Differentially private data aggregating with relative error constraint

Hao Wang¹ · Xiao Peng¹ · Yihang Xiao¹ · Zhengquan Xu² · Xian Chen¹

Received: 13 April 2021 / Accepted: 17 September 2021 / Published online: 9 October 2021
© The Author(s) 2021

Abstract

Privacy preserving methods supporting for data aggregating have attracted the attention of researchers in multidisciplinary fields. Among the advanced methods, differential privacy (DP) has become an influential privacy mechanism owing to its rigorous privacy guarantee and high data utility. But DP has no limitation on the bound of noise, leading to a low-level utility. Recently, researchers investigate how to preserving rigorous privacy guarantee while limiting the relative error to a fixed bound. However, these schemes destroy the statistical properties, including the mean, variance and MSE, which are the foundational elements for data aggregating and analyzing. In this paper, we explore the optimal privacy preserving solution, including novel definitions and implementing mechanisms, to maintain the statistical properties while satisfying DP with a fixed relative error bound. Experimental evaluation demonstrates that our mechanism outperforms current schemes in terms of security and utility for large quantities of queries.

Keywords Privacy preserving · Data fusion · Differential privacy · Relative error · Statistical properties

Introduction

In data-driven applications, such as location based services (LBSs), disease surveillance and social networks, etc., information fusion is necessary for data owners to obtain better services [1]. For example, in location based applications, behaviors aggregating one's precise position can be used to get better shopping recommendations and route planning services. In disease surveillance, gathering individual's physical data can prevent the outbreak of some diseases.

As suggested in above examples, information fusion has outstanding benefits for knowledge discovery and acquisition. But the aggregated data may contain individual's sensitive information (e.g., personal home address, health condition). Untreated data may disclose individual's privacy while data owners may be reluctant to release their true data

values due to privacy concerns. Therefore, privacy preserving data fusion has become a substantial issue in data aggregating and mining [2–4].

Early privacy preserving schemes inherently rely on the security guarantee of designed algorithms [5–7]. Their privacy strengths are decided by their designed algorithms, but not the mathematical foundation. The security and effectiveness of the algorithm is hard to be analyzed mathematically, nor can it be proved theoretically. To remedy this problem, DP proposed by Dwork [8,9] has a solid mathematical foundation, and no restrictions on adversary's background knowledge. It is a type of privacy protection method which strictly defines the strength of protection and data utility. Due to the fact that DP can provide a complete theoretical guarantee of privacy security and better data availability, it has become a significant privacy preserving framework in recent years.

The idea of DP is to preserve individual's privacy by introducing perturbation. Specifically, the noise scale is decided by privacy strength. A big privacy degree means a good protection, but injects high-level noise into the data to be protected, leading to low-level data utility (e.g., performance of clustering or predicting accuracy). In real-world applications, data curators always hope that the error of the data uploaded by users is sufficiently small. The aggregated data with large errors may have a great influence on the accuracy

✉ Xian Chen
chenxian@cqupt.edu.cn

¹ Key Laboratory of Tourism Multisource Data Perception and Decision, Ministry of Culture and Tourism, College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, No. 2, Chongwen Road, Nan'an District, Chongqing 400065, China

² State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, No. 129, Luoyu Road, Wuchang District, Wuhan 430079, China

of mining results [10]. For example, a user who wants to protect his/her precise location, uploads perturbed positions to the service provider (SP) to obtain LBSs, but the positions that the SP collected may have large errors and destroy the data utility, as illustrated in Fig. 1 and Example 1.

Example 1 Consider a scenario that Amy is at the bank, and she wants to query the restaurants within one kilometer radius around her true position x . To preserve her location privacy, Amy utilizes DP technology to perturb her true position x and obtain a perturbation x' . Then Amy sends x' to the SP and increases the query radius (e.g., 2 KM) to filter the results that her wants. In this case, after aggregating perturbed positions, SP may want to use these data to do some analysis and mining tasks (e.g., clustering or classification). However, some positions that Amy uploaded may have large errors, which greatly affect the analysis and mining results. In some extreme cases, data uploaded by users may be useless for a mining task.

