Adaptive Personalized Randomized Response Method Based on Local Differential Privacy

Dongyan Zhang, Henan University of Science and Technology, China

Lili Zhang, Henan University of Science and Technology, China*

Zhiyong Zhang, Henan University of Science and Technology, China

https://orcid.org/0000-0003-3061-7768

Zhongya Zhang, Henan University of Science and Technology, China

ABSTRACT

Aiming at the problem of adopting the same level of privacy protection for sensitive data in the process of data collection and ignoring the difference in privacy protection requirements, the authors propose an adaptive personalized randomized response method based on local differential privacy (LDP-APRR). LDP-APRR determines the sensitive level through the user scoring strategy, introduces the concept of sensitive weights for adaptive allocation of privacy budget, and realizes the personalized privacy protection of sensitive attributes and attribute values. To verify the distorted data availability, LDP-APRR is applied to frequent items mining scenarios and compared with mining associations with secrecy konstraints (MASK), and grouping-based randomization for privacy-preserving frequent pattern mining (GR-PPFM). Results show that the LDP-APRR achieves personalized protection of sensitive attributes and attribute values with user participation, and the maxPrivacy and avgPrivacy are improved by 1.2% and 4.3%, respectively, while the availability of distorted data is guaranteed.

KEYWORDS

Frequent items mining, Local differential privacy, Personalized privacy, Privacy budget, Randomized response

1. INTRODUCTION

In recent years, big data technologies have emerged in the scenario of booming network and information technology. People can discover the laws and knowledge hidden in the data from the huge amount of data through data mining algorithms, which is important for industrial development, social services, and many other fields (Chen et al., 2023). If the data is directly provided to a third party, it will lead to the leakage of personal privacy information, which will bring a great threat to personal safety and property security. In addition, if the data miner cannot provide sufficient privacy protection, it will lead to some users refusing to provide data due to a lack of trust, and the data miner will not be able

DOI: 10.4018/IJISP.335225

*Corresponding Author

to mine more accurate information due to a lack of data. Therefore, it is necessary to design a secure privacy protection scheme.

There are more existing privacy protection schemes, such as data anonymization, data encryption, data perturbation, etc. (Cormode et al., 2021), Differential Privacy is a privacy protection technique based on data perturbation, which has become a hot research topic today due to its rigorous mathematical proofs and possession of quantitative privacy protection capabilities (Zhang et el., 2023; Duan et al., 2022; Ren et al., 2022; Qian et al., 2022). In the big data environment, to prevent privacy attacks by untrustworthy third parties and attackers with arbitrary background knowledge, sensitive information needs to be more comprehensively protected, and the Local Differential Privacy (LDP) (Duchi et al., 2013) technique has emerged. Locality refers to the random perturbation of user data before it leaves a smart device, such as a cell phone, and subsequently sent to a third-party data collector, i.e., the data collector only gets a part of the true data, and the data still retains a certain utility. Since it was formally proposed in 2013, LDP technology has been greatly developed and improved, and widely deployed in practical applications, such as Microsoft, Google, Apple, other companies have embedded LDP in their applications (Arcolezi et al., 2023).

With the frequent occurrence of privacy leakage incidents, users' awareness of privacy protection is increasing, and the demand for personalized privacy protection is also growing, for which scholars have proposed many personalized privacy protection methods (Niu et el., 2021; Ma et al., 2022; Li et al., 2022; Qian et al., 2022). Among the existing solutions, GR-PPFM (Guo et al., 2021) is more relevant as it can guarantee the availability of perturbed data while providing personalized privacy protection for users. However, it ignores the fact that there are also different privacy protection needs between the user's data attributes and attribute values. For example, home address requires a higher level of privacy protection compared to gender, and infectious disease (HIV) requires a higher level of privacy protection compared to common class of diseases (flu, fever), so GR-PPFM has some limitations. In order to solve the problem of adopting the same level of privacy protection for sensitive data in the process of data collection, ignoring the fact that different users have personalized privacy protection needs for data security and usability, as well as personalized differences in the attributes and attribute values of the data itself, this paper designs a personalized random response algorithm based on local differential privacy, which determines the sensitive level of user data by a scoring strategy, introduces the concept of sensitive weight for adaptive allocation of privacy budget, realizes the personalized privacy protection of sensitive attributes and attribute values, and ensures the availability of data while meeting the user's personalized needs. The main contributions of this paper are as follows:

1. In order to solve the problem that the existing personalized privacy protection schemes only divide the privacy level based on experience and lacks the user's personalized needs, this paper designs a user scoring strategy to reasonably set the privacy level of the sensitive data based on the user scoring; in order to satisfy the different privacy needs of different users for different sensitive attributes and attribute values, this paper introduces the concept of sensitive weighting and puts forward the method of personalized privacy budget allocation, which allocates a reasonable privacy budget for different attributes and attribute values to satisfy the users' personalized needs.
2. In order to check the availability of distorted data, the proposed algorithm is applied to the frequent items mining scenario, and aiming at the problem of low accuracy of direct mining of distorted data in this scenario, the support reconstruction method is proposed, which theoretically deduces the estimation process of the true support by establishing the mathematical relationship between distorted data and true data in order to improve the accuracy of mining.
3. In order to evaluate the privacy of the proposed algorithm, it is strictly proved that the algorithm satisfies LDP theoretically. To verify data availability, the LDP-APRR method is tested on real

datasets and compared with the existing methods GR-PPFM (Guo et al., 2021) and MASK (Rizvi & Haritsa, 2002). The results show that under the conditions of the same experimental scenarios and parameters, and the same overall privacy protection level, the local differential privacy algorithm proposed in this paper improves the *maxPrivacy* and *avgPrivacy* protection level by 1.2% and 4.3%, respectively.

2. RELATED WORKS

N.I.E et al. (2018), in response to the fact that the LDP in the process of data collection and processing mainly focus on the optimization of a single privacy level, ignoring the user's multilevel privacy needs, designed a user-selectable privacy protection level histogram estimation method, which completed the estimation of different levels separately, and then optimally combined the estimation results of different levels. Niu et al. (2021) proposed a generic framework called SmartGuard to provide users with personalized privacy protection, which quantified user privacy in different scenarios, simulated the impact of different Privacy Preserving Mechanisms on several key factors, and then recommended the optimal privacy policy based on the user's preference and the current state of the user's mobile device. Li et al. (2022) explored to add High-Dimensional Gaussian noise to the model COEfficients (HDG-COE) and generated personalized locally differentially private models for query response. In HDG-COE, the model owner set the indistinguishable region containing the model as a safe region, the larger the region, the more privacy was provided by the mechanism, but the lower the model utility. Wang et al. (2022) proposed a multilevel LDP algorithm recommendation framework, which can set the user's weight for each resource according to the user's preference or establish guidance by the service provider to recommend different LDP algorithms for different users but suitable for the current environment of their own resources through multilevel management to achieve multiuser differential privacy protection. Xue et al. (2022) based on LDP by a non-uniform sampling procedure allows each user to specify his/her own privacy budget and security region based on his/her individual privacy needs, and perturbed sensitive data with privacy parameters to protect privacy. In addition to different privacy protection preferences for different users, Murakami and Kawamoto (2019) pointed out that not all sensitive data had the same level of sensitivity and propose the concept of Utility-optimized Local Differential Privacy (ULDP), which was based on utility-optimized random response and RAPPOR to provide LDP privacy protection for sensitive data only. Considering the setting that the difference between sensitive and non-sensitive data may vary from user to user, a personalized ULDP mechanism with semantic labels was proposed to estimate the distribution of personal data in a high utility way while keeping sensitive information of each user confidential. With this background, Gu et al. (2020) proposed Input-Discriminative Local Differential Privacy (ID-LDP) to discriminate different input data. Unlike differentiating different data using distance, ID-LDP quantified the degree of indistinguishability of different privacy budgets corresponding to inputs by assigning privacy protection coefficients to different data and introducing a systematically defined functional representation.

