Hindawi
Security and Communication Networks
Volume 2020, Article ID 8845038, 12 pages
https://doi.org/10.1155/2020/8845038

Research Article

Research on Selection Method of Privacy Parameter ε

Pan Jun Sun

School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, 800 Dongchuan RD, Minhang, Shanghai, China

Correspondence should be addressed to Pan Jun Sun; sunpanjun2008@163.com

Received 17 August 2020; Revised 10 September 2020; Accepted 27 September 2020; Published 23 October 2020

Academic Editor: Vincenzo Conti

Budget factor is an important factor to measure the intensity of differential privacy, and its allocation scheme has a great impact on privacy protection. This paper studies the selection of the parameter ε in several cases of differential privacy. Firstly, this paper proposes a differential privacy protection parameter configuration method based on fault tolerance interval and analyzes the adversary s fault tolerance under different noise distribution location parameters and scale parameters. Secondly, this paper proposes an algorithm to optimize the application scenarios of multiquery, studies the location parameters and scale parameters in detail, and proposes a differential privacy mechanism to solve the multiuser query scenarios. Thirdly, this paper proposes the differential privacy parameter selection methods based on the single attack and repeated attacks and calculates the upper bound of the parameter ε based on the sensitivity Δq, the length of the fault tolerance interval L, and the success probability p as long as the fault tolerance interval. Finally, we have carried out a variety of simulation experiments to verify our research scheme and give the corresponding analysis results.

1. Introduction

In recent years, with the rapid development of information technology, user data have experienced explosive growth. Personal information extracted by data mining and information collection has become a valuable resource for research and decision-making of various research institutions, organizations, and government departments [1]. The analysis and use of massive user data not only bring convenience to people s lives but also bring a great threat to user privacy protection [2].

More and more people pay attention to protecting data privacy while applying data. On the one hand, for published data, k-anonymity, l-diversity, and T-closure protect sensitive information from attacks, such as link attacks, skew attacks, and underlying knowledge attacks [3–7]. However, due to the lack of a strong attack model, they are not strong against background knowledge attack. The existing privacy protection models lack effective and strict methods to prove and quantify the level of privacy protection. Once the model parameters change, the quality of privacy protection will not be guaranteed. However, differential privacy has better resistance to the above attacks and has good privacy protection, which has been widely used by scholars [8, 9].

1.1. Motivation. Privacy protection theory and technology need to be able to prevent different attack means. What is more, with the rapid development of data analysis techniques such as data mining in recent years, attackers can extract information related to user privacy from massive data. Therefore, how to protect the privacy of user data and provide high availability data as much as possible in the process of data query, publishing, and sharing has become a research hotspot in privacy protection [10, 11].

At present, most of the proposed privacy protection schemes use anonymous fuzzy or data distortion processing (such as adding random noise) and other technologies and use mathematical regression analysis, data distortion adjustment, and noise scale parameter adjustment to reduce the error caused by noise, so as to improve the availability of data [12–14]. However, these schemes also have some shortcomings; that is, the same query results will cause the disclosure of privacy information when the query users with

different permissions and reputation levels query the sensitive data.

The differential method has become a hot research topic on many practical applications in recent years. Compared with the traditional privacy protection mode, differential privacy has its unique advantages. Firstly, the model assumes that the adversary has the greatest background knowledge. Secondly, differential privacy has a solid mathematical foundation, a strict definition of privacy protection, and a reliable quantitative evaluation method. By using the output perturbation technology to add random noise to the query output, the single record is in the dataset or not in the dataset, which has little impact on the calculation results. Even if the adversary has the maximum background knowledge, it can ensure that the adversary cannot obtain accurate individual information by observing the calculation results.

The research work of differential privacy protection mainly focuses on improving the privacy protection and data utility in the differential privacy data release, but a small amount of strict mathematical reasoning and research is conducted on the configuration method of privacy protection parameters in the specific practice of differential privacy. In practice, the dataset size, query function sensitivity, and privacy protection probability threshold should be considered in the configuration of privacy protection parameters.

Differential privacy is based on a good mathematical basis and can quantitatively describe the problem of privacy disclosure [1]. These two unique features make it different from other methods. Even in the worst case, if an adversary knows all the sensitive data except one record, it can ensure that the sensitive information will not be disclosed, because the adversary cannot judge whether the record is in the dataset from the query output [2].

The research of differential privacy protection technology mainly considers three problems: (1) how to ensure that the designed privacy algorithm meets the differential privacy to ensure that the data privacy is not leaked; (2) how to reduce the error probability to improve the availability of data; (3) in the face of different environments and attack modes, how to determine the value range of parameter ε and give a credible and reasonable reasoning proof process.

1.2. Contributions. Aiming at the above problems, to ensure the privacy and availability of sensitive data in the process of data query, solve the problem of real data information leakage in the process of data, and reduce the probability of attackers to obtain real results through differential attack and probabilistic reasoning attack, we study the differential privacy parameter selection methods in various situations; these specific contributions are as follows:

(i) We propose a differential privacy parameter configuration method based on fault tolerance interval and analyze the adversary s fault tolerance under different noise distribution location parameters and scale parameters and study the influence of the user's query permission on privacy protection parameter configuration.

(ii) We study the location parameters and scale parameters in detail and propose a differential privacy mechanism to solve the multi-user query scenarios.

(iii) For a single attack, we propose a differential privacy attack algorithm and calculate the upper bound of the parameter ε based on the sensitivity Δq, the length of the fault tolerance interval L, and the success probability p. Furthermore, we propose an attack model to achieve the security of differential privacy protection technology under repeated attacks, analyze the results of repeated attacks and the characteristics of noise distribution function to obtain the probability of noise falling into the fault-tolerant interval, deduce the probability of the adversary s successful attack by the permutation and combination method, and then obtain the selection range of parameter ε.

(iv) We design several experiments, analyze the relationship between adversary s fault tolerance and privacy parameters, derive the configuration formula of the privacy parameter ε, and configure appropriate parameters without violating the privacy probability threshold.

This paper studies the selection of the parameter ε in three cases of differential privacy. The structure of this paper is as follows. In Section 2, we introduce and analyze the research progress of correlation differential privacy parameters. In Section 3, we introduce the concept and theory of differential privacy. In Section 4, we propose a privacy parameter selection method-based fault tolerance and analyze the case of multiple scale parameters. In Section 5, we propose a differential privacy algorithm for a multi-user query. In Section 6, we introduce the query attack mode in differential privacy. In Section 7, we design relevant experiments and show the characteristics of the study through analysis and comparison. In Section 8, we summarize and propose future work.

2. Related Work

Recently, many achievements have been made in differential privacy research. At present, the research of differential privacy protection technology combines database theory, cluster algorithm, statistical knowledge, and modern cryptography [1, 2]. It defines a very strict mathematical model and provides rigorous and quantitative representation and proof of privacy leakage risk [3–7, 15]. Based on the relevance contents, this paper divides the research work of differential privacy protection into two parts.