The above example illustrates that a good protection needs high-level noise, but high-level noise may violate data availability. The perturbed data that containing much more noise have great effect on analysis and mining results. If we can limit the noise size to an acceptable range while satisfying the requirement of DP, then this dilemma can be avoided. The goal of this paper is to preserve DP while limiting the noisy error to a fixed bound.

Current approaches attempt to solve this problem from two natural solutions: one category is privacy-first mechanisms. Their idea is to reduce the error as much as possible while meeting the privacy requirement of DP. The other category is called “accuracy-first” methods. They first limit the error to a fixed bound and then design the noise form to meet the requirement of DP.

Although various prioritization solutions towards mitigating differentially private release problem with constrained error, current schemes are still afflicted with the following challenges:

- **Violating DP:** Although accuracy-first methods limit the error to a fixed bound. But the noise generated by these methods does not always conform to Laplacian distribution. While only Laplacian noise could meet the privacy requirement of DP strictly, thus accuracy-first methods violate DP.
- **Changing Statistical Properties:** Privacy-first schemes try to improve the accuracy of publishing results under the condition of satisfying DP. But they have an effect on the statistical properties, especially the variance and Mean Square Error (MSE), which are the fundamental elements for data analysis and mining.

These challenges imply that a novel error constrained mechanism for differentially private data release is in high

demand. With respect to the first challenge, to strictly satisfy the requirement of DP, we generate the needed noise in a truncated form. The truncated distribution can satisfy DP, while limiting the error to a fixed bound. For the second challenge, we observed that a linear transformation of a certain distribution still conforms to the same distribution. To maintain the statistical properties, a linear transformation can be conducted to generate noise satisfying the privacy and utility requirements.

Based on these considerations, we propose a relative error constrained differentially private data release solution. We first give the notions of RE-DP and RE-Geo-indistinguishability, as the definitions of error constrained DP in one and two dimensions respectively. Then we utilize truncated mechanisms and linear transformations to realize these two notions in practice. To the best of our knowledge, our solution is the first DP technique that renders linear transformation to maintain the statistical properties (mean, variance and MSE) under the condition of constrained noise error. Our contributions are threefold:

- **Error Constrained DP:** Standard DP has no limitation on error constraint. We extend DP to the notions of RE-DP and RE-Geo-indistinguishability, which are the expansion of DP in one and two dimensional form with error constraint respectively. RE-DP and RE-Geo-indistinguishability limits the noise error to a fixed bound in one and two dimensions respectively, while preserving ϵ -DP.
- **RE-DP:** We propose a mechanism to realize RE-DP in one dimension, including a truncated Laplace mechanism to generate the noise that satisfying DP meanwhile having relative error constraint, and a linear transformation to modify the deviation of mean, variance and MSE. By this means, our scheme does not change the statistical properties under the condition of error constraint while preserving DP.
- **RE-Geo-indistinguishability:** Except the one dimension case, we extend our mechanism to two dimension, which is usually used for location data protect. We propose a truncated planar Laplace mechanism and a linear transformation method. Truncated planar Laplace mechanism can limit the error to a fixed bound and the linear transformation keeps the statistical properties unchanged.

The rest of this paper is arranged as follows. In Section “Related work”, we introduce the mechanisms associated with our work. Then notations and preliminaries adopted in this work are described in Section “Notations and preliminaries”. Section “RE-DP” demonstrates the notion of RE-DP and corresponding implementation mechanism in one dimension. The expansion from one dimension to two dimen-

Fig. 1 Effect of true and perturbed positions on the mining results

sion, including the notion of RE-geo-indistinguishability and its realizing scheme are described in Section “RE-geo-indistinguishability”. The experimental evaluation was performed in Section “Experimental evaluation”, while conclusions and future work are in Section “Conclusions and future works”.