In addition to LDP-based personalized privacy protection methods, other personalized privacy methods have also been studied by scholars (Guo et el., 2021; Song et al., 2020; Liu et al., 2022). Guo et al. (2021) grouped the data according to the privacy protection requirements of different individuals, and set different privacy protection levels and corresponding randomization parameters for each group of data, i.e., the grouped multi-parameter randomization was used to protect the privacy data, but it adopted the same level of privacy protection for the user's attributes as well as the attribute values, ignoring the differentiation protection between the attributes as well as the attribute values. Song et al. (2020) introduced the weights of sensitive values on the basis of Conventional Randomized Response (CRR) model and introduced them into the decision-making of randomized response, proposing a Personalized Randomized Response (PRR) oriented to multisensitive values, which ensured that different groups of sensitive values can achieve the desired level of privacy protection, and realized personalized privacy protection. Liu et al. (2022) used a locally sensitive hashing method to calculate the similarity of users based on the characteristics of activity recognition data, users were trained only with similar users instead of all users, and feature incremental selection model with integrated learning was used to train the model in a personalized way.

3. PRELIMINARIES

3.1 Local Differential Privacy

In LDP, the user performs distorted data locally and subsequently sends the distorted data to the server, which uses the distorted data to mine the required information. Since the server does not get all the data of the user and hence cannot get the private information of the user, the formalization of LDP is defined as follows:

Definition 1. (ε -Local Differential Privacy) (Duchi et al., 2013) Given a privacy budget $\varepsilon \geq 0$, for a perturbation mechanism $R: T \rightarrow D$, if and only if for any inputs $x, x' \in T$, the probability of obtaining any output $y \in D$ satisfies $\Pr\left\{R\left(x\right) = y\right\} \leq e^{\varepsilon} \Pr\left\{R\left(x'\right) = y\right\}$,we say that perturbation mechanism R that satisfies ε -local differential privacy (ε -LDP).

$\varepsilon - \mathrm{LDP}$ guarantees that an arbitrary attacker cannot infer the exact original inputs from the outputs, and that all the data in T output the same result with almost the same probability when the privacy budget ε tends to zero, in other words, the smaller the privacy budget ε is, the stronger the protection of user privacy is.

In addition, Differential Privacy has two important combinatorial theorems:

Theorem 1. (Sequential Composition)(McSherry & Talwar, 2007) For the same dataset T, if the perturbation mechanisms $R_i\left(\ 1 \leq i < n\right)$ satisfies ε_i -LDP, respectively, it is said that the set of algorithms $\{R_1, R_2, \cdots, R_n\}$ of sequence combinations satisfy $\sum_{i=1}^{n} \varepsilon_i$ -LDP.

Theorem 2. (Parallel Composition) (McSherry, 2009) For the same dataset T divided into n disjoint datasets $T_i\left(1 \leq i < n\right)$, if the perturbation mechanism R_i acting on each dataset $T_i\left(1 \leq i < n\right)$ satisfies the ε_i -LDP respectively, it is said that the set of algorithms $\left\{R_1, R_2, \cdots, R_n\right\}$ of parallel combinations satisfy $\underset{1 \leq i < n}{max}\left\{\varepsilon_i\right\}$ -LDP.

3.2 Perturbation Mechanisms

From Definition 1, it can be seen that the implementation of ε -LDP requires the intervention of the perturbation mechanism R. This section describes three common perturbation mechanisms: the Randomized Response (RR) mechanism (Warner, 1965), the Generalized Randomized Response (GRR) mechanism (Kairouz et al., 2014), and the Personalized Randomized Response (PRR) mechanism (Song et al., 2020).

3.2.1 RR

Warner (1965) first proposed the use of the RR mechanism for privacy-sensitive data collection in 1965, a technique that draws on classic methods in statistical research to simulate an investigator's ability to collect valuable statistical data while trying not to invade the privacy of respondents. Each participant has a probability of giving a true answer of p, or a probability of giving the opposite answer of $1 - p$, whose mathematical expression is shown below:

$$P_{\mathrm{RR}}\left(y \mid x\right) = \begin{cases} p, & y = x \\ 1 - p, & y \neq x \end{cases}$$

where x denotes real data and y denotes distorted data.

This mechanism can only perturb data containing two candidate values (yes or no), so it is no longer applicable when perturbing data with multiple candidate values.

3.2.2 k-RR

Kairouz et al. (2014) proposed the Generalized Randomized Response (GRR) mechanism, also known as the k -RR mechanism, to solve the problem that the RR mechanism is only applicable to the data with two candidate values, which can be directly randomly distorted for the case containing a variety of candidate values. The probability of k -RR for distorted data is:

$$P_{k-\mathrm{RR}}\left(y \mid x\right) = \begin{cases} p, & y = x \\ \dfrac{1-p}{k-1}, & y \neq x \end{cases}$$

When $k = 2$, k -RR is equivalent to RR. In order for k -RR to satisfy the ε -LDP. Wang et al. (2018) stated that p should be satisfied: $p = \dfrac{e^{\varepsilon}}{e^{\varepsilon} + k - 1}$. Therefore, the k -RR mechanism satisfying the ε -LDP can be expressed as follows:

$$P_{k-\mathrm{RR}}\left(y \mid x\right) = \begin{cases} \dfrac{e^{\varepsilon}}{e^{\varepsilon} + k - 1}, & y = x \\ \dfrac{1}{e^{\varepsilon} + k - 1}, & y \neq x \end{cases}$$

This mechanism retains the true value with probability $e^{\varepsilon} / \left(e^{\varepsilon} + k - 1\right)$ and perturbs it with probability $1 / \left(e^{\varepsilon} + k - 1\right)$ to some other values.

3.2.3 PRR

Song et al. (2020) found that the k -RR mechanism adopted the same level of privacy protection when privacy protecting data, which would lead to some data being “over-protected” and some data being “under-protected”, therefore, a Personalized Randomized Response (PRR) mechanism was proposed. The PRR mechanism is proposed, which introduces sensitive weights on the basis of the k -RR mechanism to achieve personalized privacy protection for $k \geq 2$ candidate values. The probability of distorted data by PRR is:

$$P_{\mathrm{PRR}}\left(y \mid x_i\right) = \begin{cases} p\left(w_i\right), & y = x_i \\ \dfrac{1 - p\left(w_i\right)}{k-1}, & y \neq x_i \end{cases}$$

where $p(w_i)$ is $\dfrac{\exp\left(\dfrac{\mu}{w_i}\right)}{\exp\left(\dfrac{\mu}{w_i}\right)+k-1}$, the parameter μ can be set according to the privacy protection needs, and w_i $\left(i\in\left[1,k\right]\right)$ denotes the subjective sensitivity of the user or the data provider to the true candidate value x_i.

