Hindawi
Complexity
Volume 2021, Article ID 6674639, 10 pages
https://doi.org/10.1155/2021/6674639

Research Article

KPDR : An Effective Method of Privacy Protection

Zihao Shen,[1,2] Wei Zhen,[1] Pengfei Li,[1] Hui Wang,[1] Kun Liu,[1] and Peiqian Liu[1]

[1]*School of Computer Science and Technology, Henan Polytechnic University, Jiao'zuo 454000, China*
[2]*College of Computer Science and Technology, Jilin University, Chang'chun 130012, China*

Correspondence should be addressed to Hui Wang; wanghui_jsj@foxmail.com

Received 12 November 2020; Revised 28 December 2020; Accepted 29 January 2021; Published 9 February 2021

Academic Editor: Yongsheng Hao

To solve the problem of user privacy disclosure caused by attacks on anonymous areas in spatial generalization privacy protection methods, a K and P Dirichlet Retrieval (KPDR) method based on k-anonymity mechanism is proposed. First, the Dirichlet graph model is introduced, the same kind of information points is analyzed by using the characteristics of Dirichlet graph, and the anonymous set of users is generated and sent to LBS server. Second, the relationship matrix is generated, and the proximity relationship between the user position and the target information point is obtained by calculation. Then, the private information retrieval model is applied to ensure the privacy of users' target information points. Finally, the experimental results show that the KPDR method not only satisfies the diversity of $l(3/4)$, but also increases the anonymous space, reduces the communication overhead, ensures the anonymous success rate of users, and effectively prevents the disclosure of user privacy.

1. Introduction

Thanks to the emergence of mobile terminal equipment and the rapid development of location service systems, great changes have occurred in our lives, and people can buy their favourite products without leaving home. There are Taobao for dressing, Meituan for eating, Flying Pig for lodging, Didi for travelling, and strips for travelling. People can get services anytime and anywhere through various apps [1], all of which are derived from the rapid development of Location-Based Services (LBS). According to statistics [2], the global market share of LBS and Real-Time Location Systems (RTLS) will increase from 11.36 billion in 2015 to 54.95 billion US dollars in 2020, and the Compound Annual Growth (CAGR) will be 37.1%.

LBS [3, 4] refers to providing various value-added services for mobile users based on the location information of mobile devices and the information transmission of communication networks. However, as people's demand for services increases, location service providers (LSP) may leak users' privacy to criminals for their own benefit, which will threat users' property and personal safety [5]. Therefore, protecting user privacy while providing users with convenient services has become an urgent problem to be solved [6].

In the aspect of location privacy protection, spatial generalization technology based on k-anonymity has always been a hot spot for scholars. Its core idea is to generalize the real location of users and ensure that there are at least K-1 users in the anonymous area (ASR), so that LSP cannot distinguish real users from K users. At present, there are many researches on privacy protection technology based on k-anonymity [7, 8]. For example, Li et al. [9] introduced a credit mechanism on the basis of k-anonymity and set a threshold for users. When the user's credit is higher than this threshold, they can participate in the formation of k-anonymity to obtain privacy protection.

To resist the attack against ASR, Zheng et al. [10] proposed an outlier elimination clustering algorithm based on k-anonymous model; the algorithm optimized the distribution of users in anonymous groups by taking anonymous groups as the center instead of users' positions, but the anonymous areas formed were larger than the actual needs, and in many cases, the probability of attackers identifying query requesters was much higher than $1/k$. Wang et al. [11] proposed differential private K-valued method (DPKA) combined with the concept of difference privacy and k-anonymity and proposed a method for its realization. This method, however, does not consider the effect of $l(3/4)$

diversity on k-anonymity, which is vulnerable to continuous query attacks. Literature [12] proposed a k-anonymity algorithm based on the analytic hierarchy process; in the clustering process, the method always selects the record with the smallest distance to add and individually controls the clustering according to the K value to achieve the equivalent class, but when the k-anonymity area formed in densely populated places is small, the attacker can still infer the approximate location of the user. It can be seen that the process of generating anonymous regions from the anonymous space is the most vulnerable to attack by attackers.

To solve the above problems, this paper proposes a privacy protection method of KPDR based on Dirichlet graph model, which can protect users' privacy from location and query. In the aspect of protecting location privacy, k-anonymity random location hiding method based on Dirichlet graph model is adopted to ensure the security of ASR and satisfy the diversity of location l(3/4). Therefore, the probability of users being identified by attackers is less than $1/k$. In terms of protecting query privacy, due to the particularity and unreliability of LBS, attackers have a high probability of inferring the user's sensitive information according to the user's query points and causing privacy leakage. In this paper, the private information retrieval (PIR) technology with relatively high security [13, 14] is adopted, which can ensure that the trusted third-party server (TTPS) can securely retrieve the desired data from the untrusted LBS server and effectively prevent the privacy disclosure caused by the attack of LBS.

2. Propaedeutics

2.1. System Architecture. With the change of problem background and attack model, location privacy will continue to face new challenges. For example, LBS servers are vulnerable to attacks, and the risk of sensitive attribute disclosure exists objectively on the premise that LBS operators cannot be fully trusted. To ensure user privacy and service quality, this method introduces a trusted third-party server, and a trusted third-party center structure is composed of a mobile terminal, a trusted third-party server, and an LBS server, as shown in Figure 1. In the KPDR privacy protection method, the trusted third-party server and the LBS server jointly maintain a set of information points. The mobile terminal sends the query request information to the trusted third-party server, which generates Dirichlet graph according to the user query request information and its own cached information points and selects the false positions of the virtual user and the current user in K–1 D blocks according to the established rules to form a user anonymous set and send it to the LBS server. After obtaining the user set, the LBS server generates a relationship matrix according to the proximity relationship between the user target information points and k users. After that, the trusted third-party server retrieves the target information from the relational matrix and returns it to the user, which is a complete request. In the whole process of privacy protection, the data centralization is completed by using the trusted third party as the total carrier, and the privacy security requirements of users can be met as long as the security of the trusted third-party server is ensured.

