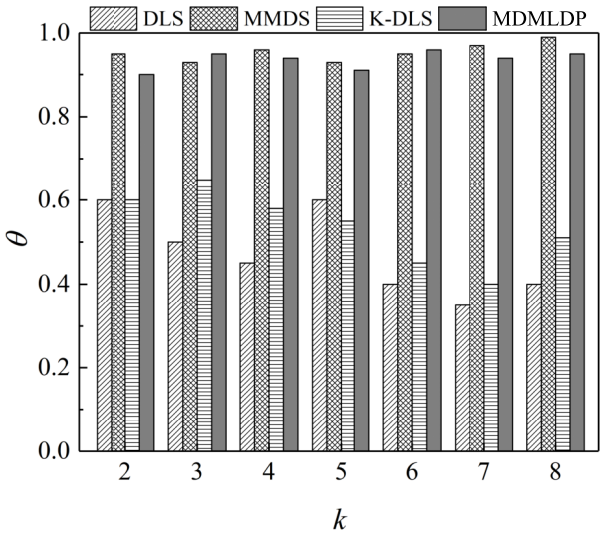

FIGURE 6. Comparison of θ-safe value

relatively high level, which can satisfy the requirement of semantic diversity. However, it is lower than MMDS because MMDS only focuses on the influence of location semantics on anonymity sets, while this paper considers location semantics when constructing secure anonymity sets, but semantic information is not the only criterion for constructing secure anonymity sets. It is obvious that the algorithm in this paper is more practical in different scenarios in real life. The θ-values of both DLP and K-DLS are at a relatively low level, which is due to the fact that they only focus on the query probability when constructing the anonymity set and do not pay attention to the case where the selected dummy location may have the same semantic information as the real location of the user.

4.2.4. *Attack Algorithm Recognition Probability.* Although the user submits a secure anonymity set, there is still a possibility that the user will be identified by the attack algorithm. The probability Q that the user's real location is recognized by the attack algorithm is calculated as follows:

$$Q = \frac{1}{\mid C_{num} \mid} = \frac{1}{k} \tag{15}$$

where C is the anonymous set obtained according to the dummy location generation algorithm, $C = \{l_1, l_2, \ldots, l_n\}$ and $\mid C_{num} \mid = k$.

To compare the advantages of considering attributes such as location, semantic information, and semantic sensitivity for user's real location protection, Equation (15) is used to measure the security of the anonymity set. The probability of being identified under semantic attacks is compared between the algorithm in this paper and DLP, MMDS, and K-DLS, as shown in Figure 6.

From Figue 7, it can be seen that as k gradually increases, the probability of each algorithm being recognized by the attack algorithm gradually decreases, which is due to the fact that the semantic variability among the dummy locations in it gradually decreases due to the expansion of the anonymity set, and the success rate of the attacker using semantic attacks is higher. DLP and K-DLS do not consider the effect of location semantic information on the security of the anonymity set when selecting dummy locations. Therefore, they have poor privacy protection and higher location recognition rate under the attack algorithm than the other two algorithms. Although MMDS pays attention to location semantic information, its attributes such as location sensitivity are not analyzed

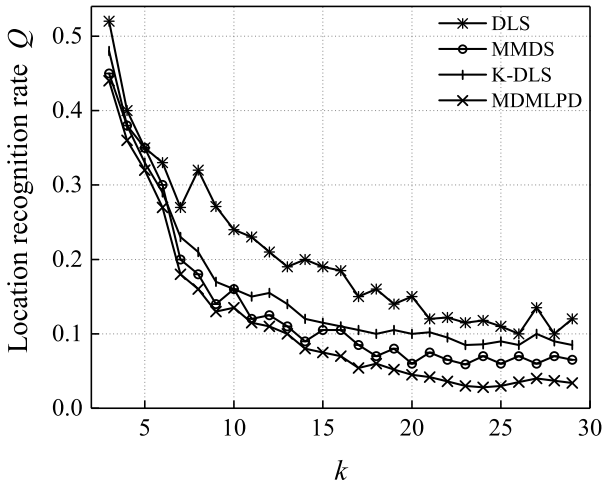

FIGURE 7. Comparison of the identification probabilities of attack algorithms

and the measurement results are not precise enough. The probability of being identified in this paper is slightly lower than the other three algorithms because users consider not only the historical query probability but also the influence of attributes such as location prevalence, semantic information, and different sensitivity levels set by each user on their security when constructing secure anonymous sets. It is difficult for an attacker to increase the probability of speculation even with edge information while conducting semantic attacks, while the algorithm in this paper satisfies the individual requirements for different users in the setting of location sensitivity levels, reflecting a certain degree of personalization.

4.2.5. *Location Entropy.* This subsection adopts location entropy to measure the effectiveness of location privacy protection. The location entropy is positively correlated with the location privacy desensitization effect, and the changes of the two are in the same direction. The larger the location entropy, the stronger the privacy protection effect and the better the desensitization effect. Figure 7 shows the comparison of the location entropy of the anonymity set generated by the method in this paper with DLP, MMDS, and K-DLS under different anonymity degrees.