4. PROPOSED METHOD

4.1 The LDP-APRR Method

Privacy budget, as an important parameter of LDP, determines the degree of privacy protection of the data on the one hand and the degree of data availability on the other hand. Song et al. designed the PRR mechanism without considering the LDP and lacked a quantitative representation of the degree of privacy protection. In this paper, we design the Local Differential Privacy-Adaptive Personalized Randomized Response (LDP-APRR) mechanism to satisfy the local differential privacy, and the formal definition of LDP-APRR is:

$$P_{\text{LDP-APRR}}\left(y\mid x_i\right)=\begin{cases}\dfrac{e^{\varepsilon\left(w_i\right)}}{e^{\varepsilon\left(w_i\right)}+k-1}, & y=x_i\\ \dfrac{1}{e^{\varepsilon\left(w_i\right)}+k-1}, & y\neq x_i\end{cases}\tag{1}$$

where y denotes the perturbation result of the true sensitivity value x_i, and $\varepsilon\left(w_i\right)$ denotes the functional relationship between the sensitivity weight w_i of the sensitivity value x_i and the privacy budget ε, as denoted in Section 4.1.2.

LDP-APRR determines the data sensitivity level through user scores, and adaptively assigns privacy budgets to sensitive data through sensitivity weights to satisfy the different privacy needs of different users with the probability of $\dfrac{e^{\varepsilon\left(w_i\right)}}{e^{\varepsilon\left(w_i\right)}+k-1}$ retaining the true value, and the probability of $\dfrac{1}{e^{\varepsilon\left(w_i\right)}+k-1}$ perturbed into other $k-1$ values.

Privacy budget as an important parameter in $\varepsilon-\text{LDP}$ mechanism, reasonable allocation of privacy budget is the key to equilibrate the degree of privacy protection and data availability. In this paper, we introduce the concept of sensitive weights, and rationally allocate privacy budgets for different users' privacy protection needs through user scoring strategies.

4.1.1 Setting of Sensitive Weights

Given that this paper utilizes the sensitive weight w_i to achieve the privacy budget adaptive allocation, the value of w_i needs to follow the following principles:

- Monotonicity. That is, the higher the sensitivity of the sensitivity value x_i, the higher the value of w_i, indicating a higher need for privacy protection.

- $\sum_{i=1}^{n} w_i = 1\left(w_i > 0\right)$. $w_i > 0$ denotes the presence of sensitive weights for all sensitive values x_i, and $\sum_{i=1}^{n} w_i = 1$ is used for adaptive allocation of the privacy budget.

4.1.2 Adaptive Allocation of Privacy Budget

Most of the existing personalized privacy protection schemes adopt the method of privacy budget fixed division for user selection for privacy protection, i.e., the privacy level is set in advance based on experience for user selection. In order to make the setting of privacy protection level more in line with the actual needs of users, drawing on the idea of user scoring strategy in B2B, different sensitive values are scored for sensitivity to determine the sensitivity level, and the privacy budget is adaptively allocated through the sensitivity weight to achieve the purpose of high utilization of the privacy budget and self-adaptation. The main steps are as follows:

Step 1: Sensitivity level determining. First, each user scores the sensitivity of each sensitive value in the set of sensitive values, and the range of the sensitivity of the sensitive values can be set as $[S_0, S_{\max}]\,(S_{\max} > S_0 \geq 0)$, the higher the sensitivity, the higher the corresponding scoring value; then, the user scoring situation is subject to the operation of removing the outliers, and according to the user scoring situation, the sensitivity is divided into $L \geq 2$ levels, and the higher the sensitivity is, the higher the level is; finally, the sensitivity of each sensitive value is determined as level $l_i \in \left\{0, 1, 2, 3, \cdots, L-1\right\}$, and when $l_i = 0$, it represents that privacy protection is not required and the data is not perturbed.

Sensitivity value weights calculating. The higher the sensitivity level, the higher the corresponding sensitive value weight. Assuming that the sensitivity level of a certain sensitive value x_i is l_i, the corresponding sensitive value weight W_{l_i} for l_i can be expressed as:

$$W_{l_i} = \frac{2l_i}{L\left(L-1\right)},\ \ l_i \in \left[1,\, L-1\right] \tag{2}$$

Once the sensitivity level l_i of the sensitivity value x_i is determined, its corresponding sensitivity value weight is: $w_i = W_{l_i}$.

Privacy budget adaptive allocating. Since the larger the sensitivity level, the higher the sensitivity weight, i.e., the more privacy protection is needed, the smaller the privacy budget assigned to it, the total privacy budget ε based on the sensitivity weight can be divided into $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{L-1}$, i.e., $\varepsilon = \sum_{i=1}^{L-1} \varepsilon_i$, so the privacy budget assigned to the sensitive weight w_i can be expressed as:

$$\varepsilon_i = \varepsilon\left(w_i\right) = W_{L-l_i} \cdot \varepsilon = \frac{2\left(L-l_i\right)}{L\left(L-1\right)} \cdot \varepsilon,\ \ l_i \in \left[1,\, L-1\right] \tag{3}$$

This user scoring strategy can adaptively calculate the sensitive value weight of each sensitive value, so as to assign a reasonable privacy budget to the sensitive value, and at the same time ensure that the privacy budget setting is more realistic. Based on the above analysis and steps, this paper proposes an adaptive personalized randomized response algorithm based on local differential privacy (LDP-APRR), and the specific perturbation process is shown in Table 1:

The algorithm first requires initial parameters to calculate the sensitivity weights corresponding to the sensitivity levels and the assigned privacy budget ε_i using the publicly available sensitivity

Table 1. Algorithm: LDP-APRR algorithm

```
Input: True data $T$, Privacy budget $\varepsilon$, Sensitivity level $SL$
Output: Distorted data $D = \left(y_i, p_i\right)$
Begin
  Initialize true data $T$
  for each attribute $X$ in $T$ do:
    $sl = SL(X), \quad L = num(sl), \quad k = len(sl)$
    for each value $x_i$ in $X$ do:
      if $sl(x_i) == 0$ then:
        $\varepsilon_i = 0, \quad p_i = 1, \quad y_i = x_i$
      else:
        $w_i = \frac{2sl(x_i)}{L\left(L-1\right)}, \quad \varepsilon_i = w_{L-sl(x_i)} \cdot \varepsilon$
        $y_i = \begin{cases} x_i, & with \quad probability \quad p_i = \frac{e^{\varepsilon_i}}{e^{\varepsilon_i}+k-1} \\ UniformRandom\left(sl \setminus \left\{x_i\right\}\right), & with \quad probability \quad q_i = \frac{1}{e^{\varepsilon_i}+k-1} \end{cases}$
      end if
    end for
  end for
  return $D$
End
```

levels and privacy budgets; secondly, it calculates the disturbance probability based on the sensitivity levels of the sensitive values; and finally, it perturbs the data based on the disturbance probability and collects the distorted data y_i and the disturbance probability p_i.

4.2 Privacy Analysis

Lemma 1. LDP-APRR algorithm satisfies ε-LDP.