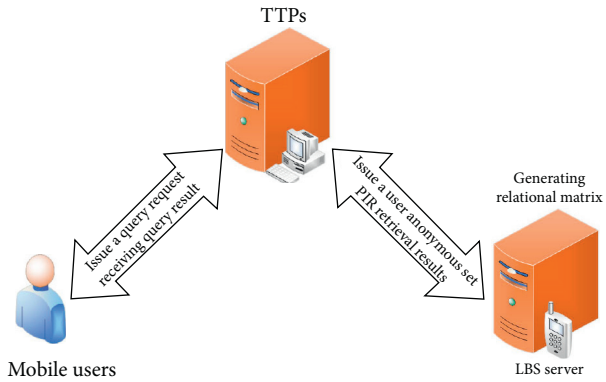

Figure 1: Communication model.

2.2. Related Definitions

Definition 1. Point of information (POI) is as follows:

$$\mathrm{POI}\left(\mathrm{S}_{\mathrm{ig}}, C_{\mathrm{la}}, l_{\mathrm{at}}\right), \tag{1}$$

where S_{ig} stands for the unique name identification of the information point, C_{la} represents the category of the information point, and l_{at} represents the coordinate information of the information point, and the introduction of information points is to enhance the ability to query and describe the user's location and improve the query efficiency.

Definition 2. Dirichlet graph is as follows: let set $D = \{D_1, D_2, D_3, \ldots, D_n\}$ be a set of n information points on the plane, where

$$\forall D_i, D_j \in D, E_{\mathrm{ve}} \in V_i \mid \mathrm{S}\left(E_{ve}, D_i\right) < \mathrm{S}\left(E_{ve}, D_j\right), \quad i \neq j, \tag{2}$$

is the Dirichlet diagram, in which $\mathrm{S}(E_{ve}, D_j)$ is the Euclidean distance from point E_{ve} to point D_i, E_{ve} is any point in D block, and V_i is any single polygon in Dirichlet graph, which is called D-block as shown in Figure 2.

The feature of Dirichlet graph is that there is a focus in each D block, and the distance from the inner point of each D block to this focus is smaller than that from other D blocks, such as $S(E_{ve}, D_1) < S(E_{ve}, D_j)$, $j \neq 1$. The distance from the point on the boundary of block D to the focus that generates this boundary is equal; by using the characteristics of Dirichlet graph, the trusted third-party server can find the nearest information point to the user more quickly after receiving the user query request, which is more efficient than$K$ (3/4)NNalgorithm.

Definition 3. Client request is as follows: $U_{\mathrm{client}}(I_{\mathrm{dent}}, L_{\mathrm{oc}}, C_{\mathrm{la}}, U_{\mathrm{time}}, \lambda)$ represents the request information sent by the user's mobile terminal. The field I_{dent} represents the unique identification number of the user; the field L_{oc} represents the position when the user initiates the

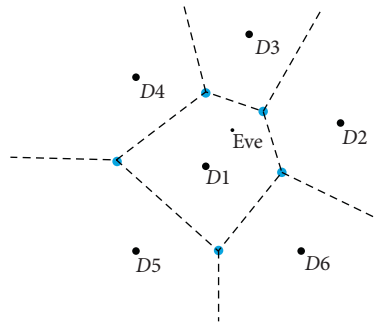

Figure 2: Block D in the Dirichlet diagram.

request; C_{la}stands for POI category; U_{time} represents the time point when the user sends the request; λrepresents a large odd prime number.

Definition 4. User anonymous set is as follows: Users$(ID_{\text{gate}}, C_{\text{la}}, C_{\text{time}}, L, X_{\text{inf}})$, where ID_{gate} represents the unique identification number of user anonymous set; C_{la}represents the information point category of user anonymous set; C_{time} indicates the time when the anonymous set of users is sent out; the query set $X_{\text{inf}} = \{x_1, x_2, \ldots, x_u, \ldots, x_k\}$ has quadratic residuals of K−1 modular$\lambda$and quadratic nonresidual x_u of one modularλ; the location set $L = \{L_1, L_2, \ldots, L_u, \ldots, L_k\}$ represents the location of each user in the user anonymous set, where L_u represents the random false location in the D block to which the real user belongs. The position parameter L must satisfy the following equation:

$$\exists L_i \in D_i, \forall L_j \notin L_i, \quad i \neq j, \tag{3}$$

where parameter D_i represents the D block in the Dirichlet diagram. In this way, when the trusted third-party server sends the user anonymity set to the LBS server, the user location in the user anonymity set must be randomly selected from different D blocks.

3. Privacy Protection Method of KPDR Based on K-Anonymous

3.1. Location Hiding Algorithm Based on Dirichlet Graph Model

3.1.1. Dirichlet Construction Based on POI. Before the privacy protection process starts, the LBS server keeps the Dirichlet graph based on the same category information points in TTPS and LBS servers synchronously. As shown in Algorithm 1, taking the information point as the base point, the Delaunay Triangulation Algorithm is used to generate the triangulation and then determine the circumscribed circle center of each triangle in the triangulation and finally connect the adjacent circle centers to construct the Dirichlet diagram model.

From the above algorithm, we can see that the algorithm complexity calculation of Algorithm 1 is divided into four parts. The first part is to construct Delaunay triangular network with the complexity of $O(n^2)$. The second part computes the center of the triangle peripheral circle, and the complexity is $O(n)$. In the third part, the complexity of finding triangles with three adjacent sides is $3O(n)$. The fourth part draws the Dirichlet diagram; the complexity is $O(n)$. Therefore, the algorithm complexity of generating Dirichlet graph focusing on POI of the same kind is $O(n^2) + O(n) + 3O(n) + O(n) = O(n^2)$.

As shown in Figure 3, each polygon represents a D block, and the focus of each D block is the information point. When the trusted third-party receives the request sent by the user, it will divide the corresponding Dirichlet graph according to the position L_{oc} in the mobile terminal request U_{client}.

3.1.2. The Processing of TTP Server to the User Sending Service Request U_{client}. When the trusted third-party server receives a user request for U_{client}, it will first determine whether the request is initiated by the same user again according to the unique user identification number ID_{gate} in U_{client}. If it is the first time, the trusted third-party server will determine the rule, generate an anonymous set of users, and send it to the LBS server. If it is not initiated for the first time, the trusted third-party server will calculate according to the two positions L in the latest service request information U_{client} sent by the user; when the latest user position L has left the last D block, it will regenerate the latest user anonymous set and send it to the LBS server. Updating user anonymous sets in time can effectively prevent joint attacks and location inference attacks. If the location L sent by the user multiple times is the same as the location sent for the first time, then the trusted third-party server will form a time series set according to the time U_{client} initiated in U_{client} each time the user requests $U_{\text{client}} = \{t_1, t_2, \ldots, t_n\}$; calculate the value of ε:

$$\varepsilon(U_{\text{time}}) = \left(\sum_{k=1}^{n} \left| t_{k+1} - t_k \right|^2 \right)^{(1/2)}. \tag{4}$$

Supposing the normal load of the trusted third-party server is∂, when $\partial \geq \varepsilon$, the trusted third-party server will regenerate the user anonymity set and send it to the LBS server; otherwise, it will directly send the last generated user anonymity set. This method reduces the load pressure of trusted third-party servers to a certain extent. At the same time, when the server load is low, updating the user anonymity set with high frequency can effectively resist continuous query attacks and associated attacks and enhance anonymity.