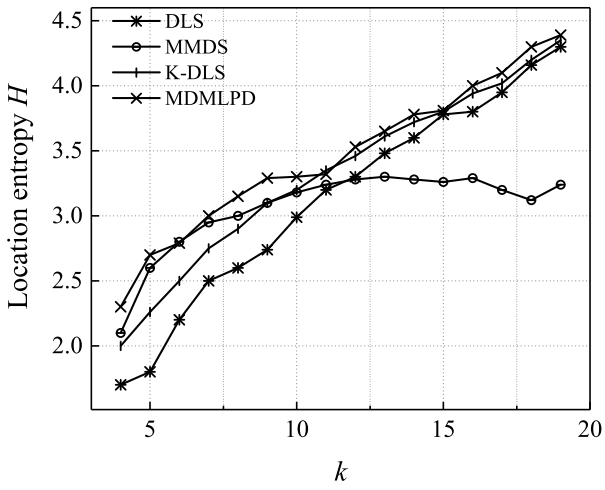

FIGURE 8. Comparison of location entropy

It can be seen from Figue 8 that with the increase of k, the location entropy value of each algorithm is increasing. However, it is not difficult to find that the location entropy of MDMLPD algorithm is higher than that of the other three algorithms. This is because MDMLPD algorithm comprehensively considers the possible impact of various attribute values such as location semantics, location universality, location sensitivity, etc. In this way, the uncertainty of the user's real location is enhanced, and the security of the anonymity set is guaranteed, which is beneficial to the protection of the user's location privacy.

5. **Conclusion.** In this paper, the MDMLPD algorithm integrates five aspects, namely, geographic distance, semantic similarity, POI score, location prevalence, and semantic sensitivity level, to achieve location privacy desensitization. The core idea of this algorithm is to select the optimal dummy location and then build a secure anonymity set by multi-attribute decision model. Firstly, the initial filtering is performed based on the user's historical query probability for each location point; secondly, a hierarchical structure model is established using hierarchical analysis to calculate the weights corresponding to the five attributes; finally, a multi-attribute decision model is used to filter out the optimal dummy locations and construct a more rational and indistinguishable secure anonymous set. The experiments compare MDMLPD algorithm with related algorithms in four aspects, including anonymity set generation time, semantic diversity, attack algorithm recognition probability, and location entropy. The experimental results show that the secure anonymity set constructed using the MDMLPD algorithm has a higher location entropy value, which reduces the risk of user location privacy data leakage and achieves location privacy desensitization. However, in this paper, privacy protection is only studied on the unit placement point, and the change of location over time is not considered. Most of the user locations are presented as trajectories in a short period of time, so the privacy protection of user trajectories will be a subsequent research direction.

Acknowledgment. This work was supported in part by the Key scientific research projects of colleges and universities in Henan Province under Grant 23A520033, the Doctoral Scientific Fund of Henan Polytechnic University under Grant B2022-16 and B2010-32, and the Youth Fund of Henan Polytechnic University(Q2014-05).

REFERENCES

[1] T.-Y. Wu, Q. Meng, L. Yang, X.-L. Guo, and S. Kumari, "A provably secure lightweight authentication protocol in mobile edge computing environments," *The Journal of Supercomputing*, vol. 78, pp. 13893–13914, 2022.

[2] J.-MT. Wu, Q. Teng, S. Huda, Y.-C. Chen, and C.-M. Chen, "A Privacy Frequent Itemsets Mining Framework for Collaboration in IoT Using Federated Learning," *ACM Transactions on Sensor*, 2022.

[3] H.-J. Yang, P. Vijayakumar, J. Shen, and B. Gupta, "A location-based privacy-preserving oblivious sharing scheme for indoor navigation," *Future Generation Computer Systems*, vol. 137, pp. 42–52, 2022.

[4] Y. Yan, F. Xu, A. Mahmood, Z.-Y. Dong, and Q.-Z. Sheng, "Perturb and optimize users' location privacy using geo-indistinguishability and location semantics," *Scientific Reports*, vol. 12, no. 4, 20445, 2022.

[5] H.-B. Jiang, L. Li, P. Zhao, F.-Z. Zeng, Z. Xiao, and A. Iyengar, "Location Privacy-preserving Mechanisms in Location-based Services: A Comprehensive Survey," *ACM Computing Surveys*, vol. 54, no. 4, pp. 1–36, 2022.

[6] T.-Y. Wu, X.-L. Guo, Y.-C. Chen, S. Kumari, C.-M. Chen, "Amassing the Security: An Enhanced Authentication Protocol for Drone Communications over 5G Networks," *Drones*, 6(1), 10, 2022.

[7] V. Stephanie, M. Chamikara, I. Khalil, and M. Atiquzzaman, "Privacy-preserving location data stream clustering on mobile edge computing and cloud," *Information Systems*, vol. 107, 9944460, 2022.

[8] Y. Liu, J. Tian, P. Tian, Y.-M. Du, and S. Li, "A Random Sensitive Area Based Privacy Preservation Algorithm for Location-Based Service," *Wireless Personal Communications*, vol. 119, no. 2, pp. 1179–1192, 2021.