Proof: Let x_i, x_i' denote different sensitive values under the same sensitive attribute, respectively. x_i, x_i' undergoes the LDP-APRR mechanism to get the perturbation result y.

From the LDP-APRR algorithm and Eq. (1), it can be obtained that

$$\begin{cases} \Pr\left(y \mid x_i\right) = \frac{e^{\varepsilon_i}}{e^{\varepsilon_i}+k-1} \\ \Pr\left(y \mid x_i'\right) = \frac{1}{e^{\varepsilon_i}+k-1} \end{cases}$$

Then, it follows from Definition 1 that

$$\frac{\Pr\left\{R(x_i)=y\right\}}{\Pr\left\{R(x_i')=y\right\}} \leq \frac{\Pr\left\{y \mid x_i\right\}}{\Pr\left\{y \mid x_i'\right\}} = \frac{e^{\varepsilon_i} / \left(e^{\varepsilon_i}+k-1\right)}{1 / \left(e^{\varepsilon_i}+k-1\right)} = e^{\varepsilon_i}$$

That is, each algorithm $R_i\left(1 \le i \le L-1\right)$ satisfy ε_i -LDP respectively.

Then by the Sequential Composition property of ε -LDP, the above algorithmic sets $\left\{R_1, R_2, \cdots, R_{L-1}\right\}$ of sequence combinations satisfy $\sum_{l=1}^{L-1} \varepsilon_i$ -LDP. In other words, the level of privacy protection is the sum of all privacy budgets $\sum_{l=1}^{L-1} \varepsilon_i$ -LDP.

Another factorization of Eq. (3) and $\sum_{i=1}^{n} w_i = 1\left(w_i > 0\right)$ yield that

$$\sum_{l=1}^{L-1} \varepsilon_i = \sum_{l=1}^{L-1} \left(w_{L-l} \cdot \varepsilon\right) = \varepsilon \cdot \sum_{l=1}^{L-1} w_{L-l} = \varepsilon \cdot 1 = \varepsilon$$

That is, the level of privacy protection is ε -LDP.

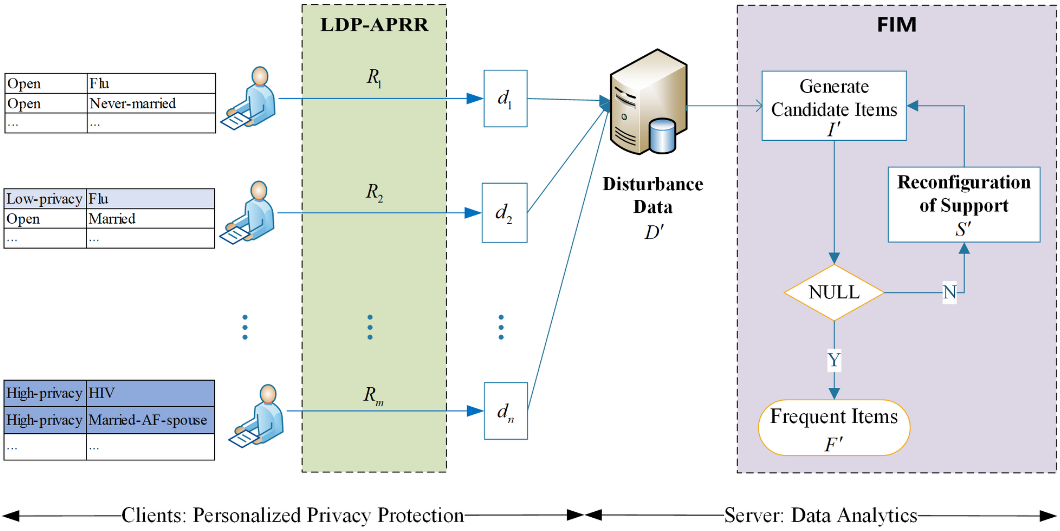
It follows that the LDP-APRR satisfies ε -LDP.

5. FREQUENT ITEMS SUPPORT RECONSTRUCTIONG BASED ON LDP-APRR

Data mining technology can mine the laws or knowledge with potential value from massive data, which can provide more effective services for users, among which Frequent Items Mining (FIM) is one of the hotspots in data mining research (Wang et al., 2018; Wang et al., 2022), the core of which is to find out the frequently occurring items in massive data. Due to the risk of data leakage in the mining process, in order to ensure data privacy and availability, this paper takes it as an application scenario of LDP-APRR, and its application process is shown in Figure 1.

Since the premise of FIM is that the item set support must be obtained, and the LDP-APRR has already perturbed the original dataset in a personalized way, the support must be reconstructed from the distorted data before frequent items mining can be carried out. Since the distorted data is obtained by perturbing the real data according to a certain probability, the support degree of the real data can be reconstructed according to the distorted data and its disturbance probability, and the specific process can be reconstructed according to Lemma 2.

Figure 1. Frequent items mining process based on LDP-APRR



Lemma 2. The number of items in the true data set is determined by the disturbance probability matrix and the number of items in the distorted data, which can be expressed as follows:

$$C^T = P^{-1} \cdot C^D$$

Where, C^T, C^D are the number of items in the true dataset T and the distorted dataset D, respectively, and P is the disturbance matrix.

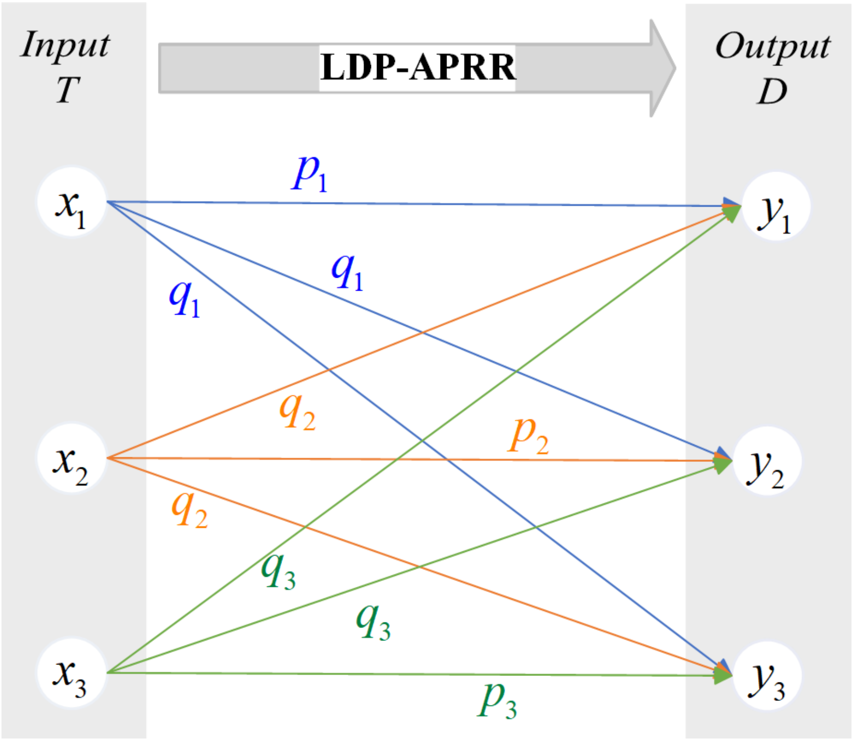
Proof: Suppose a sensitive attribute in T has k sensitive attribute values $\left\{x_1, x_2, \cdots, x_k\right\}$, denote the sensitive attribute values under the corresponding attributes in D as $\left\{y_1, y_2, \cdots, y_k\right\}$, the probability that x_i is perturbed to y_i is denoted as p_i, and the probability that the perturbation is y_j $\left(j \in [1, i) \cup (i, k]\right)$ is denoted as q_i.