2.1. Research on the Basic Theory of Differential Privacy. How to reduce the noise of dataset on the premise of differential privacy: Yi and Zhabin [16] proposed a data publishing algorithm based on wavelet transform, which can effectively reduce the size of ε parameter and improve the accuracy of the histogram counting query. Park and Hon [10] studied parameter ε to protect differential privacy and

introduced a new attack index to capture the relationship between attack probability and privacy assurance. Yao [12] introduced the concept of α-mutual information security and showed that statistical security meant mutual information security. Du and Wang [13] proposed a query model and implemented differential privacy by Laplace noise. Tsou and Chen [17] quantified the disclosure risk and linked the differential privacy with k-anonymity. Zhang and Liu [18] proposed a privacy-preserving decision tree classification model based on differential privacy mechanism, through the Laplace mechanism and index mechanism, which provided users with a secure data access interface and optimized the search scheme to reduce the error rate.

Lin et al. [19] proposed an optimized differential private online transaction scheme for online banking, which set consumption boundary with additional noise, and selected different boundaries while satisfying the definition of differential privacy. Besides, they provided a theoretical analysis to prove that the scheme can meet the differential privacy restriction. The choice of a privacy mechanism usually does not have a significant impact on performance but is critical to maintaining the usability of the result. Goryczka and Xiong [20] described and compared distributed data aggregation methods with security and confidentiality, studied the secure multiparty addition protocol, and proposed a new effective Laplace mechanism, ensuring the security of computation, the minimum communication traffic, and the high reliability of the system. Kang and Li [21] proposed a new framework based on the concept of differential privacy, by purposefully adding noise to locally perturb its training parameters, which achieved a compromise between the convergence performance and privacy protection level.

Li et al. [22] focused on the linear query function based on Laplacian mechanism and proposed a method to determine the upper bound of the number of linear queries from the perspective of information theory. Huang and Zhou [23] proposed a differential privacy mechanism to optimize the number of queries in multi-user scenarios and analyzed the distortion of data distribution and the absolute value of noise in terms of utility. Ye and Alexander [15] studied the minimax estimation problem under the restriction of the discrete distribution in the privacy of differential privacy, under the given conditions, considering the structure ε- privacy level of the optimal problem of the privatization program, minimizing expected estimated losses.

2.2. Application of Differential Privacy. Differential privacy has a wide range of applications. Cheng et al. [11] realized the private publishing of high-dimensional data and determined the optimal parameters by non-overlapping coverage. The studies in [14, 24] introduced differential privacy to protect data privacy and prevented the adversary from inferring important sensitive information. Due to the high complexity and multi-dimension of data, [25] proposed a data partition technology and further used the interactive differential privacy strategy to resist the privacy leakage. Based on noise estimation and Laplace mechanism, the work in [26] studied the trade-off relationship between privacy and utility, derived the optimal differential privacy mechanism, and effectively adapted to the needs of personalized privacy protection.

Zhang et al. [27] formally studied the issue of privacy-preserving set-value data publishing on hybrid cloud, provided a complete system framework, and designed a new data partition mechanism, further setting up query analysis tools that can be automatically switched on the structure of the query optimization of hybrid cloud data query, ensuring the confidentiality of data. In a voting system, users can report their desired parameter values to the selector mechanism. Without limiting user preferences, [28] struck a balance between protecting personal privacy and returned accurate results through the parameter epsilon control.

Sun and Tay [29] constructed an optimization framework that combined local variance privacy and inferential privacy measures and proposed a two-stage local privacy mapping model that can achieve information privacy and local variance privacy within a predetermined budget. Cao and Yoshikawa [30] studied the potential privacy loss of a traditional differential privacy mechanism under time dependence, analyzed the privacy loss of adversaries with time dependence, and designed a fast algorithm to quantify the time privacy leakage. Based on the differential privacy model, the study in [31] constructed a privacy protection method based on clustering and noise and proposed a privacy measurement algorithm based on adjacency degree, which can objectively evaluate the privacy protection strength of various schemes and prevent graph structure and degree attacks.

In the cloud service, the study in [32] proposed a priority ranking query information retrieval scheme to reduce the query overhead on the cloud. The higher-ranking query can retrieve a higher percentage of matching files; users can retrieve files on demand by selecting different levels of queries. Sun and Wang [33] proposed a weight calculation system based on the classification regression tree method, which combined differential privacy and decision tree method, and used differential private small-batch gradient descent algorithm to track privacy loss and prevented adversary from invading personal privacy. Chamikara et al. [34] proposed a recognition protocol, which used different privacy to disturb the featured face and stored the data in a third-party server, which can effectively prevent attacks such as member inference and model memory attacks.

To determine the reasonable release time of dynamic positioning data, the study in [35] designed an adaptive sampling method based on proportional integral derivative controller and proposed a heuristic quadtree partition method and a privacy budget allocation strategy to protect the difference privacy of published data, which improved the accuracy of statistical query and improved the availability of published data. There is often a trade-off between privacy and mining results. Xu and Jiang [36] described the interaction between users in the distributed classification scenario, constructed a Bayes classifier, and proposed an algorithm that allowed users to change their privacy budget;

users can add noise to meet different privacy standards. Yin and Xi [37] combined practicability with privacy to establish a multi-level location information tree model and used the index mechanism of differential privacy to noise the access frequency of selected data.

3. Basic Concepts

Here, this paper will introduce some concepts of differential privacy and related theories.

Definition 1 (Adjacent dataset) [1]. Given the dataset D and D' with the same attribute structure, when the number of records difference is 1, the datasets D and D' are called adjacent datasets.

Definition 2 (Differential privacy) [1]. A random algorithm A satisfies ε differential privacy, if and only if, for any two sets D, D' and any output S with only one tuple difference, the following conditions are met:

$$\left|\frac{\text{Prob}(A(D)) \in S}{\text{Prob}(A(D')) \in S}\right| \leq e^{\varepsilon}, \tag{1}$$

where ε is a constant number of the user. Both D and D' differ by at most one tuple; e is a natural logarithm constant. When the parameter ε is small enough, it is difficult for an adversary to distinguish whether the query function acts on D or on D' for the same output S.

Definition 3 (Global sensitivity) [1]. There is a function q: $D \longrightarrow R^d$; the global sensitivity Δq of function q is expressed as follows:

$$\Delta q = \max_{D,D'} \left\| q(D) - q(D') \right\|_1, \tag{2}$$

where D and D' are adjacent datasets, d is the dimension of function q, and $\|q(D) - q(D')\|_1$ is the 1-order norm distance between $q(D)$ and $q(D')$.

Definition 4 (Laplace mechanism) [1]. It adds independent noise to the true answer and uses $\text{Lap}(b)$ to represent the noise from Laplace distribution with a scaling parameter b.