3.1.3. Generation Rules of TTP Server for User Anonymous Set Users. After receiving the user request, the trusted third-party server will first save all information in U_{client} according to the identification number ID_{gate} in U_{client} requested by the user and then find the nearest information point from the server cache according to the real location L of the user in U_{client} to generate the Dirichlet diagram of the same information point. Then, according to the location L of the real user, a fake location point is randomly selected from the D block to which it belongs to replace the real user location, and K−1 fake location points are selected from different D

Input: POI List
Output: Dirichlet diagram focusing on POI
(1) Initialize the *triangle list*
(2) Determine the super triangle
(3) Add super triangle vertices to the end of the *POI List*
(4) Add the super triangle to the *triangle list*
(5) **for** each sample point in the *POI List*
(6) Initialize the *edge buffer*
(7) **for** each triangle currently in the *triangle list*
(8) Calculate the triangle circumcircle center and radius
(9) **if** the point lies in the triangle circumcircle then
(10) Add three triangle edges to the *edge buffer*
(11) Remove the triangle from the *triangle list*
(12) **endif**
(13) **end for**
(14) Delete all doubly specified edges from the *edge buffer*, this leaves the edges of the enclosing polygon only
(15) Add to the *triangle list* all triangles formed between the point and the edges of the enclosing polygon
(16) **end for**
(17) Remove any triangles from the triangle list that use the super triangle vertices
(18) Remove the super triangle vertices from the *POIList*
(19) Connect and get Dirichlet

Algorithm 1: Dirichlet construction based on POI.

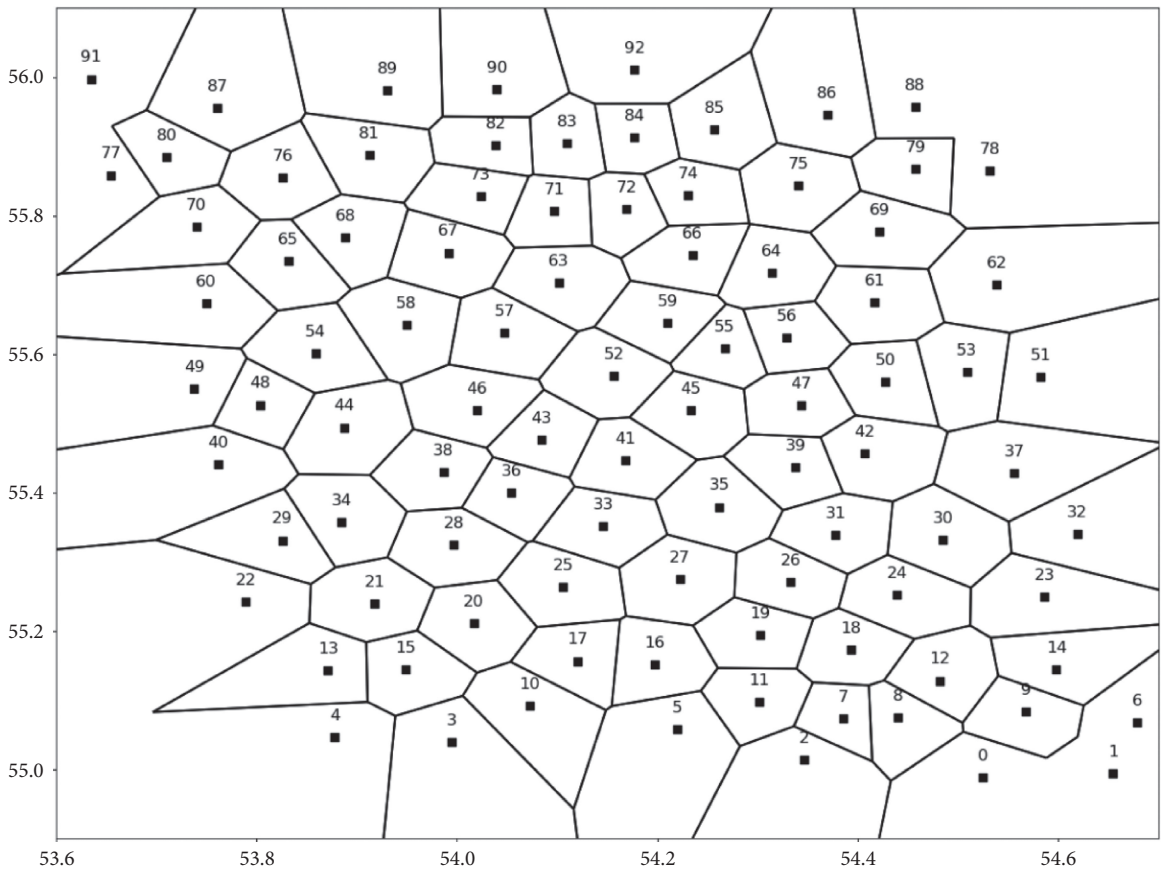

Figure 3: Dirichlet diagram.

blocks in this Dirichlet diagram in a fully random manner. A total of K false users belonging to different D blocks are generated to form a user anonymity set and sent to the LBS server, where the information point category C_{la} in the user request U_{client} is equal to the C_{la} in the user anonymity set Users.

The use of Dirichlet graph divides the continuous anonymous space into discrete D blocks; compared with the anonymous interval generated by the traditional K-anonymous method, it achieves the function of dividing the space. The advantage of this is that this method of randomly dividing the space will enhance anonymity and protect the security of ASR; the second is that it will not affect the quality of service while replacing the real location of the user with a fake location; the third is that it effectively avoids adding noise in the privacy protection method of the United States, the risk of privacy leakage caused by the impractical location of the noise.