When k=3, the perturbation process is shown in Figure 2.

Where the number of x_1, x_2, x_3 are denoted as C_1^T, C_2^T, C_3^T in turn, and the number of perturbation data y_1, y_2, y_3 are denoted as C_1^D, C_2^D, C_3^D, then the following relation can be established according to the perturbation process in Fig. 2:

$$\begin{cases} C_1^D = p_1 \cdot C_1^T + q_2 \cdot C_2^T + q_3 \cdot C_3^T \\ C_2^D = q_1 \cdot C_1^T + p_2 \cdot C_2^T + q_3 \cdot C_3^T \\ C_3^D = q_1 \cdot C_1^T + q_2 \cdot C_2^T + p_3 \cdot C_3^T \end{cases}$$

Figure 2. LDP-APRR based k-element symmetric channel modeling (k=3)



$$\Rightarrow C^D = P \cdot C^T \tag{4}$$

where P C^T and C^D are shown below:

$$P = \begin{bmatrix} p_1 & q_2 & q_3 \\ q_1 & p_2 & q_3 \\ q_1 & q_2 & p_3 \end{bmatrix} = \begin{bmatrix} p_1 & \frac{1-p_2}{2} & \frac{1-p_3}{2} \\ \frac{1-p_1}{2} & p_2 & \frac{1-p_3}{2} \\ \frac{1-p_1}{2} & \frac{1-p_2}{2} & p_3 \end{bmatrix}, C^D = \begin{bmatrix} C_1^D \\ C_2^D \\ C_3^D \end{bmatrix}, C^T = \begin{bmatrix} C_1^T \\ C_2^T \\ C_3^T \end{bmatrix}$$

From Eq. (4), the calculation of the number of items C^T requires the inverse calculation of P. The number of items C^T is:

$$C^T = P^{-1} \cdot C^D \tag{5}$$

Similarly, the support reconstruction can be computed by Eq. (5) for the K-element attribute 1-item set, when P, C^T, and C^D are:

$$P = \begin{bmatrix} p_1 & q_2 & \cdots & q_k \\ q_1 & p_2 & \cdots & q_k \\ \vdots & \vdots & & \vdots \\ q_1 & q_2 & \cdots & p_k \end{bmatrix}, C^D = \begin{bmatrix} C_1^D \\ C_2^D \\ \vdots \\ C_k^D \end{bmatrix}, C^T = \begin{bmatrix} C_1^T \\ C_2^T \\ \vdots \\ C_k^T \end{bmatrix}$$

6. EXPERIMENTS

6.1 Data and Parameter Settings

In order to verify the availability of distorted data, the method in this paper is applied to frequent items mining and compared with GR-PPFM and MASK in the same application scenario on the real shopping basket count set *Basket*. *Basket* is the real shopping basket data of a food supermarket, the average length of the transactions is 3, the total number of transactions is 940, and the total number of items is 11, including fruitveg, freshmeat, dairy, cannedveg, cannedmeat, frozenmeal, beer, wine, softdrink, fish and confectionery. Since the data utility in this paper is validated in a FIM scenario, the publicly available dataset basket for this scenario is used, which can be downloaded directly from https://github.com/PineCoffee/basket. And the TransactionEncoder function is utilized to convert the dataset into a Boolean two-dimensional array, where each row represents a transaction and each column represents a commodity, and if a transaction contains a certain commodity, then the corresponding element is 1, otherwise it is 0.

In the experiment, LDP-APRR and all comparison methods used the same scenario, evaluation criteria and data. The detailed description of these methods is as follows:

- LDP-APRR:This method determines the data sensitivity level by user scoring strategy, and adaptively allocates privacy budget by sensitive weight to provide different levels of privacy protection for sensitive data attributes and attribute values.
- GR-PPFM:This method groups users according to the privacy protection requirements of different users, sets different privacy protection levels and corresponding randomized parameters for each group of user data, and provides different levels of privacy protection for users.
- MASK:This method provides the same level of privacy protection for the data using random numbers generated from predefined distribution functions based on simple probability distortion of the user data.

The experiment classifies privacy protection into five levels $\{0,1,2,3,4\}$, according to the Information Security Level Protection Management Measures promulgated by the State Secrets Bureau on June 22, 2007, which represent Open, Restricted, Secret, Confidential, and Top-secret, respectively. According to the concept of the information entropy in information theory, the smaller the proportion of information, the more information it contains. A reasonable sensitivity level is assigned to each product according to the frequency of its occurrence. In real life, the sensitivity level can be set according to the user scoring strategy designed in 4.1.2. According to Eq. (2) and (3), the privacy budget for each sensitivity level is adaptively assigned, as shown in Table 2.

In the actual questionnaire survey or product purchase, the privacy protection demands of the answers “Yes” and “Purchased” are often higher than those of the answers “No” and “Not purchased”. Therefore, this paper makes a secondary division on the basis of allocating the privacy budget for each level. Since there are only two sensitive value types of “Purchase” and “Not purchased” in this experimental data set, namely “1” and “0”, there are only two types of sensitive value. As a result, in the secondary allocation of privacy budget, the sensitivity level is directly divided into two types, the secondary allocation of privacy budget is completed according to formula (4) and (5), and the disturbance probability is calculated according to formula (3), where the average disturbance probability is:

$$\overline{p} = \frac{1+1}{2}\cdot\frac{4}{11} + \frac{0.92+0.77}{2}\cdot\frac{2}{11} + \frac{0.86+0.71}{2}\cdot\frac{2}{11} + \frac{0.77+0.65}{2}\cdot\frac{2}{11} + \frac{0.65+0.57}{2}\cdot\frac{1}{11} \\ = 0.84$$

Table 2. Parameter settings of dataset

Sensitivity Level (L)	Proportion (ρ)	Sensitivity Values (x_i)	Disturbance Probability (p_i)
0 (Open)	4/11	0	1
		1	1
1 (Restricted)	2/11	0	0.92
		1	0.77
2 (Secret)	2/11	0	0.86
		1	0.71
3 (Confidential)	2/11	0	0.77
		1	0.65
4 (Top-secret)	1/11	0	0.65
		1	0.57

6.2 Experimental Results

6.2.1 Privacy Protection Evaluation

In this section, we mainly analyze the comparison of the privacy protection performance of GR-PPFM, MASK and LDP-APRR methods, and evaluates them from four aspects: the degree of privacy protection, exposed information, personalization and whether or not the LDP is satisfied.

6.2.1.1 The Degree of Privacy Preservation

Definition 2. (*privacy*) (Rizvi & Haritsa, 2002) The degree of privacy protection corresponding to the perturbation probability p, is denoted as:

$$privacy = 1 - R(p) \tag{6}$$

where $R(p) = aR_1(P) + (1-a)R_0(P)$, $R_1(P)$ denotes the probability that a "1" can be reconstructed from the perturbed data, $R_0(P)$ denotes the probability that a "0" can be reconstructed from the perturbed data, and a denotes the proportion of the data accounted for by "1".