For a function $q; D \longrightarrow R$ over a dataset D, the mechanism A provides the ε-differential privacy:

$$A(D) = q(D) + \text{Lap}\left(\frac{\Delta q}{\varepsilon}\right). \tag{3}$$

For query q on the database D, the random algorithm A returns $q(D) + x$ to the user based on a query result $q(D)$ and adds the noise x to satisfy the Laplace distribution. In the theory of probability and statistics, the probability density function of variable x is expressed as follows:

$$f(x \mid \mu, b) = \frac{1}{2b} e^{-|(x-u/b)|}. \tag{4}$$

$$\begin{aligned} F(x) = \int_{-\infty}^{x} f(\mu) \mathrm{d}\mu &= \begin{cases} \frac{1}{2} e^{-((x-\mu)/b)}, & x < \mu, \\ 1 - \frac{1}{2} e^{-((x-\mu)/b)}, & x \geq \mu, \end{cases} \\ &= \frac{1}{2} + \frac{1}{2} \text{sign}(s-\mu)\left(1 - e^{-(|x-\mu|/b)}\right). \end{aligned} \tag{5}$$

This is the Laplace distribution, μ is the position parameter, and $b > 0$ is the scale parameter, and x is the sample value that satisfies the $f(\mu, b)$ Laplace distribution: $x \infty f(\mu, b)$, $b = (\Delta q/\varepsilon)$; notice that the larger the ε, the smaller the b. For the convenience of discussion, $\mu = 0$; the expectation and variance are μ and $2b^2$, respectively. The implementation of ε-differential privacy algorithm is relatively simple. From Laplace distribution $f(\mu, b)$, the location parameter μ does not affect the adversary, while the parameter $b = (\Delta q/\varepsilon)$ directly affects the vulnerability of the attack. When the parameter b is smaller, the sampling data x is closer to the location parameter μ; on the contrary, when the parameter b is large enough, the sampling data x is equal to the average distribution on $(-\infty, +\infty)$, which is very difficult for the adversary.

Definition 5 ($(\boldsymbol{\alpha}, \boldsymbol{\beta})$ – useful) (see [1, 38]). A mechanism A meets the (α, β) – useful; it has the formulafd6

$$\text{Prob}\left(\left|A_i(D) - A_j(D)\right| \leq a\right) > 1 - \beta, \tag{6}$$

where α and β are the accuracy parameters and A_j is the private algorithm of A_i.

Theory 1 (Sequential composition theory [2]). For $A_1, A_2, \ldots, A_k$, they satisfyε_1-difference privacy, ε_2-difference privacy, and ε_k-differential privacy. When they are applied to the same dataset, publishing results $t = <t_1, t_2, \ldots, t_k>$ meet the $\sum_{i=1}^{k} \varepsilon_i$-differential privacy, $t_1 = A_1(D), t_2 = A_2(D), \ldots, t_k = A_k(D)$.

Theory 2 (Parallel composition theory [2]). A $\text{Prob}(|A_i(D) - A_j(D)| \leq a) > 1 - \beta$ dataset D is divided into k units, $D_1, D_2, \ldots, D_k$, respectively, so that $A_1, A_2, \ldots, A_k$ can satisfy $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ differential privacy, $t = \langle t_1, t_2, \ldots, t_k \rangle$ can satisfy $\max_{i \in [1,2,\ldots,k]} \varepsilon_i$-differential privacy.

Theory 3 (Medium convexity theory [38]). Given that two algorithms A_1 and A_2 satisfy ε-differential privacy, for any probability $p \in [0, 1]$, A_p is used as a mechanism. It uses the algorithm A_1 with the probability p and uses the A_2 algorithm with the probability $1 - p$; then the A_2 mechanism satisfies the ε-differential privacy.

4. Privacy Parameter Selection Based on Fault Tolerance

The query value of the adversary is generated based on the real value; the distribution of noise directly affects the probability of the adversary obtaining the real information.

4.1. Privacy Fault Tolerance. For some query functions, if the noise x is distributed in $[-L, L]\,(L > 0)$, the adversary can infer the true value $f(x)$ with a large probability and then analyze whether a specific record is in or not in the dataset. In this paper, $[-L, L]$ is called the fault tolerance interval, and the corresponding fault tolerance is $\mathrm{fatl}(x)$.

According to the Laplace definition, the probability that the random noise x lies in the fault tolerance $F(x)$ can be obtained by $F(L) - F(-L)$. Thus, the mathematical expression of the adversary s fault tolerance $\mathrm{fatl}(x)$ is obtained as follows:

$$\mathrm{fatl}(x) = F(L) - F(-L) = \int_{-L}^{L} f(\mu)\mathrm{d}\mu = \begin{cases} \frac{1}{2}\left(\exp\left(\frac{\mu + L}{b}\right) - \exp\left(\frac{\mu - L}{b}\right)\right), & \mu \le -L \le L, \\ 1 - \frac{1}{2}\left(\exp\left(\frac{\mu - L}{b}\right) + \exp\left(\frac{-L - \mu}{b}\right)\right), & -L \le \mu \le L, \\ \frac{1}{2}\left(\exp\left(\frac{L - \mu}{b}\right) - \exp\left(\frac{-L - \mu}{b}\right)\right), & -L \le L \le \mu. \end{cases} \tag{7}$$

Through this mathematical theory analysis, we can select appropriate privacy parameters ε and add noise that meets the requirements of differential privacy protection, to prevent the adversary s probabilistic reasoning attack.

4.2. Analysis of Privacy Parameter. When the adversary s fault tolerance level satisfies the privacy probability threshold, the appropriate scale parameter value can be obtained. In this method, the privacy probability threshold $\mathrm{PT}_{\mathrm{pr}} \in (0, 1)$ is determined by the privacy attribute, which means that the adversary s probabilistic inference attack will not exceed the privacy protection threshold.

To meet the requirements of privacy protection, the scale parameter b can meet the formula

$$\mathrm{fatl}(x) = F(L) - F(-L) = \int_{-L}^{L} f(\mu)\mathrm{d}\mu \le \mathrm{PT}_{\mathrm{pr}}. \tag{8}$$

The mathematical expression of fault tolerance $\mathrm{fatl}(x)$ has many forms according to the different position parameters μ.

(1) When $\mu \le -L \le L$, we can get the formula

$$\begin{aligned} \mathrm{fatl}(x) = \frac{1}{2}\left(\exp\left(\frac{\mu + L}{b}\right) - \exp\left(\frac{\mu - L}{b}\right)\right) &\le \mathrm{PT}_{\mathrm{pr}} \\ \longrightarrow \exp\left(\frac{\mu + L}{b}\right) &\le 2\mathrm{PT}_{\mathrm{pr}} + \exp\left(\frac{\mu - L}{b}\right) \\ \longrightarrow \frac{\mu + L}{b} &\le \frac{\mu - L}{b} + \frac{2\mathrm{PT}_{\mathrm{pr}}}{\exp(\mu - L/b)} \\ \longrightarrow \frac{\exp(\mu - L/b)}{b} &\le \frac{\mathrm{PT}_{\mathrm{pr}}}{L} \\ \longrightarrow \frac{\mu - L}{b} - \ln b &\le \frac{\mathrm{PT}_{\mathrm{pr}}}{L} \\ \longrightarrow b^2 - \left(1 - \ln\frac{\mathrm{PT}_{\mathrm{pr}}}{L}\right)b + L - \mu &\ge 0. \end{aligned} \tag{9}$$