3.2. Query Privacy Protection Algorithm Based on Private Information Retrieval. The function of the private information retrieval technology is to ensure that the private information of the retrieval initiator will not be leaked when the information retriever initiates a retrieval to the server. There are currently two mainstream private information retrieval methods: one is private information retrieval based on information theory, and the other is private information retrieval based on computational theory. The private information retrieval method based on information theory needs to send all service information from the LBS server to the mobile terminal. Although the user service quality and the absolute security of user privacy are guaranteed, the transmission cost is too large and still only stays at the theoretical level. Therefore, the current mainstream use of private information retrieval methods is based on computing power.

The problem model based on the intractable quadratic residue hypothesis is a common method of private information retrieval technology based on computational theory. In the private information retrieval protocol of quadratic residue model, the server generates a relational matrix from the data in the database, and the retrieval target of the trusted third-party server is one bit of data in the matrix. The mobile terminal initiates a query to the server according to its own private information. When the LBS server receives the query information, it performs modular multiplication on each row of elements in the matrix to obtain a query result and then returns the query result to the mobile terminal to complete a retrieval.

After the LBS server receives the user anonymity set Users, it will generate a relationship matrix according to the information point category C_{la}; this relationship matrix contains the proximity relationship between n information points and the K user positions.

Definition 5. Quadratic residue is as follows: set a large odd prime number $\lambda > 1, \mu \in Z$, and $1 \le \mu \le \lambda, (\mu, \lambda) = 1$; for the basic quadratic congruence $\Upsilon^2 \equiv \mu (\mathrm{mod}\lambda)$, if there is $\Upsilon \in Z$ that satisfies this congruence, then it is said that μ is a quadratic residue modulo λ; otherwise, it is called quadratic nonresidual.

Definition 6. $\lambda$is a large odd prime number, $(\mu, \lambda) = 1$, by the quadratic residue Euler discriminant conditions which are as follows:

(i) $\mu$is the necessary and sufficient condition of the quadratic residue modulo λas $\mu^{(\lambda-1/2)} \equiv 1 (\mathrm{mod}\lambda)$.

(ii) μ is the necessary and sufficient condition of quadratic nonresidual modulo λas $\mu^{(\lambda-1/2)} \equiv -1 (\mathrm{mod}\lambda)$.

Definition 7. The relation matrix generated in LBS server is $A_k^n = \begin{pmatrix} NR_{11} & \cdots & NR_{1k} \\ \cdots & O & \cdots \\ NR_{n1} & \cdots & NR_{nk} \end{pmatrix}$, in which $NR_{ij} \in \{0, 1\}$, the values of NR_{ij} represent the proximity relationship between the ith POI in the relation matrix and the jth user in the query set X_{inf}, 1 represents proximity, 0 represents alienation, n represents the number of information points, and k is the number of users in anonymous set.

Definition 8. $X_{inf} \otimes A_k^n = \Psi = \{f(\varphi) | f(\varphi) = \prod_{m=1}^{k} x_m \cdot NR_{\varphi m}, x_m \in X_{inf}, NR_{\varphi m} \in A_k^n\}$, in which X_{inf} is the query set $X_{inf} = \{x_1, x_2, L, x_u, L, x_k\}$in the anonymous set Users and $\mathbf{A_k^n}$ is the relational matrix.

Definition 9. $h(m, \varphi) = x_m a_{\varphi m}$, where $x_m \in X_{inf}, NR_{\varphi m} \in A_k^n$. If and only if $h(m, \varphi)$result is 0, it is recorded as 1.

Definition 10. λ is a large odd prime number, $(\mu, \lambda) = 1$. Legendre symbol is defined as follows:

$$\left(\frac{\mu}{\lambda}\right) = \begin{cases} 1, & \text{If } \mu \text{ is a quadratic residue of modular } \lambda, \\ -1, & \text{If } \mu \text{ is a quadratic non residue of modular } \lambda. \end{cases} \tag{5}$$

In the quadratic residue theory, the attacker cannot figure out whether μ is a quadratic residue modulo λ without a given factorization of a large odd prime number λ.The trusted third-party server calculates the quadratic residue of K-1 module $\lambda$and the quadratic nonresidual x_u of one module in advance according to the large odd prime number in the user request R to form the query set X_{inf}, send it to LBS server, and x_u correspond to the real user to be queried. After the LBS server receives the user anonymity set sent by the trusted third-party server, it generates the relationship matrix $\mathbf{A_k^n}$ according to the type of information point C_{at} and performs the $X_{inf} \otimes A_k^n$ operation. Because the LBS server cannot identify the secondary nonresidual in X_{inf}, it returns the result set Ψ to the trusted third-party server.

According to Definition 7, it can be seen that the relationship matrix $\mathbf{A_k^n}$ records the neighbor relationship between K users and n information points, and the returned result set $\Psi$is a set composed of $f(\varphi)$. When μ and ν are quadratic residuals of modulo λ, it can be seen from

Definition 5 that $\Upsilon^2 \equiv \mu(\mathrm{mod}\lambda)$ is equivalent to $\Upsilon^2 \equiv \nu(\mathrm{mod}\lambda)$, and Definition 10 has $(\mu/\lambda) = (\nu/\lambda)$. Available from Definition 6, $(\mu/\lambda) \equiv \mu^{(\lambda-1/2)}(\mathrm{mod}\lambda)(\nu/\lambda) \equiv \nu^{(\lambda-1/2)}(\mathrm{mod}\lambda)$, $(\mu\nu/\lambda) \equiv (\mu\nu)^{(\lambda-1/2)}(\mathrm{mod}\lambda)$. So, there are $(\mu\nu/\lambda) \equiv (\mu\nu)^{(\lambda-1/2)} = \mu^{(\lambda-1/2)}\nu^{(\lambda-1/2)} \equiv (\mu/\lambda)(\nu/\lambda)(\mathrm{mod}\lambda)$, and because the Legendre symbol in Definition 10 has a value range of ±1 and λ is a large odd prime number, there is $(\mu\nu/\lambda) = (\mu/\lambda)(\nu/\lambda)$.

To sum up, there are inferences: when λ is a large odd prime number, μ and ν are relatively prime to λ; if both μ and ν are quadratic residues modulo λ, then $\mu\nu$ is also a quadratic residue modulo λ; if one of μ and ν is a quadratic residue of modulo λ, and the other is a quadratic nonresidual of modulo λ, then $\mu\nu$ is a quadratic nonresidual modulo λ. In the result set Ψ, we have the following.