Related work

Existing utility error constrained differentially private data aggregating mechanisms can be classified into two categories. The former is accuracy-first methods. They first set a fixed noise bound and design the noise, whose size is within the bound, to satisfy the requirement of DP. The other type is privacy-first methods. These methods first make the noise satisfy the requirement of DP, then they improve the data utility using a variety of means.

Privacy-first

When publishing various sensitive data (such as search logs [11,12], machine learning results [13–15], etc.), there are a lot of technologies for enforcing DP. Nonetheless, none of these technologies can optimize errors of published data. Instead, they can not reduce noisy results’ variance or specific application indicators, such as the accuracy of classification [16]. Next, we will discuss some typical methods of privacy priorities.

Barak et al. [17] designed a technology to publish the marginals of a given data set. However, their purpose is not only to improve the accuracy of published data, but also to make noisy results more user-friendly. Their method can guarantee each marginal number is non-negative and all margins are consistent. That is to say, the sum of each marginal number and that of the others should be equal. To extend the scenario of Barak et al.’s work [17], Kasiviswanathan et al. [18] proved several lower bounds of the relative error of noisy marginal counting. However, they did not come up with any specific algorithm for the release margin. Instead, Blum et al. [19] presented a specific mechanism to publish single dimensional data, which has a good performance on

relative error in terms of various counting query even facing the worse case. Moreover, literature [20] proposed an method to improve the performance range of Blum et al.’s scheme. To address multidimensional data, Xiao et al. [21] designed a method for multidimensional data publishing.

Li et al. [22] summarized the methods in [20,21]. They proposes an optimal scheme, which can minimize the relative error of any given query. In fact, the method proposed by Li et al. can also address marginal data. Since we can consider the marginal count data as the result to a specific count query. Nonetheless, as these methods only decrease the relative error of count data, they still generate a big relative error for the smaller count data.

Privacy first mechanisms enforce the noise satisfy ϵ -differential privacy. They attempt to minimize the noisy relative error to improve the utility of data. Nevertheless, the optimal results not always meet the accuracy requirement. In addition, the deviations of mean, variance and MSE are not discussed in these works, which have significant effect on the mining results.

Accuracy-first

Ligett et al. [23] proposed a framework in theory, which takes care of accuracy first. It has a high-level utility while preserving the privacy of empirical risk minimization (ERM). We can utilize it to search the privacy degree space and find the most experienced algorithm which satisfies the requirement of accuracy. While the number of privacy levels searched only produces logarithmic overhead.

To apply the framework proposed by Ligett et al. [23] in practice, Arinjita et al. [24] improved the utility of the publication methods based on Kronecker graphs. Instead of anonymizing the social network as a whole, they anonymize each cluster of the network separately, and combine the sanitized results thereafter.

Kasra et al. [25] extend the scenario of Arinjita et al.’s [24] work. They used their developed error bounds to provide guidelines to calibrate privacy levels to keep filter error within pre-specified bounds.

The most relevant work is the truncated Gaussian mechanism proposed by Liu [26]. According to the definition of

Table 1 Summary of related work

Categories	References	Key ideas	Scenarios
Privacy-first	Barak [17]	Publish the marginals of a given data set	Sum
	Kasiviswanathan [18]	Prove lower bounds of relative error	Count
	Blum [19]	Publish single dimensional data	Count
	Hay [20]	Improve performance range of Blum [12]	Histogram
	Xiao [21]	Publish multidimensional data	Count
	Li [22]	Minimize the relative error of any given query	Count
Accuracy-first	Ligett [23]	Find the most experienced algorithm	ERM
	Arinjita [24]	Anonymize each cluster of the network separately	Social network
	Kasra [25]	Keep filter error within pre-specified bounds	Stream
	Liu [26]	Generalized Gaussian (GG) distribution family	Sum
	Our solution	Render linear transformation to maintain statistical property	Count and Histogram

global sensitivity in DP, she demonstrates that the widely used Laplacian distribution is a specific instance of generalized Gaussian (GG) distribution family. She discusses how GG mechanism meets the theoretical requirements of DP under the pre-specified privacy degree, and studies the relationships between GG distribution and Laplacian distribution. In her work, she indicates that the statistical properties could be the same with original data only if the bound is symmetrical. But she has not given the method to make the properties unchanged.