(2) When $b > 0$, by solving the formula (8), we can get the formula

$$b \ge \frac{1 - \ln\left(\mathrm{PT}_{\mathrm{pr}}/L\right) + \sqrt{\left(1 - \ln\left(\mathrm{PT}_{\mathrm{pr}}/L\right)\right)^2 - 4 * (L - \mu)}}{2}. \tag{10}$$

(3) When $-L \le \mu \le L$, we can get the formula

$$\begin{aligned}
\mathrm{fatl}(x) &= 1-\frac{1}{2}\left(\exp\left(\frac{\mu-a}{b}\right)+\exp\left(\frac{-a-\mu}{b}\right)\right)\le \mathrm{PT}_{\mathrm{pr}}\\
&\longrightarrow \exp\left(\frac{2\mu}{b}\right)+1\ge 2\left(1-\mathrm{PT}_{\mathrm{pr}}\right)\exp\left(\frac{\mu+L}{b}\right)\\
&\longrightarrow \exp\left(\frac{2\mu}{b}\right)\frac{\mu+L}{b}+\ln 2\left(1-\mathrm{PT}_{\mathrm{pr}}\right)\\
&\longrightarrow \frac{2\mu}{b}\ge \ln(\mu+L)-\ln b+\frac{b\ln 2\left(1-\mathrm{PT}_{\mathrm{pr}}\right)}{\mu+L}\\
&\longrightarrow b(\ln b)-\frac{b^2\ln 2\left(1-\mathrm{PT}_{\mathrm{pr}}\right)}{\mu+L}-b\ln(\mu+L)+2\mu\ge 0\\
&\longrightarrow \left[1-\frac{\ln 2\left(1-\mathrm{PT}_{\mathrm{pr}}\right)}{u+L}\right]b^2-[1+\ln(u+L)]b+2u\ge 0.
\end{aligned} \tag{11}$$

(4) when $b>0$, by solving the above inequality, we can obtain the formulafd12

$$b\ge\frac{[1+\ln(\mu+L)]+\sqrt{[1+\ln(\mu+L)]^2-8\mu\left[1-\left(\left(\ln 2\left(1-\mathrm{PT}_{\mathrm{pr}}\right)\right)/\mu+L\right)\right]}}{2\left[1-\left(\left(\ln 2\left(1-\mathrm{PT}_{\mathrm{pr}}\right)\right)/\mu+L\right)\right]}. \tag{12}$$

(5) When $-L\le L\le\mu$, formula (8) can be rewritten as follows:fd13

$$b^2-\left(1+\ln\frac{L}{\mathrm{PT}_{\mathrm{pr}}}\right)b+L+\mu\ge 0. \tag{13}$$

Budget parameter ε configuration can be expressed as follows:

$$\varepsilon\le\begin{cases}
\dfrac{2\Delta f}{\left(1-\ln\left(\mathrm{PT}_{\mathrm{pr}}/L\right)+\sqrt{\left(1-\ln\left(\mathrm{PT}_{\mathrm{pr}}/L\right)\right)^2-4(L-\mu)}\right)}, & \mu\le -L\le L,\\
\dfrac{2\Delta f\left[1-\left(\ln 2\left(1-\mathrm{PT}_{\mathrm{pr}}\right)/u+L\right)\right]}{\left(1+\ln(\mu+L)+\sqrt{[1+\ln(\mu+L)]^2-8\mu\left[1-\left(\ln 2\left(1-\mathrm{PT}_{\mathrm{pr}}\right)/u+L\right)\right]}\right)}, & -L\le\mu\le L\\
\dfrac{2\Delta f}{\left(1+\ln\left(L/\mathrm{PT}_{\mathrm{pr}}\right)+\sqrt{\left(1+\ln\left(L/\mathrm{PT}_{\mathrm{pr}}\right)\right)^2-4(L+\mu)}\right)}, & -L\le L\le\mu.
\end{cases} \tag{14}$$

From the above analysis, we can deduce the selection range of privacy parameter ε under different location parameters, scale parameters, and privacy probability thresholds.

In this paper, the value range of query authority is set as [0, 1]. To configure smaller privacy protection budget parameters to users with low query rights, the privacy budget

parameter $\varepsilon' = \varepsilon P_b$ is set. Based on this, the configuration method of privacy parameter ε' under different query permissions can be obtained by the following formula:

$$\varepsilon' = \varepsilon P_b \le \begin{cases} \dfrac{2\Delta f P_b}{\left(1 - \ln\left(\mathrm{PT}_{\mathrm{pr}}/L\right) + \sqrt{\left(1 - \ln\left(\mathrm{PT}_{\mathrm{pr}}/L\right)\right)^2 - 4(L-\mu)}\right)}, & \mu \le -L \le L, \\ \dfrac{2\Delta f\left[1 - \left(\left(\ln 2\left(1 - \mathrm{PT}_{\mathrm{pr}}\right)\right)/\mu + L\right)\right]P_b}{\left(1 + \ln(\mu + L) + \sqrt{[1 + \ln(\mu + L)]^2 - 8u\left[1 - \left(\left(\ln 2\left(1 - \mathrm{PT}_{\mathrm{pr}}\right)\right)/\mu + L\right)\right]}\right)}, & -L \le \mu \le L, \\ \dfrac{2\Delta f P_b}{\left(1 + \ln\left(L/\mathrm{PT}_{\mathrm{pr}}\right) + \sqrt{\left(1 + \ln\left(L/\mathrm{PT}_{\mathrm{pr}}\right)\right)^2 - 4(L+\mu)}\right)}, & -L \le L \le \mu. \end{cases} \tag{15}$$

Through this privacy parameter configuration method, the privacy protection probability threshold can be set, and the appropriate privacy parameter ε can be selected according to the query function and the fault tolerance, so as to achieve the privacy protection and ensure the maximization of data utility.

5. Differential Privacy of Multiuser Query

In this section, we continue to study the location parameters and scale parameters and propose a differential privacy mechanism to solve the multi-user query.

Assume that the number of users is m, and the query number of each user is k. The query set is $Q = \{q_{ij} | i \in [m],\ j \in [k]\}$; the results for i^{th} user are covered with scale parameter $b = (\Delta q/\varepsilon)$ and location parameter $u = u_i$. The u_i is randomly chosen from the interval $[\mu - L, \mu + L]$.

According to Definition 3, the global sensitivity is Δq, the $\widehat{r}_{ij} = r_{ij} + x_{ij}$ is the noisy value of the query q_{ij} by the database D, the r_{ij} is the real value of the q_{ij}, and x_{ij} is noise with $b = (k\Delta q/\varepsilon)$ and $\mu = \mu_i$. The $\widehat{r}'_{ij} = r'_{ij} + x'_{ij}$ is the noisy value answer of the query q_{ij} by the database D', x'_{ij} is the noisy value for q_{ij} by the database D', and r'_{ij} is the real value for q_{ij} by the database D'.

Theory 4. For the database D and query set Q, the mechanism A is ε-differential privacy.