[9] J.-H. Liu, and S.-X. Liu, "All-dummy k-anonymous privacy protection algorithm based on location offset," *Computing*, vol. 104, no. 8, pp. 1739–1751, 2022.

[10] X.-Y. Guo, W.-M. Wang, H.-P. Huang, Q. Li, and R. Malekian, "Location Privacy-Preserving Method Based on Historical Proximity Location," *Wireless Communications and Mobile Computing*, vol. 2020, 8892079, 2020.

[11] I. Natgunanathan, N. Nisha, Y. Xiang, and S. Yi, "Smart-Area-Selection Based Location Privacy Enhancement," *IEEE Systems Journal*, vol. 16, no. 2, pp. 2020–2031, 2022.

[12] X.-Y. Guo, W.-M. Wang, H.-P. Huang, Q. Li, and R. Malekian, "Heterogeneous deniable authenticated encryption for location-based services," *Plos One*, vol. 16, no. 2, e0244978, 2021.

[13] C.-M. Chen, Z. Tie, E.-K. Wang, M.-K. Khan, S. Kumar, and S. Kumari, "Verifiable dynamic ranked search with forward privacy over encrypted cloud data," *Peer-To-Peer Networking and Applications*, 14, 2977-2991, 2021.

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

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

[16] G. Sun, V. Chang, M. Ramachandran, Z.-L. Sun, G.-M. Li, H.-F. Yu, and D. Liao, "Efficient location privacy algorithm for Internet of Things (IoT) services and applications," *Journal of Network and Computer Applications*, vol. 89, pp. 3–13, 2017.

[17] Y. Yang, X.-H. Hu, and Y.-W. Du, "The K-Anonymous Dummy Location Selection Algorithm Based on Historical Query Probability," *Computer Engineering*, vol. 48, no. 2, pp. 147–155, 2022.

[18] T. Hara, "Dummy-based Location Anonymization for Controlling Observable User Preferences," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–7.

[19] J. Wang, C.-R. Wang, J.-F. Ma, and H.-T Li, "Dummy location selection algorithm based on location semantics and query probability," *Journal on Communications*, vol. 41, no. 3, pp. 53–61, 2021.

[20] D.-D. Yang, B.-M. Ye, W.-Y. Zhang, H.-Y. Zhou, and X.-B. Qian, "KLPPS: A k-Anonymous Location Privacy Protection Scheme via Dummies and Stackelberg Game," *Security and Communication Networks*, vol. 2021, 9635411, 2022.

[21] M.-H. Min, W.-H. Wang, L. Xiao, Y.-L. Xiao, and Z. Han, "Reinforcement learning-based sensitive semantic location privacy protection for VANETs," *China Communications*, vol. 16, no. 6, pp. 244–260, 2021.

[22] C.-Y. Yin, X.-K. Ju, Z.-C. Yin, and J. Wang, "Location recommendation privacy protection method based on location sensitivity division," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, 266, 2020.

[23] Y.-B. Jiang, Y.-F. Zhu, X. Du, and T. Jin, "The implicit network inferred from users' residences and workplaces enhancing collaborative recommendation on smartphones," *Physica A: Statistical Mechanics and its Applications*, vol. 535, 122255, 2019.

[24] Z.-P. Gui, Y.-Z. Sun, L. Yang, D.-H. Peng, F. Li, H.-Y. Wu, C. Guo, W.-F. Guo, and J.-Y Gong, "LSI-LSTM: An Attention-aware LSTM for Real-time Driving Destination Prediction by Considering Location Semantics and Location Importance of Trajectory Points," *Neurocomputing*, vol. 440, pp. 72–88, 2021.

[25] W.-H. Li, C. Li, and Y.-L. Geng, "APS: Attribute-aware privacy-preserving scheme in location-based services," *Information Sciences*, vol. 527, pp. 460–476, 2020.

[26] J.-J. Wang, Y.-L. Han, X.-Y. Yang, T.-P. Zhou, and J.-Y Chen, "A new group location privacy-preserving method based on distributed architecture in LBS," *Security and Communication Networks*, vol. 2019, UNSP 2414687, 2019.

[27] C. Nunez-Gomez, and V. Garcia-Font, "HyperNet: A conditional k-anonymous and censorship resistant decentralized hypermedia architecture," *Expert Systems with Applications*, vol. 208, 118079, 2022.

[28] W.-J. Wang, J.-M. Zhan, and J.-S. Mi, "A three-way decision approach with probabilistic dominance relations under intuitionistic fuzzy information," *Information Sciences*, vol. 582, pp. 114–145, 2021.

[29] K.-D. Goepel, "Comparison of judgment scales of the analytical hierarchy process—A new approach," *International Journal of Information Technology & Decision Making*, vol. 18, no. 2, pp. 445–463, 2019.

[30] J. Li, X. Pei, X.-J. Wang, D.-Y. Yao, Y. Zhang, and Y. Yue, "Transportation Mode Identification with GPS Trajectory Data and GIS Information," *Tsinghua Science and Technology*, vol. 26, no. 4, pp. 403–416, 2021.