Since the data protection of "Yes" and "Purchased" is of more valuable when filling out a questionnaire or purchasing a product, the privacy protection in this paper only considers the case of "1", so a is 1, and the privacy protection degree is $privacy = 1 - R_1(p)$. Let the random perturbation probability be p, the average support of the item is s, then $R_1(P)$ can be expressed as follows:

$$R_1(P) = \frac{p^2 s}{(1-p)(1-s) + ps} + \frac{(1-p)^2 s}{p(1-s) + (1-p)s} \tag{7}$$

In this paper, the data are categorized into five groups $g = \{1,2,3,4,5\}$ according to the sensitivity level, and different levels correspond to different privacy budgets, perturbation probabilities, and different degrees of privacy protection, and Guo et al., (2021) defined the *minPrivacy*, the *maxPrivacy*, the *avgPrivacy*, and the *overallPrivacy*, as follows:

Definition 3. (*minPrivacy*) The degree of privacy protection corresponding to the lowest privacy level, denoted as:

$$minPrivacy = \min\{privacy(g) \mid g = 1,2,3,4,5\} \tag{8}$$

Definition 4. (*maxPrivacy*) The degree of privacy protection corresponding to the hightest privacy level, denoted as:

$$maxPrivacy = \max\{privacy(g) \mid g = 1,2,3,4,5\} \tag{9}$$

Definition 5. (*avgPrivacy*) Average of the degree of grouped privacy protection corresponding to multiple privacy levels, denoted as:

$$avgPrivacy = \sum\nolimits_{g=1}^{5} \rho_g privacy(g) \tag{10}$$

where ρ denotes the percentage of privacy levels.

Definition 6. (***overallPrivacy***) The degree of privacy protection corresponding to the smallest privacy level, denoted as:

$$overallPrivacy = 1 - R_1(\overline{p}) \tag{11}$$

The average support of items in the dataset *basket* is 27.08%, the average support of items is generally calculated from the real data, and in principle the real data is not available, the same real s is unknown, the practical application can be estimated by sampling the value of s. Since this experiment only measures the degree of privacy protection for the sensitive value of "1", the degree of privacy protection of this paper's algorithm under the condition of known sensitivity level and privacy budget can be obtained by substituting $s = 27.08$ and $p_1 = 1$, $p_2 = 0.77$, $p_3 = 0.71$, $p_4 = 0.65$, $p_5 = 0.57$, $\overline{p} = 0.84$ into Eq. (6)-(11), where the *minPrivacy* is 0, the *maxPrivacy* is 71.8%, the *avgPrivacy* is 40.2%, and the *overallPrivacy* is 43.4%.

The corresponding perturbation probabilities $p_1' = 1$, $p_2' = 0.9$, $p_3' = 0.8$, $p_4' = 0.7$, $p_5' = 0.6$, $\overline{p'} = 0.84$ for each level in the GR-PPFM method are brought into the same formula for calculation, and the *minPrivacy* is 0, the *maxPrivacy* is 70.6%, the *avgPrivacy* is 35.9%, and the *overallPrivacy* is 43.4%. The MASK method has only one privacy protection level with a perturbation probability of $p'' = 0.84$, which is the same as the average perturbation frequency of GR-PPFM and LDP-APRR. Therefore, the overall privacy protection degree is 43.4% like GR-PPFM and LDP-APRR, and the specific results are shown in Table 3.

The single-parameter randomization MASK corresponds to only one degree of privacy protection, and in the experiments, the perturbation probability $p'' = 0.84$ for MASK, the average perturbation probability $\overline{p'} = 0.84$ for the grouped multi-parameter randomization GR-PPFM, and the average perturbation probability $\overline{p} = 0.84$ for the method proposed in this paper, LDP-APRR, i.e., the average perturbation frequency of the above three methods is the same, so the *overallPrivacy* is the same; GR-PPFM and LDP-APRR have the same minimum privacy level, so the *minPrivacy* is the same. Under the condition that the three methods have the same degree of the *overallPrivacy*, the satisfying local differential privacy method proposed in this paper improves the *maxPrivacy* by 1.2% and the *avgPrivacy* by 4.3%.

6.2.1.2 Privacy Protection Difference

The algorithm proposed in this paper is different from GR-PPFM and MASK methods in the degree of privacy protection, but there are also differences in personalization.GR-PPFM method only

Table 3. Privacy of LDP-APRR vs. GR-PPFM, MASK

	LDP-APRR	**GR-PPFM**	**MASK**
minPrivacy (%)	0	0	43.4
maxPrivacy (%)	71.8	70.6	
avgPrivacy (%)	40.2	35.9	
overallPrivacy (%)	43.4	43.4	
Whether personalized attributes	√	×	×
Whether personalized attribute values	√	×	×
Whether LDP is satisfied	√	×	×
the exposed information	D, ε	$D, (p_1, \rho_1), (p_2, \rho_2), (p_3, \rho_3), (p_4, \rho_4), (p_5, \rho_5)$	D, p

considers the personalization of the user, ignoring that there are different privacy protection needs for the attributes of the sensitive data and their attribute values. MASK method adopts a single level of privacy protection as a whole, ignoring the user's personalization needs. In this paper, on the basis of user participation in sensitivity scoring and comprehensive assessment of sensitivity level, personalized perturbation of sensitive attributes and values is carried out, which integrates user, attribute and attribute value considerations, and the algorithm proposed in this paper is more fine-grained compared to GR-PPFM coarse-grained personalization.

Since all three algorithms use reconstruction mining for the distorted data, there is a certain risk of exposing the privacy information. MASK supports reconstruction needs to know perturbation probability p, and since all individuals use the same p, the miner knows exactly how each individual corresponds to the randomization parameter, i.e., by specifying an individual at random, the miner knows exactly what parameter the individual has used for the randomization transformation. GR-PPFM requires the knowledge of perturbation probability p, its randomization parameters and its percentage ρ. The local differential privacy algorithm in this paper only needs to expose the total privacy budget ε, which is less information compared to the other two methods and has higher security.

6.2.2 Mining Accuracy Evaluation

The research goal of this paper is to improve the usability of distorted data while protecting data privacy, and the reconstruction of distorted data deviates from the real data, which leads to errors in the mining results of frequent items, so the mining results are evaluated using the Identity Error.