Proof. For the D, D' and the i^{th} user s query q_{ij}, the location parameter is μ_i, so it can get the formula

$$\mathrm{Prob}\{\widehat{r}_{ij} = A(D)\} = P\{r_{ij} + x_{ij}\} = P\{x_{ij} = A(D) - r_{ij}\}. \tag{16}$$

x_{ij} meets the Laplace distribution; it can get the formula

$$\mathrm{Prob}\{x_{ij} = A(D) - r_{ij}\} = \frac{\varepsilon}{2k\Delta q} e\left(\frac{-\varepsilon\left|A(D) - r_{ij} - \mu_i\right|}{k\Delta q}\right). \tag{17}$$

For the adjacent database, it can get the formula

$$\mathrm{Prob}\{\widehat{r}'_{ij} = A(D')\} = \frac{\varepsilon}{k\Delta q} e^{\left(-\varepsilon\left|A(D)\widehat{r}'_{ij} - \mu_i\right|/k\Delta q\right)}. \tag{18}$$

For the i^{th} user's query q_{ij}, it can get the formula

$$\begin{aligned} \frac{\mathrm{Prob}\{x_{ij}\}}{\mathrm{Prob}\{x'_{ij}\}} &= \frac{(\varepsilon/2k\Delta q)e\left(-\varepsilon\left|A(D) - r_{ij} - \mu_i\right|/k\Delta q\right)}{(\varepsilon/2k\Delta q)e\left(-\varepsilon\left|A(D') - r'_{ij} - \mu_i\right|/k\Delta q\right)} \\ &= e\left(\frac{-\varepsilon\left|A(D) - r_{ij} - \mu_i\right|}{k\Delta q} - \frac{-\varepsilon\left|A(D') - r'_{ij} - \mu_i\right|}{k\Delta q}\right) \\ &\le e^{\left((\varepsilon/k\Delta q)\left(\left|\widehat{r}_{ij} - \widehat{r}'_{ij}\right|\right)\right)} = e^{(\varepsilon/k)}. \end{aligned} \tag{19}$$

In Algorithm 1, there are some denotations. The database is denoted by D and its global sensitivity is ΔD. for the query q_{ij} of the i^{th} user, the privacy budget is (ε/k). According to Theory1 of differential privacy for the query set Q, this mechanism is ε differential privacy. □

6. Research of the Attack Model

In the actual application scenario, users often face attack problems of different privacy. This section is divided into two parts: single attack and repeated attack.

6.1. Single Attack. Assume that there are only two potential input sets in the worst case, this section discusses how to guess the real value $q(D)$ according to the $q(D) + x$. An adversary puts forward a query question q against the attack object. The database owner gets the result $q(D)$ according to the query question and returns it to the adversary after adding the noise x. The adversary needs to make a judgment by the result $q(D) + x$; an attack object is not in the collection. Each noise x satisfies the Laplace distribution, so it is impossible for the adversary to accurately guess this x. Considering the characteristics of query functions, the adversary can only guess that x falls in a certain range. To describe the above phenomenon, the probability of x in interval $[\mu - L, \mu + L]$ decreases with the increase of b, which can reflect the difficulty of the adversary.

Require: the number of user is m
The number of query for each user is k
The query set is Q
The interval is $[\mu - L, \mu + L]$
The database D and its global sensitivity is Δq
The privacy budget is ε
Ensure: the set of answer $\{\hat{r}_{ij}\}$ for queries
(1) For each user $i \in [n]$ do
(2) Choose μ_i from $[\mu - L, \mu + L]$ for i^{th} user
(3) Set the i^{th} user's noise distribution $\text{lap}((k\Delta q/\varepsilon), \mu_i)$
(4) For each query $q_{ij} \in Q$ do
(5) The answer $\hat{r}_{\text{ij}} = q_{\text{ij}}(D) + \text{lap}((k\Delta q/\varepsilon), \mu_i)$
(6) End

Algorithm 1: Multi-user query.

Lemma 1. *If the Laplace distribution is used to add noise x to $q(D)$, then the probability of $q(D) + x$ in the interval $(-\infty, Q(D) + \mu + L)$ is expressed as $1 - (1/2)e^{(-L\varepsilon/\Delta q)}$.*

Proof. Based on Definition 3, the probability of $q(D) + x$ falling in the interval $(-\infty, q(D) + \mu + L)$ is equal to the probability of x in the interval $(-\infty, \mu + L)$. Therefore, from the Laplace function, according to $b = (\Delta q/\varepsilon)$, the probability of x in the interval $(-\infty, \mu + L)$ is expressed as $F(\mu + L) = (1/2) + (1/2)(1 - e^{(-L/b)}) = 1 - (1/2)^{(-L/b)}$. □

Lemma 2. *The probability of an adversary s success in Algorithm 2 is $1 - (e^{(-\varepsilon/2)}/2)$.*

Proof. Assume that $q(D) = m$ or $q(D) = m + 1$; there are two intervals $(-\infty, m + 0.5)$ and $[m + 0.5, +\infty)$ for $q(D) + x$. From Theory1, if $q(D) = m$, the probability of $q(D) + x$ in the $(-\infty, m + 0.5)$ is $1 - (e^{(-\varepsilon/2)}/2)$. If $q(D) = m + 1$, the probability of $q(D) + x$ in the $[m + 0.5, +\infty)$ is the same.

Therefore, according to $q(D) + x$, the probability of success is $1 - (e^{(-\varepsilon/2)}/2)$; if $q(D) + x$ falls into the $(-\infty, m + 0.5)$ interval, then $q(D) = m$; otherwise, $q(D) = m + 1$. Note: $q(D) = m$ means the adversary is not in the original data; $q(D) = m + 1$ means that the adversary is in the original data. For a common query, it can deduce the probability $1 - (1/2)e^{(-L\varepsilon/\Delta q)}$.

With Lemma 2 and Algorithm 2, when the adversary s success probability $p \le (1 - (1/2)e^{(-L\varepsilon/\Delta q)})$ is solved, it can obtain the upper bound of the ε that meets the formulafd20

$$\varepsilon \le \frac{\ln 2(1 - p)\Delta q}{L}. \tag{20}$$

The upper bound of the parameter ε in formula (20) is independent of the dataset, which is related to the query function $(\Delta q, L)$ and the adversary s success probability p. □

6.2. Repeated Attack. Although differential privacy is the latest technology to protect personal privacy, it has an obvious defect in the Laplace mechanism. If the adversary can perform the same query function infinitely, he can infer the real query result by observing which point the query results concentrate on. Therefore, it is necessary to study the limit of the number of query times.

Input: $A(q(D)) = x + q(D)$
Output present or absence
/* Laplace distribution $f(\mu, b)$, and $q(D) \in \{m, m + 1\}$
(1) $y = x + q(D)$
(2) $y \in [m + 0.5, +\infty]$
(3) Return present
(4) Else
(5) Return absence

Algorithm 2: Single attack query.