When $f(i)$is a quadratic nonresidual of module λ, it shows that $h(u, i) = x_u \cdot NR_{iu} = x_u$; that is,$NR_{iu} = 1$; that is, the user to be queried is adjacent to the ith POI.

When $f(i)$ is still the quadratic residue of module λ, it shows that $h(u, i) = x_u \cdot NR_{iu} = 1$; that is, $NR_{iu} = 0$; that is, the user to be queried is distant from the ith POI.

According to the result set Ψreturned by the LBS server, the trusted third-party server can obtain the proximity relationship between the real user and each information point. After determining the proximity relationship, the user can be guided to the next step.

The mainstream privacy protection strategy based on an independent architecture is to send the processed data information to the LBS server to ensure that the user's private information will not be leaked. However, when the user has high requirements for service quality, the LBS server can only send processed data providing service, and such service quality is at a loss. The application of private information retrieval technology solves the problems of information loss caused by factors such as the complexity of the network environment and the uncertainty of user behavior.

4. Discussion on K Values in KPDR

In the traditional K-anonymous privacy protection method, the user's privacy protection degree and service quality are affected by the K value. When the value of K is larger, the degree of privacy protection of the user is higher, and the quality of service is lower; when the value of K is smaller, the quality of service of the user is higher, but the user is susceptible to link attacks and privacy leakage. Therefore, choosing a K value that can balance the user's service quality, and the degree of privacy protection is the key to the traditional K-anonymous privacy protection method.

In the KPDR method, the selection of the K value is slightly different. The larger the value of K, the larger the user's anonymity set, and the higher the user's privacy protection. However, because of the application of private information retrieval technology to protect query privacy, the user needs to traverse the entire relationship matrix for each query, so that the user's request service efficiency will be affected; the smaller the value of K, the smaller the user anonymity set, the faster the traversal speed of the relationship matrix, and the higher the quality of service provided to users. Therefore, the degree of user privacy protection, service request efficiency, and server computing power are all related to the value of K. With the rapid development of the computer industry, the computing power of the computer has been significantly enhanced, which is enough to cope with the calculation amount of K taking a larger value. However, if K takes a very large value or the amount of concurrent user query requests is particularly high, the server still using this query will fail because of insufficient computing power and downtime or too long computing time.

It is assumed here that the computing power of the computer is unlimited. Given μ and ν in Definition 5, whenμ is the quadratic residue of modulo λ,μ takes one of the series: $(-(\lambda-1/2))^2, (-(\lambda-1/\lambda-1)+1)^2, \ldots, (-1)^2, (1)^2, \ldots, ((\lambda-1/2)-1)^2, (\lambda-1/2)^2$; the simplified residue system with the smallest absolute value of modulo λ is $-(\lambda-1/2), -(\lambda-1/2)+1, \ldots, -1, 1, \ldots, (\lambda-1/2)-1, (\lambda-1/2)$. Because $(-\alpha)^2 = \alpha^2$,μ is the quadratic residue of modulo λ if and only if the value is one of $(1)^2, \ldots, ((\lambda-1/2)-1)^2, (\lambda-1/2)^2$. And because when $1 \le \alpha \le \beta \le (\lambda-1/2)$,$\alpha^2 \equiv \beta^2(\mathrm{mod}\lambda)$, so all quadratic residuals of modulo λ are $(1)^2, \ldots ((\lambda-1/2)-1)^2, (\lambda-1/2)^2$, a total of $(\lambda-1/2)$. Thus, there are $(\lambda-1)-(\lambda-1/2) = (\lambda-1/2)$ quadratic nonresidues of module λ.

Because the trusted third-party needs to choose K-1 quadratic residue of modulo λ and a quadratic nonresidual of modulo λ, the value of K needs to satisfy $K \le (\lambda + 1/2)$; because each query needs a quadratic nonresidual of modulo λ, $2 \le K$ needs to be satisfied.

To sum up, the value of K is related to the computing power of the computer and positively correlated with the degree of privacy protection, and the theoretical value of K is $2 \le K \le (\lambda + 1/2)$.

5. Security Analysis

With the increasing number of users using LBS service, criminals have increased attacks on users' privacy. This section will analyze the security of KPDR method in the face of various attacks.

5.1. Resist Attacks Based on Geographic Location Information. The attack based on geographic location information is mainly due to the incompleteness of privacy protection technology, which leads to many unrealistic false positions in the generated ASR. When criminals find that a large number of false locations are distributed among lakes and cliffs, these locations can be easily excluded, which increases the probability of the user's true location leaking. The KPDR method based on the ASR generated by the actual POI can resist this attack method, because the actual POI position will not be in the lake or cliff, and if the V block generated based on the POI contains similar lakes, the KPDR method only one false location will be distributed in the area, avoiding the generation of a large number of invalid false locations, and the impact on the leakage probability of the user's true location is almost zero.

5.2. Resist Inference Attacks Based on User Background Knowledge. Attacks based on users' information background knowledge refer to privacy attacks launched by attackers on the basis of mastering users' basic information, such as interests and habits. When the KPDR method responds to the request service initiated by the user, the user request is divided into a Dirichlet graph each time, and the type of user request is different, and the generated Dirichlet graph will be different. In the entire privacy protection process, the user's basic information is never exposed, the attacker cannot infer the user's requested service information, and the KPDR method can resist such attacks very well.

5.3. Resist Continuous Multiquery Attack. Continuous multiquery attack means that when a user continuously requests a service for a period of time, the attacker infers the next position of the user according to the current moving speed of the user and the generated ASR results. In KPDR, when the user makes a continuous query, TTPs will judge the user's position every time. Every time the user initiates a query, new false information will be regenerated according to the new V block. Every false information and ASR update make it impossible for the attacker to analyze any information of the user in time. Therefore, KPDR method has a good effect on the attack of continuous multiquery and effectively protects the privacy of the user.

5.4. Resist Monitor Attacks by Attackers. Monitoring attacks are mainly aimed at privacy protection methods using distributed point-to-point architecture. In this architecture, users spontaneously form anonymous groups through P2P protocol, and attackers can impersonate ordinary users to participate in anonymous group construction. If attackers monitor users' requests in anonymous groups, they can monitor users' private information by analyzing the returned results. The difference is that KPDR adopts the trusted third-party center architecture and TTPs as the overall carrier to complete centralized data processing. Users do not communicate or interfere with each other when requesting services, and attackers cannot listen to any request information from other users.