Definition 7. (Identity Error θ) (Rizvi & Haritsa, 2002) Percentage of frequent item set mining error, divided into θ^- and θ^+:

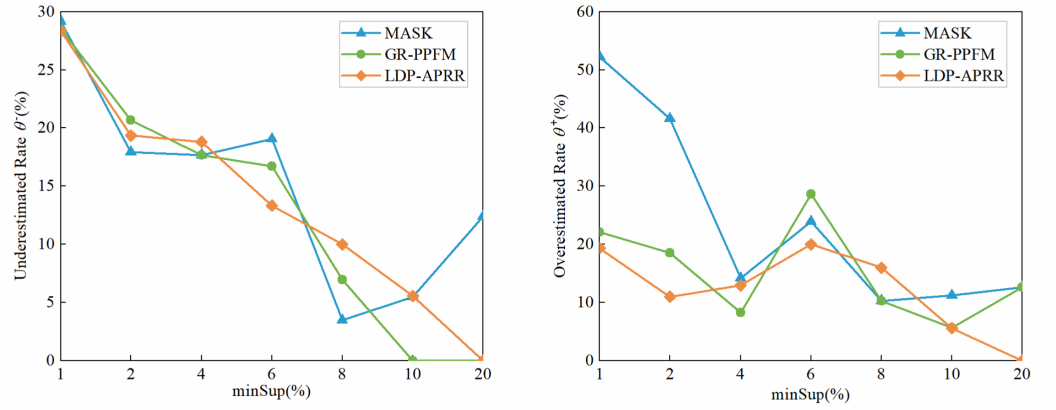
$$\theta^- = \frac{|F - \bar{F}|}{|F|} * 100\%, \ \theta^+ = \frac{|\bar{F} - F|}{|F|} * 100\%$$

where θ^+ denotes the frequent item overestimated rate, which represents the proportion of items that was originally not frequent items and then mistakenly recognized as frequent to the true frequent items; θ^- denotes the frequent item underestimated rate, which represents the proportion of items that were originally frequent items and then mistakenly recognized as not frequent to the true frequent items.

This experiment focuses on verifying the usability of the LDP-APRR method to synthesize the perturbed data, the experimental parameters are set as shown in Table 2, and the effect of the support threshold minSup on the Identity Error (Figure 3) is analyzed through several experiments.

Figure 3(a) shows that the frequent item overestimated rate decreases with the increase of the minimum support threshold minSup, and the decreasing trend is consistent with GR-PPFM, which indicates that the algorithm proposed in this paper is still usable and has a lower error rate under the premise of ensuring data privacy. Figure 3(b) shows that the frequent items underestimated rate decreases with the increase of the minimum support threshold minSup, and the overall error rate is lower, indicating that the algorithm proposed in this paper has a higher mining accuracy. Among them, the increase of frequent items is mainly due to the fact that more "0" data are perturbed to "1", which leads to the increase of "Not purchased" or "No" answer in the original data. "No" in the original data is perturbed to "Purchased" or "Yes". In real life, users are more concerned about the privacy of information related to them, while the privacy protection demand for unrelated information is lower, so the perturbation frequency of "0" data can be set at a higher level to improve data usability.

Figure 3. Experiment error of mask, GR-PPFM and LDP-APRR on real-world data



(a)The effect of minSup on the Underestimation Rate of frequent items

(b)The effect of minSup on the Overestimation Rate of frequent items

7. CONCLUSION

Facing the sensitive data collection process, different users have personalized privacy protection needs for data security and usability, as well as personalized differences in the attributes and attribute values of the data itself. MASK adopts uniform randomization parameters and applies the same level of privacy protection level to all users, GR-PPFM takes into account the variability of privacy protection needs of different users and adopts grouped multi-parameter randomization of data protection to provide personalized privacy protection for users. However, GR-PPFM ignores the fact that the attributes and attribute values of the data itself also have the variability of privacy protection needs, for this reason, this paper proposes an adaptive personalized stochastic response algorithm based on local differential privacy, which determines the sensitivity level through user scores, adaptively allocates privacy budgets through sensitivity weights, and integrally considers the multifaceted needs of the user, attributes, and attribute values to construct a multifaceted, fine-grained personalized privacy protection strategy, and synthesize a privacy dataset that satisfies differential privacy. Through experiments, it is verified that the three methods that are compared, have the same average perturbation probability, same overall Privacy, same minimum privacy level. And at the same degree of the overall Privacy, the proposed method in this paper improves the *maxPrivacy* by 1.2% and the *avgPrivacy* by 4.3% which shows minimal impact in the performance. This algorithm can be widely promoted and applied to scenarios such as medical, financial, and other sensitive data collection involving privacy, etc. However, our algorithm also has some limitations. The reconstruction of perturbed data deviates from the real data, which leads to errors in the mining results of frequent itemsets, such as the accuracy of frequent itemset identification in Figure 3. Therefore, how to improve the privacy data availability and whether the algorithm in this paper can be applied to other personalized privacy-preserving analysis algorithms will be our next research goal.

ACKNOWLEGMENT

Competing Interests

All authors of this article declare there are no competing interest.

Funding Agency

This work was supported by National Natural Science Foundation of China under Grant 61972133, Henan Province Natural Science Fund under Grant 232300420148, Project of Leading Talents in Science and Technology Innovation in Henan Province under Grant 204200510021, Program for Henan Province Key Science and Technology under Grant 222102210177 and 232102211060, Joint Fund Project of Science and Technology Research and Development Plan of Henan Province under Grant 232103810042, Henan Province University Key Scientific Research Project under Grant 23A520008.

REFERENCES

Arcolezi, H. H., Gambs, S., Couchot, J. F., & Palamidessi, C. (2023). On the risks of collecting multidimensional data under local differential privacy. [PVLDB]. *Proceedings of the VLDB Endowment International Conference on Very Large Data Bases*, *16*(5), 1126–1139. doi:10.14778/3579075.3579086

Chen, Y., Gan, W., Wu, Y., & Philip, S. Y. (2023). Privacy-preserving federated mining of frequent itemsets. *Information Sciences*, *625*, 504–520. doi:10.1016/j.ins.2023.01.002

Cormode, G., Maddock, S., & Maple, C. (2021). Frequency estimation under local differential privacy. [PVLDB]. *Proceedings of the VLDB Endowment International Conference on Very Large Data Bases*, *14*(11), 2046–2058. doi:10.14778/3476249.3476261

Duan, J., Ye, Q., & Hu, H. (2022, May). Utility analysis and enhancement of LDP mechanisms in high-dimensional space. In *Proceedings of 2022 IEEE 38th International Conference on Data Engineering (ICDE)* (pp. 407-419). IEEE. doi:10.1109/ICDE53745.2022.00035

Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2013). Local privacy and statistical minimax rates. In *Proceedings of 2013 IEEE 54th Annual Symposium on Foundations of Computer Science* (pp. 429-438). IEEE. doi:10.1109/FOCS.2013.53

Gu, X., Li, M., Xiong, L., & Cao, Y. (2020). Providing input-discriminative protection for local differential privacy. In *Proceedings of 2020 IEEE 36th International Conference on Data Engineering (ICDE)* (pp. 505-516). IEEE. doi:10.1109/ICDE48307.2020.00050

Guo, Y., Tong, Y., & Su, Y. (2021). Privacy preserving frequent pattern mining based on grouping randomization. *Journal of Software*, *32*(12), 3929–3944. doi:10.13328/j.cnki.jos.006101

Kairouz, P., Oh, S., & Viswanath, P. (2014). Extremal mechanisms for local differential privacy. *Advances in Neural Information Processing Systems*, *4*, 2879–2887.