According to the above sections, an adversary can obtain $q(D) + x_1, q(D) + x_2, \ldots, q(D) + x_n$ results after N times of attacks.

Lemma 3. *If the adversary attacks N times and adds noise $x_1, x_2, \ldots, x_N$ to $q(D)$ by Laplace distribution, the probability of n times $q(D) + x_i$ in $(-\infty, q(D) + \mu + L)$ is expressed as follows:*

$$C_N^n \left(1 - \frac{1}{2}e^{(-L\varepsilon/\Delta q)}\right)^n \left(\frac{1}{2}e^{-(L\varepsilon/\Delta q)}\right)^{N-n}. \tag{21}$$

Proof. According to Definition 2, it can be known in a query that the probability of $q(D) + x$ in the $(-\infty, q(D) + \mu + L)$ is expressed as follows:

$$F(\mu + L) = 1 - \frac{1}{2}e^{(-L\varepsilon/\Delta q)}. \tag{22}$$

If there are n times in the interval $(-\infty, q(D) + \mu + L)$, from the binomial distribution function, the probability of n in the N times of repeated attacks is $C_N^n (1 - (1/2)e^{(-L\varepsilon/\Delta q)})^n ((1/2)e^{-(L\varepsilon/\Delta q)})^{N-n}$.

In Algorithm 3, $\Delta q = 1$ is the normal query, $\mu = 0$; the half-length of the fault-tolerant interval $L = 0.5$. After

making N times of attack, the adversary can judge whether the attack object is in the result set. □

Lemma 4. *According to Algorithm 3, the adversary performs $N(N = 2n + 1)$ times of query, and the probability of success is expressed as follows:*

$$\sum_{i=1}^{n+1} C_{2n+1}^{n+1}\left(1 - \frac{1}{2}e^{(-\varepsilon/2)}\right)^{n}\left(\frac{1}{2}e^{-(\varepsilon/2)}\right)^{N+1-i}. \tag{23}$$

Proof. Let $N = n + 1$; assume that $q(D) = y$ or $q(D) = y + 1$. Considering two intervals $(-\infty, y + 0.5]$ and $(y + 0.5, +\infty]$, a times falls into $[y + 0.5, +\infty)$ and b times falls into $(-\infty, y + 0.5]$ after$N$ times of attacks. According to Lemma 4, if $q(D) = y + 1$, the probability of $a > b, a = (n + 1, n + 2, \ldots, 2n + 1)$ is $\sum_{i=1}^{n+1} C_{2n+1}^{n+1}(1 - (1/2)e^{(-\varepsilon/2)})^{n+i}((1/2)e^{-(\varepsilon/2)})^{N+1-i}$.

If $q(D) = y$, then $a < b$; the probability of $b = (n + 1, n + 2, \ldots, 2n + 1)$ is $\sum_{i=1}^{n+1} C_{2n+1}^{n+1}(1 - (1/2)e^{(-\varepsilon/2)})^{n+i}((1/2)e^{-(\varepsilon/2)})^{N+1-i}$.

Therefore, the probability of a successful attack is expressed as $\sum_{i=1}^{n+1} C_{2n+1}^{n+1}(1 - (1/2)(e^{-\varepsilon/2}))^{n+i}((1/2)e^{-(\varepsilon/2)})^{n+1-i}$.

$q(D) = y$ indicates that the attack object is not in the original dataset, and $q(D) = y + 1$ indicates that the attack object is in the original dataset. □

7. Experiment Simulation Analysis

The experimental environment: Intel core i7-7500, CPU 2.9 GHz, 8 GB memory, Windows 10 operating system, MATLAB 2015b. The experiment uses UCI machine learning dataset, which contains 48842 records of US census data with 14 attributes. Here, we select five attributes in Table 1: education, marital status, occupancy, native country, and work class.

7.1. Fault-Tolerant Experiment. To express the problem more intuitively, according to the configuration method of privacy parameter in Section 4, the parameter ε is analyzed qualitatively and quantitatively.

In Figure 1, $PT_{pr} = 0.7$, when the location parameter μ is outside the fault tolerance interval $(\mu \le -L \le L$ or $-L \le L \le \mu)$, the adversary s fault tolerance on the fault interval $[-L, L]$ is low; the adversary cannot effectively obtain the real information in the dataset. This is because the location parameter is large, the data distortion is serious, and the data availability is low.

When the location parameter μ is in $-L \le \mu \le L$, the adversary s fault tolerance is higher, which has reference significance for privacy protection analysis. According to Figure 1, this paper analyzes the impact of different interval lengths on the adversary s fault tolerance when the location parameter μ is within the fault tolerance interval.

In Figure 2, the configuration of ε is related to the location parameter μ and fault tolerance interval $[-L, L]$ of noise distribution. Under the same fault tolerance interval, when the position parameter ε is taken as 0, the adversary s fault tolerance is larger. Under the same location parameters, the larger the fault tolerance interval, the greater the fault tolerance level. The maximum privacy parameter value can be obtained without violating the privacy protection probability threshold PT_{pr}.

In Figure 3, the smaller the query authority is, the smaller the upper limit of privacy protection budget parameters is. By limiting the upper limit of privacy protection budget parameters, different values can be configured for query users with different query permission ranges.

Input $A(q(D)) = X + q(D)$
Output present or absence
/* Laplace(μ, b) distribution, and $q(D) \in \{y, y + 1\}$ */
(1) $a = 0, b = 0, i = 1$
(2) While $i < N$
(3) Begin
(4) $m = x + q(D)$
(5) If $m \in [y + 0.5, +\infty]$
(6) $a = a + 1$
(7) Else
(8) $b = b + 1$
(9) $i = i + 1$
(10) End
(11) If $a > b$
(12) Return present
(13) Else
(14) Return absence

Algorithm 3: Repeated attack algorithm (RAA).

Table 1: Attributes of dataset.

Attribute	KL-divergence	Standard deviation
Education	0.010183	0.013650
Marital status	0.016147	0.019767
Occupation	0.001250	0.001680
Native country	0.121652	0.068400
Work class	0.088199	0.088485

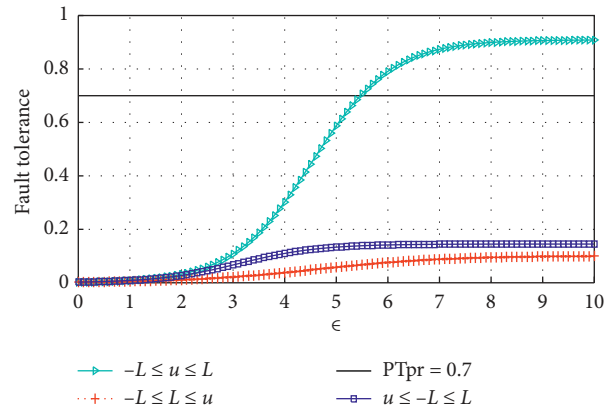

Figure 1: Fault tolerance values of adversary under different parameters.