6. Experimental Results Analysis

The experiment makes a detailed comparison between the KPDR method proposed in this paper and the privacy protection method (GRAM), which is also based on the principle of K- anonymity. The GRAM [15] method constructs a protection graph that satisfies the anonymity requirements of (k, l) identifies vertices in the protection graph, satisfies users' privacy requirements by constantly adding vertices and edges, alleviates the contradiction between privacy protection and quality of service, and has some advantages over traditional k-anonymity methods, but because the GRAM method cannot rule out all redundant edges in the process of adding vertices. It not only reduces the efficiency of anonymity, but also has some shortcomings. This paper will analyze the difference, advantages, and disadvantages between KPDR and GRAM through experiments. Because GRAM has carried out data experiments with the traditional k-anonymity method in terms of security and efficiency, this paper will not repeat it in the data comparison but will explain it in the theoretical analysis.

6.1. Analysis of Computing and Communication Overhead

6.1.1. Computational Overhead Analysis of KPDR Method. In the KPDR method, based on the POI data in the geographic information system, different kinds of POIs are generated into Dirichlet diagrams by using Delaunay Triangulation Algorithm. Considering that the update of POI data in real life is not frequent, the strategy of sacrificing storage space is adopted to reduce the computing overhead of the server. The Dirichlet diagrams divided by different kinds of POI are stored in TTPs in advance, and when updating the POI data, only the Dirichlet diagrams generated by the corresponding POI categories need to be recalculated, which greatly reduces the computational overhead of TTPs. In the process of privacy protection, using Dirichlet graph to segment the interval, TTPs need to traverse the proximity relationship between POI and users, and the complexity is $O(n)$. Although the computational overhead increases linearly, combined with the classification of POI before, the search cardinality has been greatly reduced, which improves the computational efficiency of TTPs and reduces the computational overhead on the premise of ensuring security. When selecting the false position of the user, the calculation cost is related to the value of K; because of the characteristics of the Dirichlet graph, the nearest neighbor calculation is not required. Although the calculation cost will increase with the increase of K, the overall cost will not be generated with excessive changes.

6.1.2. Analysis of Communication Cost of KPDR Method. In the KPDR method, Dirichlet graphs divided according to different types of POI are jointly maintained by LBS and TTPs. LBS accepts user anonymity data packets and responds to user requests. Therefore, the size of communication overhead is related to the speed of LBS processing user requests, especially in the face in the case of multiple users and high concurrency; the throughput of LBS directly affects the quality of service for users. In the process of forming an anonymous set, the communication overhead increases with the increase of the size of the anonymous set. Due to the use of the quadratic residue hypothesis model in this paper, LBS accepts the generation of the relation matrix of the user anonymous set, although the proximity relationship between two POI and K users is recorded in the relation matrix; the TTPs does not need to index all proximity relationships; it only needs to retrieve a neighbor relationship between the user and the POI. This not only reduces the communication overhead to a certain extent, but also ensures that the overall communication overhead will

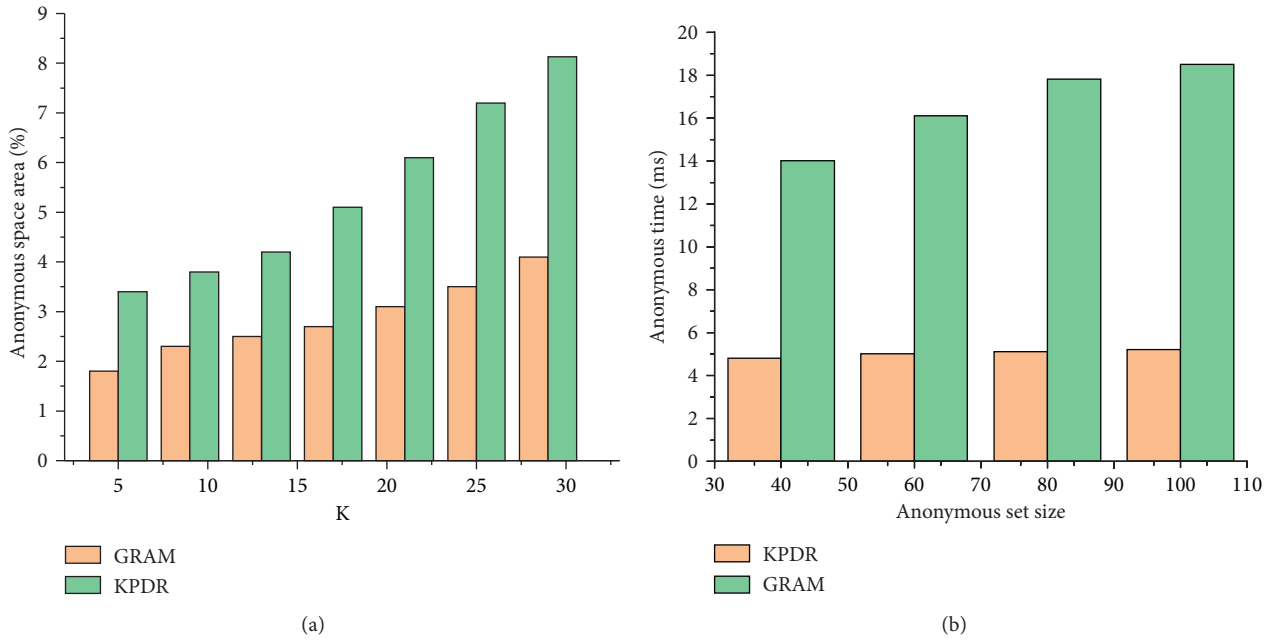

Figure 4: Comparison of anonymous space and anonymous time. (a) Anonymous space area comparison. (b) Anonymous time comparison.

not increase significantly with the increase of the size of user anonymous sets.

6.2. Experimental Comparative Analysis. The experiment uses a dataset of Beijing's POI category for catering services to verify the performance of the KPDR algorithm. The data comes from the POI set of AutoNavi Map, which contains about 10,821 POIs. The algorithm is implemented using Python 3.8.3 programming. The environment is configured as processor Intel (*R*) Core (TM) i7-4710HQ CPU @ 2.50 GHz (8 CPUs), ~2.5 GHz, memory 4 GB, graphics card NVIDIA GeForce GTX 850M, operating system Windows 10 Professional Edition. The Forbidden City Museum is the center of the circle and the distribution of POIs within a 5,000-meter radius after being scaled down.