Li, X., Yan, H., Cheng, Z., Sun, W., & Li, H. (2022). Protecting regression models with personalized local differential privacy. *IEEE Transactions on Dependable and Secure Computing*, *20*(2), 960–974. doi:10.1109/TDSC.2022.3144690

Liu, S., Wang, J., & Zhang, W. (2022). Federated personalized random forest for human activity recognition. *Mathematical Biosciences and Engineering*, *19*(1), 953–971. doi:10.3934/mbe.2022044 PMID:34903021

Ma, B., Wang, X., Ni, W., & Liu, R. P. (2022). Personalized location privacy with road network-indistinguishability. *IEEE Transactions on Intelligent Transportation Systems*, *23*(11), 20860–20872. doi:10.1109/TITS.2022.3179501

McSherry, F., & Talwar, K. (2007). Mechanism design via differential privacy. In *Proceedings of 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)* (pp. 94-103). IEEE. doi:10.1109/FOCS.2007.66

McSherry, F. D. (2009). Privacy integrated queries: An extensible platform for privacy-preserving data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data* (pp. 19-30). ACM. doi:10.1145/1559845.1559850

Murakami, T., & Kawamoto, Y. (2019). Utility-Optimized local differential privacy mechanisms for distribution estimation. In *Proceedings of 28th USENIX Security Symposium (USENIX Security 19)* (pp. 1877-1894). USENIX Association.

Nie, Y., Yang, W., Huang, L., Xie, X., Zhao, Z., & Wang, S.N. I. E. (2018). A utility-optimized framework for personalized private histogram estimation. *IEEE Transactions on Knowledge and Data Engineering*, *31*(4), 655–669. doi:10.1109/TKDE.2018.2841360

Niu, B., Li, Q., Wang, H., Cao, G., Li, F., & Li, H. (2021). A framework for personalized location privacy. *IEEE Transactions on Mobile Computing*, *21*(9), 3071–3083. doi:10.1109/TMC.2021.3055865

Qian, W., Shen, Q., Wu, P., Dong, C., & Wu, Z. (2022). Research progress on privacy-preserving techniques in big data computing environment. *Chinese Journal of Computers*, *45*(4), 669–701. doi:10.11897/SP.J.1016.2022.00669

Ren, X., Shi, L., Yu, W., Yang, S., Zhao, C., & Xu, Z. (2022). LDP-IDS: Local differential privacy for infinite data streams. In *Proceedings of the 2022 International Conference on Management of Data* (pp. 1064-1077). ACM. doi:10.1145/3514221.3526190

Rizvi, S. J., & Haritsa, J. R. (2002). Maintaining data privacy in association rule mining. In *Proceedings of the 28th International Conference on Very Large Databases* (pp. 682-693). Morgan Kaufmann. doi:10.1016/B978-155860869-6/50066-4

Song, H., Luo, T., Wang, X., & Li, J. (2020). Multiple sensitive values-oriented personalized privacy preservation based on randomized response. *IEEE Transactions on Information Forensics and Security*, *15*, 2209–2224. doi:10.1109/TIFS.2019.2959911

Wang, H., Li, X., Bi, W., Chen, Y., Li, F., & Niu, B. (2022). Multi-level local differential privacy algorithm recommendation framework. *Journal of Communication*, *43*(8), 52–64. doi:10.11959/j.issn.1000−436x.2022106

Wang, T., Li, N., & Jha, S. (2018). Locally differentially private frequent itemset mining. In *Proceedings of 2018 IEEE Symposium on Security and Privacy (SP)* (pp. 127-143). IEEE. doi:10.1109/SP.2018.00035

Wang, Z., Zhu, Y., Wang, D., & Han, Z. (2022). FedFPM: A unified federated analytics framework for collaborative frequent pattern mining. In *Proceedings of IEEE INFOCOM 2022-IEEE Conference on Computer Communications* (pp. 61-70). IEEE. doi:10.1109/INFOCOM48880.2022.9796719

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, *60*(309), 63–69. doi:10.1080/01621459.1965.10480775 PMID:12261830

Xue, Q., Zhu, Y., & Wang, J. (2022). Mean estimation over numeric data with personalized local differential privacy. *Frontiers of Computer Science*, *16*(3), 1–10. doi:10.1007/s11704-020-0103-0

Zhang, Y., Zhou, Y., Zhou, Y., & Yuan, J. (2023). Mean Estimation Mechanisms under (ε, δ)-Local Differential Privacy. *Dianzi Yu Xinxi Xuebao*, *45*(3), 765–774. doi:10.11999/JEIT221047

Dongyan Zhang is currently pursuing her master's degree in College of Information Engineering, Henan University of Science and Technology, Henan International Joint Laboratory of Cyberspace Security Applications and Henan Intelligent Manufacturing Big Data Development Innovation Laboratory, Luoyang, China. Her research interests focus on cyber security and privacy protection.

Lili Zhang (Member, CCF), received her Master and Ph.D. degrees in Cryptography from Xidian University, China. She is currently working in College of Information Engineering, Henan University of Science and Technology. She is a core member of Henan International Joint Laboratory of Cyberspace Security Applications and Henan Intelligent Manufacturing Big Data Development Innovation Laboratory. Her main research interests are big data technology, privacy computing, cryptographic security protocols and so on. Recent years, she has published over 15 scientific papers and participated in editing 3 books in the above research fields, and also holds 20 authorized patents.

Zhiyong Zhang (IEEE Senior Member, ACM Senior Member) received his Master and Ph.D. degrees in Computer Science from Dalian University of Technology and Xidian University, P. R. China, respectively. He has been ever post-doctoral fellowship at School of Management, Xi'an Jiaotong University, China, and visiting professor of Computer Science Department of Iowa State University, US. Nowadays, he is Director of Henan International Joint Laboratory of Cyberspace Security Applications, Director of Henan Intelligent Manufacturing Big Data Development Innovation Laboratory, Vice-Dean of College of Information Engineering, and full-time Henan Province Distinguished Professor at Henan University of Science and Technology, China. His research interests include cyberspace security and privacy computing, cyber-physical system and industrial internet security, as well as multimodel large model and social intelligence. Recent years, he has published over 150 scientific papers in IEEE TDSC, IEEE TCSS, IEEE TBD, etc, and edited 7 books in the above research fields, and also holds above 20 authorized patents. He is Chair of IEEE MMTC DRMIG, IEEE Systems, Man, Cybernetics Society Technical Committee on Soft Computing, World Federation on Soft Computing Young Researchers Committee, Committeeman of China National Audio, Video, Multimedia System and Device Standardization Technologies Committee. And also, he is editorial board member and associate editor of IEEE Access (IEEE), Human-centric Computing and Information Sciences (Springer), Multimedia Tools and Applications (Springer), Journal of Big Data (Springer), and leading guest editor or co-guest editor of Applied Soft Computing (Elsevier), Computer Journal (Oxford) and Future Generation Computer Systems (Elsevier). And also, he is Chair/Co-Chair and TPC Member for numerous international conferences/ workshops on cyber security and privacy computing, big data and artificial intelligence.

Zhongya Zhang, received his Ph.D.in University of Chinese Academy of Sciences and Institute of Software Research, Chinese Academy of Sciences, majoring in cyberspace security. He is currently working in College of Information Engineering, Henan University of Science and Technology. He is a core member of Henan International Joint Laboratory of Cyberspace Security Applications and Henan Intelligent Manufacturing Big Data Development Innovation Laboratory. His main research interests: quantum computing, analysis and design of symmetric cryptographic algorithms.