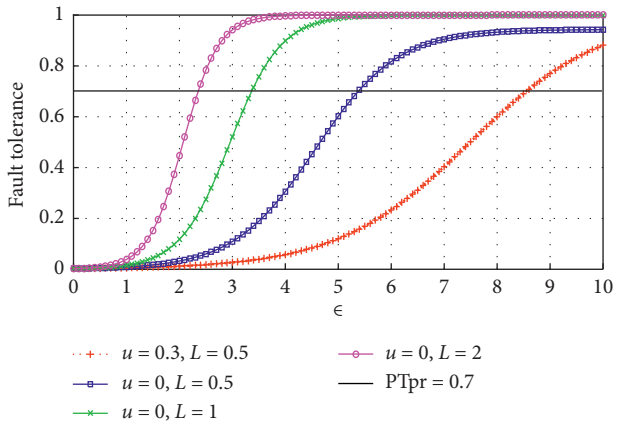

Figure 2: Fault tolerance values of adversary under different intervals.

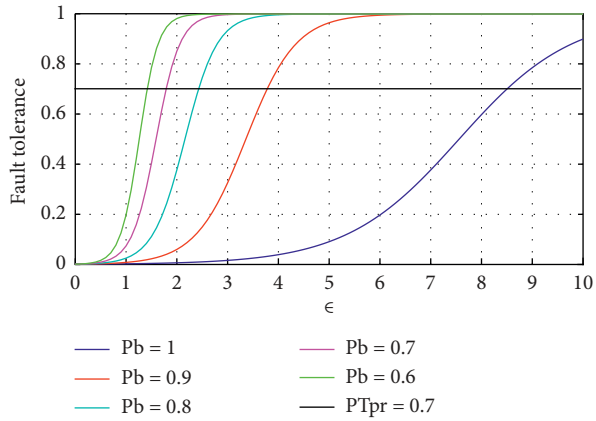

Figure 3: Fault tolerance values under different query conditions.

7.2. Query Success Rate Experiment. ε is an important factor to measure the intensity of privacy protection. Its different allocation schemes have a great impact on the error of the privacy protection algorithm. Next, we give the query probability in different conditions to verify the role of our parameters.

In Figure 4, in the interval $[-\infty, \mu + L]$, with the increasing of the value of ε, the probability of $q(D) + x$ falling in the given interval also increases.

Figure 5 shows the probability curve of different values of the privacy parameter ε in the interval $[\mu - L, \mu + L]$. It can be seen from the figure that, in the range of $[\mu - L, \mu + L]$, the probability of $q(D) + x$ falling into the interval $[\mu - L, \mu + L]$ will decrease with the increase of the value of ε; that is, the probability of $q(D) + x$ falling in the interval $[\mu - L, \mu + L]$ will decrease with the increase of ε.

Figure 6 shows the probability curve image of the noise value falling in the interval $[-\infty, \mu - L]$ with different privacy parameters. As can be seen from Figure 6, with the increase of the privacy parameter ε, the probability of $q(D) + x$ falling into a given interval becomes smaller.

In Figure 7, under the same privacy budget ε, the probability of attack success increases with the number of

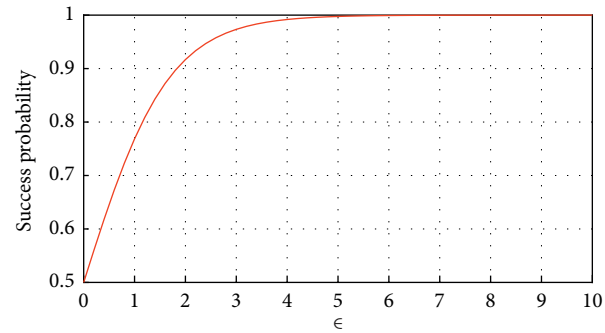

Figure 4: ε in interval $[-\infty, \mu + L]$ and attack success probability.

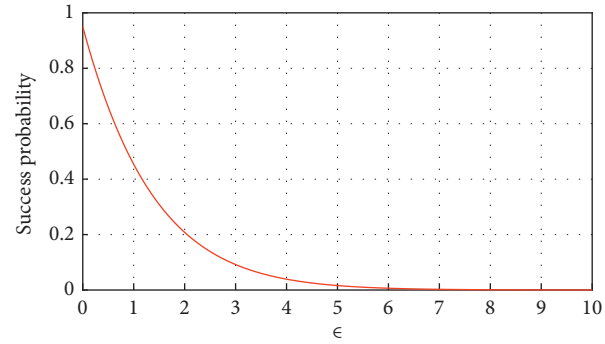

Figure 5: ε in interval $[\mu - L, \mu + L]$ and attack success probability.

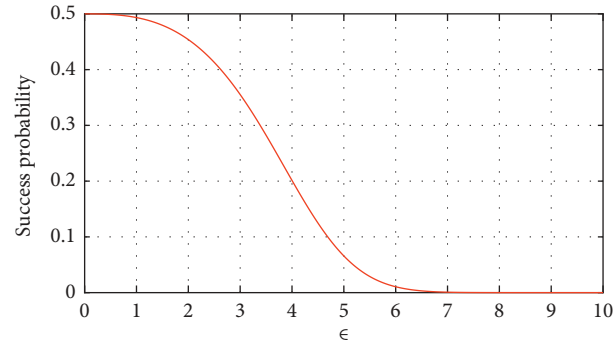

Figure 6: ε in interval $[-\infty, \mu - L]$ and attack success probability.

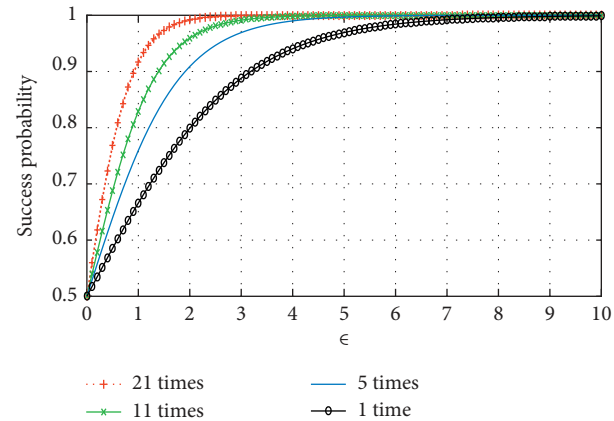

Figure 7: The success probability of count query under different attack times.

attacks; with the increase of ε, the success rate reaches 1; furthermore, the selection range of parameters can be deduced by formula (23).

8. Conclusion

This paper studies the selection of the parameter ε in several cases of differential privacy. Firstly, this paper proposes a differential privacy parameter configuration method based on fault tolerance interval and analyzes the adversary s fault tolerance under different noise distribution location parameters and scale parameters. Secondly, this paper proposes an algorithm to optimize the application scenarios of multi-query and proposes a differential privacy mechanism to solve the multi-user query scenarios. Thirdly, this paper proposes the differential privacy parameter selection methods of several attack models and calculates the upper bound of the parameter ε based on the sensitivity Δq, the length of the fault tolerance interval L, and the success probability p. Finally, we have carried out a variety of simulation experiments to verify our research scheme and given the corresponding analysis results.