The aforementioned schemes limit the error to a fixed bound and then change the form of noise to meet the privacy requirement of DP. But these methods have the risk of violating DP.

Summary

The privacy and accuracy are always a couple of contradiction in private data publishing. As shown in Table 1, state-of-the-art methods try to solve this contradiction from two aspects. Nonetheless, privacy first schemes can not guarantee enough utility with a certain relative error. Although accuracy first methods can preserve good utility, the noise generated may not conform to Laplacian distribution, and the statistical properties are changed. Therefore, in this paper, our goal is to provide an practical solution to aggregate data with constrained relative error while preserving comprehensive privacy. Specially, we aim to address the following challenges:

- How to define the relative error constrained DP in one and two dimension respectively?
- How to guarantee the privacy requirement of DP with a relative error constrained noise?
- How to maintain the statistical properties (mean, variance and MSE) when we use an error constrained noise?

Notations and preliminaries

This section introduces the basic concepts for solving our problem. In particular, we first give the necessary symbols for our work. Secondly, we describe the preliminaries of DP and its implementation mechanism. Finally, an expansion of DP in two dimension, i.e. Geo-indistinguishability is demonstrated.

Notations

We use several notations to denote the symbols we needed in one and two dimensions respectively in this paper. In one dimension, we use a dataset D , to denote sensitive instances. To protect sensitive data, data owner utilizes a randomized perturbation mechanism M to generate and publish a perturbed query result $Q' = M(D)$. In two dimension, we use p and p' to denote the true and perturbed two dimensional data respectively. Table 2 lists the main notations used in this paper.

Differential privacy

DP [8] is a state-of-the-art privacy preservation model which can guarantee the security of indistinguishability. Essentially, it is a noisy perturbation privacy preserving mechanism. By adding perturbation to raw data or statistical results, DP can guarantee that changing a single record's value has minimal effect on the output results. Thus, DP can preserve the privacy of data to be protected, while supporting mining results well. Definition 1 is its formalized form.

Definition 1 (ϵ -DP [8]) Considering two adjacent datasets, D and D' , which have the same admeasurement, but differ one record to be protected. If the random perturbation mechanism M makes every set of results S satisfy the following equation

Table 2 Notations

Symbol	Description
D	One dimensional dataset contains sensitive data.
M	Privacy preserving mechanism.
p, p'	True and perturbed sensitive data in two dimension.
α	relative error bound.
A	Coefficient of the truncated Laplace mechanism.
w	Parameter of the linear transformation in one dimension.
B	Coefficient of the truncated Planar Laplace mechanism.
ω	Parameter of the linear transformation in two dimension.
r, θ	Perturbed radius and angle in two dimension.
γ	Intermediate variable to realize the linear transformation.
u	Random noise generated in one dimension.
o	Origin in Cartesian coordinate system.

Fig. 2 Probability density function of random algorithm M on the statistical output of D and D'

on D and D' , then M satisfies ϵ -DP.

$$Pr[M(D) \in S] \leq e^\epsilon \times Pr[M(D') \in S], \quad (1)$$

where $S \subseteq \text{Range}(M)$, $\text{Range}(M)$ is the value range of random algorithm M . $Pr[\cdot]$ indicates probability density function (PDF) and ϵ represents privacy budget parameter.

A smaller ϵ is related with high-level privacy. Figure 2 shows the PDF of random algorithm M on the statistical output of D and D' .

Privacy budget ϵ is mainly limited by random algorithm M . Laplace mechanism is usually used to realize M . The Laplace mechanism is defined as follows.