6.2.1. Comparison between Anonymous Space and Anonymous Time. As shown in Figure 4(a), the number of POIs in the KPDR method is fixed, and the value of K is continuously increased. As the area of the anonymous space becomes larger, the degree of privacy protection of users will be higher, but no matter what value K takes, the area of anonymous space of KPDR is always larger than that of GRAM, and as the value of K becomes larger, the area of anonymous space that differs between the two methods increases. As shown in Figure 4(b), the number of POIs is fixed to ASR; the KPDR method uses the characteristics of Dirichlet graph model and does not need to judge by the algorithm of the nearest distance. The LBS server stores the Dirichlet graph under the current POI division, which is updated only when the POI is changed, while the anonymous time of GRAM method increases significantly because it needs to meet the (*k*, l) mechanism. Therefore, when the scale of anonymous set is increased, the difference of anonymous time between the two methods will be greater.

6.2.2. Comparison of Anonymous Success Rate and Communication Overhead. As shown in Figure 5(a), the ASR is fixed. With the continuous increase of K value, the anonymous success rate of the two methods remains at a relatively high level, but the anonymous success rate of the KPDR method is still higher than that of GRAM method. The GRAM method needs to continuously add edges to the base map to protect user privacy. Each edge addition must be recalculated and K integrations are required, so the anonymous success rate will be lower. As shown in Figure 5(b), the average communication cost of both methods increases with the increase of K value, and the increase of KPDR method is relatively slow, because the increase of K value indicates that users need more location information to construct anonymous areas when requesting services; when the GRAM method increases the value of K and when K reaches a certain node value, it will add a vertex corresponding to the edge on the protection graph, so the GRAM algorithm increases gently and jumps. From the results, the average communication cost of this method is lower than that of GRAM method.

6.2.3. Comparison of Influence of Different Values of POI and K on KPDR Method. As shown in Figure 6(a), taking the number of POIs as 1500, 3000, 4500, 6000, 7500, and 9000, you can see that the anonymous time increases with the increase in the number of POIs; at the same time, keeping the POI value unchanged and increasing K value, the anonymous time will also increase slightly. Although the anonymity time of this scheme will increase with the

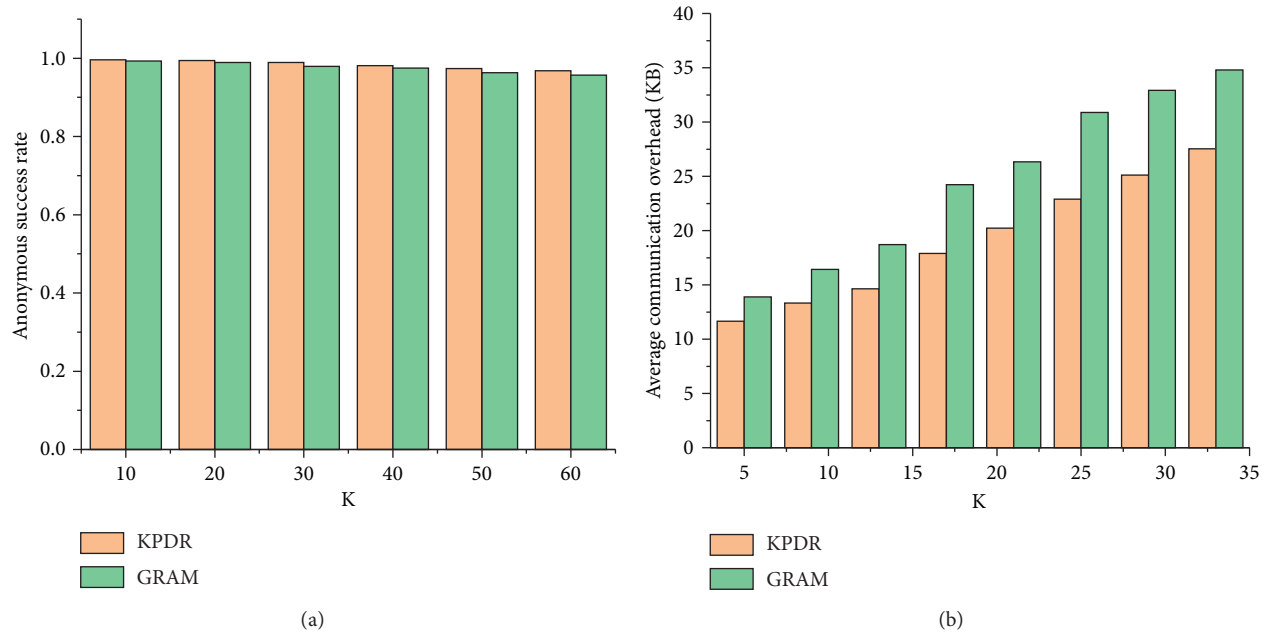

FIGURE 5: Average communication overhead. (a) Anonymous success ratio comparison. (b) Average communication overhead comparison.