The research of ε is limited not only to choosing a proper privacy parameter value in the Laplace mechanism but also to choosing a reasonable ε in exponential mechanism and calculating an ideal parameter value by the method of probability and statistics.

Data Availability

The original data of this article is confidential, but the processed data (data used to support the research in this article) have been given in Section 7 of the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proceedings of the 48 th Annual IEEE Symposium on Foundations of Computer Science*, 2017.

[2] Q. A. Arain, H. Memon, I. Memon, M. H. Memon, R. A. Shaikh, and F. A. Mangi, "Intelligent travel information platform based on location base services to predict user travel behavior from user-generated GPS traces," *International Journal of Computers and Applications*, vol. 39, no. 3, pp. 155–168, 2017.

[3] Q. A. Arain, Z. Deng, I. Memon et al., "Privacy protection with dynamic pseudonym-based multiple mix-zones over road networks," *China Communications*, vol. 14, no. 4, pp. 89–100, 2017.

[4] Q. A. Arain, M. A. Uqaili, Z. Deng et al., "Clustering based energy efficient and communication protocol for multiple mix-zones over road networks," *Wireless Personal Communications*, vol. 95, no. 2, pp. 411–428, 2017.

[5] I. Memon and H. T. Mirza, "MADPTM: mix zones and dynamic pseudonym trust management system for location privacy," *International Journal of Communication Systems*, vol. 31, no. 17, 2018.

[6] I. Memon, L. Chen, and Q. Ali Arain, "Pseudonym changing strategy with multiple mix zones for trajectory privacy protection in road networks," *International Journal of Communication Systems*, vol. 31, no. 1, 2017.

[7] I. Memon and Q. Ali Arain, "Dynamic path privacy protection framework for continuous query service over road networks," *World Wide Web*, vol. 20, no. 4, pp. 639–672, 2017.

[8] P. J. Sun, "Security and privacy protection in cloud computing: discussions and challenges," *Journal of Network and Computer Applications*, vol. 160, Article ID 102642, 2020.

[9] X. Yang, T. Wang, and X. Ren, "Survey on improving data utility in differentially private sequential data publishing," *IEEE Transactions on Big Data*, 2017.

[10] C. Park and D. Hon, "An attack-based evaluation method for differentially private learning against model inversion attack," *IEEE Access*, vol. 7, 2019.

[11] X. Cheng, P. Tang, S. Su, R. Chen, Z. Wu, and B. Zhu, "Multi-party high-dimensional data publishing under differential privacy," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 8, pp. 1557–1571, 2019.

[12] Y. Yao, "A generalized constraint of privacy: α-mutual information security," *IEEE Access*, vol. 7, 2019.

[13] M. Du and K. Wang, "A differential privacy-based query model for sustainable fog data centers," *IEEE Transactions on Sustainable Computing*, vol. 4, no. 2, 2019.

[14] Z. Heng and T. Wang, "Differentially private high-dimensional data publication in Internet of Things," *IEEE Internet of Things Journal*, vol. 7, no. 4, 2020.

[15] M. Ye and B. Alexander, "Optimal schemes for discrete distribution estimation under locally differential privacy," *IEEE Transactions on Information Theory*, vol. 64, no. 8, 2018.

[16] Q. Yi and L. Zhaobin, "An effective data privacy protection algorithm based on differential privacy in Edge computing," *IEEE Access*, vol. 7, 2019.

[17] Y.-T. Tsou and H.-L. Chen, "RoD: Evaluating the risk of data disclosure using noise estimation for differential privacy," *IEEE Transactions on Big Data*, 2018.

[18] L. Zhang and Y. Liu, "Efficient privacy-preserving classification construction model with differential privacy technology," *Journal of Systems Engineering and Electronics*, vol. 28, no. 1, pp. 170–178, 2017.

[19] J. Lin, J. Niu, and X. Liu, "Protecting your shopping preference with differential privacy," *IEEE Transactions on Mobile Computing*, 2020.

[20] S. Goryczka and L. Xiong, "A comprehensive comparison of multiparty secure additions with differential privacy," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 5, 2017.

[21] W. Kang and J. Li, "Federated learning with differential privacy: algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020.

[22] X. Li, H. Li, H. Zhu, and M. Huang, "The optimal upper bound of the number of queries for Laplace mechanism under differential privacy," *Information Sciences*, vol. 503, pp. 219–237, 2019.

[23] W. Huang and S. Zhou, "Optimizing query times for multiple users scenario of differential privacy," *IEEE Access*, vol. 7, 2019.

[24] C. Li and Z. Pan, "Differentially private distributed online learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 8, 2018.

[25] H. Zhang and Z. Zhou, "Towards privacy-preserving publishing of set-valued data on hybrid cloud," *IEEE Transactions on Cloud Computing*, vol. 6, no. 2, 2018.

[26] Q. Geng and P. Viswanath, "The optimal noise-adding mechanism in differential privacy," *IEEE Transactions on Information Theory*, vol. 62, no. 2, 2016.

[27] H. Zhang, Z. Zhou, and L. Ye, "Towards privacy-preserving publishing of set-valued data on hybrid cloud," *IEEE Transactions on Cloud Computing*, vol. 6, no. 2, 2018.

[28] N. Kohli and L. Paul, "Epsilon voting: mechanism design for parameter selection in differential privacy," in *Proceedings of the 2018 IEEE Symposium on Privacy-Aware Computing*, Washington, DC, USA, September 2018.

[29] M. Sun and W. P. Tay, "On the relationship between inference and data privacy in decentralized IoT networks," *IEEE Transactions on Information Forensics and Security*, vol. 15, 2020.

[30] Y. Cao and M. Yoshikawa, "Quantifying differential privacy in continuous data release under temporal correlations," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 7, 2019.

[31] H. Huang and D. Zhang, "Privacy-preserving approach PBCN in social network with differential privacy," *IEEE Transactions on Network and Service Management*, vol. 17, no. 2, 2020.

[32] Q. Liu, C. Chiu, and Tan, "Towards differential query services in cost-efficient clouds," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 6, 2014.

[33] Z. Sun and Y. Wang, "Differential privacy for data and model publishing of medical data," *IEEE Access*, vol. 7, 2019.

[34] M. A. P. Chamikara, P. Bertok, I. Khalil, D. Liu, and S. Camtepe, "Privacy preserving face recognition utilizing differential privacy," *Computers & Security*, vol. 97, Article ID 101951, 2020.

[35] Y. Yan and L. Zhang, "Dynamic release of big location data based on adaptive sampling and differential privacy," *IEEE Access*, vol. 7, 2019.

[36] L. Xu and C. Jiang, "Privacy-accuracy trade-off in differentially-private distributed classification: a game theoretical approach," *IEEE Transactions on Big Data*, 2017.

[37] C. Yin and J. Xi, "Location privacy protection based on differential privacy strategy for big data in industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, 2018.

[38] D. Kifer and B.-R. Lin, "Towards an axiomatization of statistical privacy and utility," in *Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems of Data-PODS 10*, 2010.