Definition 2 (Laplacian Mechanism [4,9]) Let $f(\cdot)$ be the statistical function of the output result. The noisy samples $Z \sim \text{Lap}(\lambda)$ obeying Laplacian distribution can ensure the random perturbed result $M(D) = f(D) + Z$ satisfy ϵ -DP, where λ is the scale of Laplacian distribution. The PDF of

Laplacian distribution is formalized by the following formula

$$\rho(z) = \frac{1}{2\lambda} \exp\left(-\frac{|z|}{\lambda}\right). \quad (2)$$

The scaling parameter λ is decided by the sensitivity function Δf and privacy protection intensity ϵ :

$$\lambda = \frac{\Delta f}{\epsilon}, \quad (3)$$

where Δf is the largest effect of a single record on the statistical results.

$$\Delta f = \max_{D'} \|f(D) - f(D')\|_1. \quad (4)$$

For example, consider a dataset whose sensitivity is 1. Based on the concept of DP, the noise (added to the real answer) distributed according to $\text{Lap}(1/\epsilon)$ is enough to guarantee ϵ -DP.

Geo-indistinguishability

DP can preserve one's privacy strictly with an ϵ parameter. But DP and Laplace mechanism can only address one dimensional data. In real-world applications, the data always are high dimension. For example, location data are commonly applied in LBSs and are fundamental type in big data. Based on the idea of “indistinguishability”, Andres et al. proposed the notion of “Geo-indistinguishability”, to address two dimensional data.

Definition 3 (Geo-indistinguishability [27]) For all observations S , a mechanism satisfies ϵ -Geo-indistinguishability if:

$$\frac{Pr(S|p)}{Pr(S|p')} \leq e^{\epsilon r} \quad \forall r > 0 \forall p, p' : d(p, p') \leq r. \quad (5)$$

The above definition ensures that points within distance 1 can produce observations with limited probabilities. If two points are farther, the probabilities to produce S are more different. It is very similar as the definition of differential privacy, which requires two databases that differ a single record value to produce the same answer with similar probabilities.

Because the Laplace mechanism can only process one dimensional data, Geo-indistinguishability uses a distribution defined on a plane. In addition, they must utilize straight distance $|p - \mu|$ to replace Euclidean plane distance $d(p, \mu)$. However, only $|p - \mu|$ needs to be replaced by $d(p, \mu)$ in (1) results, leading to the natural expansion of the Laplacian distribution from one dimension to two dimensions.

Definition 4 (Planar Laplacian Mechanism [27]) Given the privacy budget $\epsilon \in \mathbb{R}^+$, and individual's real location $p \in \mathbb{R}^2$, the PDF of our privacy preserving algorithm, on another point $p' \in \mathbb{R}^2$, is:

$$D_\epsilon(p)(p') = \frac{\epsilon^2}{2\pi} e^{-\epsilon d(p, p')}, \quad (6)$$

where $\epsilon^2/2\pi$ is a normalization parameter.

They call this function the planar Laplace operator, centered on p . It should be noted that the introduction of planar Laplacian operator on any vertical plane, which passes through its center draws a graph that is proportional on the linear Laplace operator.

RE-DP

In this section, we first propose the definition of RE-DP, which satisfies DP with constrained relative error. Then we demonstrate our implement mechanism based on truncated Laplace distribution, which can guarantee DP while limiting the perturbed error into a fixed bound. Then we make a linear transformation to keep the mean, variance and MSE the same as original data to reserve its utility.

Definition of RE-DP

DP in Definition 1 only can address one dimensional data. As we have discussed in “Introduction” section, the SP may always has requirement on the perturbed error. In this case, the definition of DP is not appropriate. In this section, We propose the notion of RE-DP with relative error constraint. First of all, we give the definition of relative error.