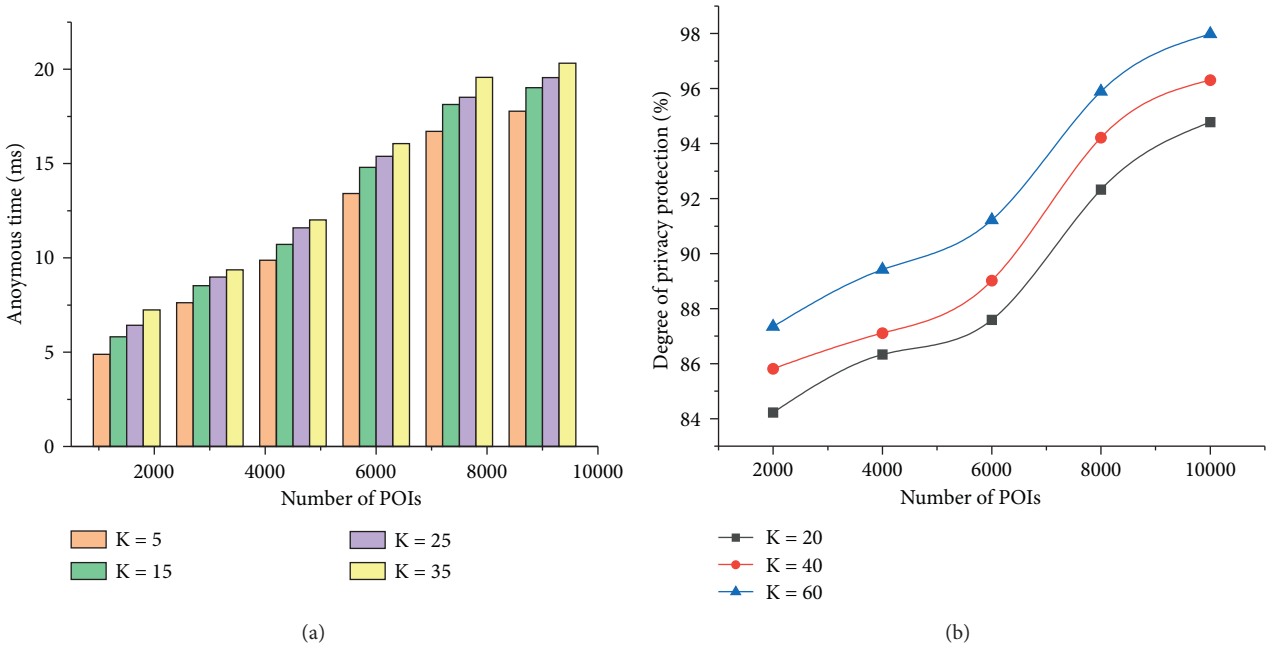

FIGURE 6: Comparison of the influence of POI and K on this scheme. (a) Influence of POI and K on anonymous time. (b) Influence of POI and K on privacy protection degree.

increase of the number of POI and K values, the overall anonymity efficiency is still controlled at a good level. In real life, when there are so many similar POIs, the coverage area is large enough, and such anonymity efficiency is enough to ensure the quality of service for users, which also proves the superiority of this method. As shown in Figure 6(b), take the number of POIs as 2000, 4000, 6000, 8000, and 10000 to test the degree of privacy protection. It can be seen from the experiment that when the K value is equal; the more POIs are generated, the larger the coverage area of the Dirichlet graph is, the higher the dispersion when constructing user anonymity sets, and the higher the degree of privacy protection of users; when the number of POIs is equal, the K value becomes greater, the degree of privacy protection of users is higher, and the overall degree of privacy protection is maintained at a relatively high level.

The GRAM method has proved its superiority compared to the traditional privacy protection method. The experimental results show that the KPDR method proposed in this paper has better security performance and anonymity efficiency than the GRAM method. By storing the Dirichlet graph on the LBS server, the space is exchanged in time, which avoids service congestion due to a large number of user requests, further improves the user's privacy security, and increases the practicability of the method.

7. Conclusion

In this paper, a KPDR method based on K-anonymity mechanism is proposed. By using Dirichlet graph model and quadratic residue theory model, it can effectively resist link attacks and continuous query attacks and solve the problem that anonymous areas are vulnerable to attacks. With the advent of the era of big data as a service provider, we must fully consider the possibility of a large number of requests from users, so the next step will be to improve the query efficiency of users when the concurrent amount of service requests is high.

Data Availability

The data come from the POI set of AutoNavi Map, which contains about 10,821 POIs.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant no. 61300216.

References

[1] L. Zhang, J. Li, S. Yang, B. Wang, and X. Bian, "A novel attributes anonymity scheme in continuous query," *Wireless Personal Communications*, vol. 101, pp. 943–961, 2018.

[2] Location Based Services (LBS) and Real-Time Location Systems (RTLS) Market-Global Forecast to 2020, https://www.digitaljournal.com/pr/2758079.

[3] Y. Sun, M. Chen, L. Hu, Y. Qian, and M. M. Hassan, "ASA: against statistical attacks for privacy-aware users in location based service," *Future Generation Computer Systems*, vol. 70, pp. 48–58, 2016.

[4] W. He, "Research on LBS privacy protection technology in mobile social networks," in *Proceedings of the 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Chongqing, China, 2017.

[5] Z. Lei, H. Lili, L. Desheng, L. Jing, J. Qingfeng, and Y. Qi, "An attribute generalization mix-zone without privacy leakage," *IEEE Access*, vol. 7, pp. 57088–57099, 2019.

[6] Y. Zhang, Q. Chen, and S. Zhong, "Privacy-preserving data aggregation in mobile phone sensing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 980–992, 2016.

[7] R. Wang, X. Jie, Z. Lin, and R. Si, "An improved algorithm of individuation k-anonymity for multiple sensitive attributes," *Wireless Personal Communications an Internaional Journal*, vol. 95, pp. 2003–2020, 2017.

[8] Y. Yuji and I. Kouichi, "K-presence-secrecy: practical privacy model as extension of k-anonymity," *Ice Transactions on Information & Systems*, vol. 100, pp. 730–740, 2017.

[9] X. Li, M. Miao, H. Liu, J. Ma, and K.-C. Li, "An incentive mechanism for k-anonymity in LBS privacy protection based on credit mechanism," *Soft Computing*, vol. 21, no. 14, pp. 3907–3917, 2017.

[10] L. Zheng, H. Yue, Z. Li, X. Pan, M. Wu, and F. Yang, "K-Anonymity location privacy algorithm based on clustering," *IEEE Access*, vol. 6, pp. 28328–28338, 2018.

[11] J. Wang, Z. Cai, Y. Li, D. Yang, J. Li, and H. Gao, "Protecting query privacy with differentially private k-anonymity in location-based services," *Personal & Ubiquitous Computing*, vol. 22, pp. 453–469, 2018.

[12] K. Wang, W. Zhao, J. Cui, Y. Cui, and J. Hu, "A K-anonymous clustering algorithm based on the analytic hierarchy process," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 76–83, 2019.

[13] H. Sun and S. A. Jafar, "The capacity of private information retrieval with colluding databases," in *Proceedings of the IEEE Global Conference on Signal & Information Processing (GlobalSIP)*, pp. 941–946, Washington, DC, USA, November 2017.

[14] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4075–4088, 2017.

[15] R. Mortazavi and S. H. Erfani, "GRAM: An Efficient (K, l) Graph anonymization method," *Expert Systems with Applications*, vol. 153, pp. 113454–113463, 2020.