Definition 5 (Relative Error) Denote z as the random noise corresponding to the query result s_z , where $s_z \in S$, then z/s_z is denoted as the relative error, limited by the bound α , i.e. $|z/s_z| \leq \alpha$, where z is generated with ϵ -RE-DP. Then we can say that the relative error of z is α .

Fig. 3 PDFs of DP and RE-DP with $\lambda = 1$

Different with the definition of DP, RE-DP considers the constrained relative error. RE-DP must limit the relative error to a fixed bound meanwhile satisfying ϵ -DP. We give its definition as following:

Definition 6 (RE-DP) Considering two adjacent datasets, D and D' , which have the same admeasurement, but differ one record to be protected. Then the random perturbation mechanism M satisfies ϵ -DP if M makes every set of results S on adjacent datasets D and D' satisfy

$$\begin{aligned} \Pr[M(D) \in S | |z/s_z| \leq \alpha] \\ \leq e^\epsilon \Pr[M(D') \in S | |z/s_z| \leq \alpha], s_z \in S, \end{aligned} \quad (7)$$

where $S \subseteq \text{Range}(M)$, $\text{Range}(M)$ is the value range of random algorithm M . $\Pr[\cdot]$ indicates probability density function (PDF) and ϵ represents privacy budget parameter.

Truncated Laplace Mechanism

Definition 6 gives the formal definition of RE-DP. Then we propose a truncated Laplace mechanism to realize RE-DP in practice. The truncated Laplace mechanism is shown in Definition 7.

Definition 7 (Truncated Laplace Mechanism) A Laplace noise z that conforms to the following distribution satisfies RE-DP.

$$f(z) = \frac{1}{2\lambda(1 - e^{-\alpha s_z/\lambda})} e^{-\frac{|z|}{\lambda}}, z \in [-\alpha s_z, \alpha s_z]. \quad (8)$$

Definition 7 gives the form of noise which could provide bounded relative error constraint. Compared to the standard Laplace distribution, the PDF of truncated Laplace distribution has an extra factor $(1 - e^{-\alpha s_z/\lambda})^{-1}$. The function of

factor is to make the cumulative distribution function of truncated Laplace be 1. Besides, the relative error is limited by the bound. We can implement this noise form in a post-hoc way by discarding the out of range cleaned results of the conventional Laplace mechanism until the inbound value is obtained. Figure 3 is the diagram of PDFs of DP and RE-DP with $\lambda = 1$. The differences between them are the limitation on x axis and the peak of RE-DP is higher than that of DP.

Statistical properties

In the above section, we have given the methods to generate noise satisfying DP with relative error constraint. However, the challenging issue is that the statistical properties of the data intended to protect have been changed after applying truncated Laplace mechanism. In this paper, we investigate three commonly used properties in data mining tasks, which are mean, variance and MSE. Since standard Laplacian mechanism is a symmetrical distribution with mean 0, and the methods we used are also symmetrical. Thus, our mechanism does not changed the mean of noise, as shown in Theorem 1.

Theorem 1 *The mean of the random variables generated by our truncated Laplacian mechanism is 0.*

Proof Assume z is a random variable who conforms to the truncated Laplacian distribution. Then the mean of z is:

$$\begin{aligned} E(z) &= \int_{-\alpha s_z}^{\alpha s_z} z f(z) dz \\ &= \int_{-\alpha s_z}^{\alpha s_z} \frac{z}{2\lambda(1 - e^{-\alpha s_z/\lambda})} e^{-\frac{|z|}{\lambda}} dz \end{aligned} \quad (9)$$

Let $\frac{z}{\lambda} = y$, then we have

$$\begin{aligned} E(z) &= \int_{-\alpha s_z}^{\alpha s_z} \frac{\lambda}{2(1 - e^{-\alpha s_z/\lambda})} y e^{-|y|} dy \\ &= 0. \end{aligned} \quad (10)$$

Theorem 1 demonstrates that the mean of our proposed truncated Laplacian mechanism is the same with that of standard Laplacian mechanism. In addition, we have known that the variance of standard Laplacian distribution is $2\lambda^2$, Theorem 2 demonstrates the variety of variance using our proposed mechanism. $\square$

Theorem 2 *Given a random variable z that conforms to the truncated Laplacian distribution, the variance of z is:*

$$E(z^2) = \frac{\lambda^2}{1 - e^{-\alpha s_z/\lambda}} \left[2 - e^{-\alpha s_z/\lambda} \left(\left(\frac{\alpha s_z}{\lambda} \right)^2 + 2 \frac{\alpha s_z}{\lambda} + 2 \right) \right]. \quad (11)$$

Proof The calculation equation of variance of random variable z is:

$$\begin{aligned} E(z^2) &= \int_{-\alpha s_z}^{\alpha s_z} z^2 f(z) dz \\ &= \int_{-\alpha s_z}^{\alpha s_z} \frac{z^2}{2\lambda(1 - e^{-\alpha s_z/\lambda})} e^{-\frac{|z|}{\lambda}} dz \end{aligned} \quad (12)$$

Let $\frac{z}{\lambda} = y$, then we have

$$\begin{aligned} E(z^2) &= \int_{-\alpha s_z/\lambda}^{\alpha s_z/\lambda} \frac{\lambda^2}{2(1 - e^{-\alpha s_z/\lambda})} y^2 e^{-|y|} dy \\ &= \int_0^{\alpha s_z/\lambda} \frac{\lambda^2}{1 - e^{-\alpha s_z/\lambda}} y^2 e^{-y} dy \\ &= \frac{\lambda^2}{1 - e^{-\alpha s_z/\lambda}} e^{-y} (-y^2 - 2y - 2) \Big|_0^{\alpha s_z/\lambda} \\ &= \frac{\lambda^2}{1 - e^{-\alpha s_z/\lambda}} \left[2 - e^{-\alpha s_z/\lambda} \left(\left(\frac{\alpha s_z}{\lambda} \right)^2 + 2 \frac{\alpha s_z}{\lambda} + 2 \right) \right]. \end{aligned} \quad (13)$$

Theorem 2 indicates that the variance of truncated Laplace random variable is decided by the bound α , which may affect quantity of applications. According to the properties of Laplace distribution, a linear transformation of Laplace variables still conforms to Laplace distribution. We utilize this property to generate truncated Laplace noise to make the variance not changed. $\square$

Theorem 3 *Given a random variable z that conforms to the truncated Laplacian distribution, with a bound of $[-\gamma, \gamma]$. Denote $u = wz$, if we want to u is bounded by $[-\alpha s_z, \alpha s_z]$ and has the same variance with standard Laplacian distribution, then w , γ and α should satisfy the following two conditions:*

$$\begin{aligned} w\gamma &= \alpha s_z, \\ w^2 A &= 1 - e^{-\gamma/\lambda}, \end{aligned} \quad (14)$$

where $A = 1 - \frac{1}{2} e^{-\gamma/\lambda} [(\gamma/\lambda)^2 + 2\gamma/\lambda + 2]$.

Proof If $u = wz$, $u \in [-\alpha s_z, \alpha s_z]$, then the PDF of u is:

$$\begin{aligned} f(u) &= \frac{1}{|w|} f_z \left(\frac{u}{w} \right) \\ &= \frac{1}{w} \frac{1}{2\lambda(1 - e^{-\gamma/\lambda})} e^{-\frac{|u|}{\lambda w}} du. \end{aligned} \quad (15)$$

Then the mean of u is:

$$\begin{aligned} E(u) &= \int_{-\alpha s_z}^{\alpha s_z} u f(u) du \\ &= \int_{-\alpha s_z}^{\alpha s_z} \frac{1}{2w\lambda(1 - e^{-\gamma/\lambda})} u e^{-\frac{|u|}{\lambda w}} du \\ &= 0. \end{aligned} \quad (16)$$

and the variance $E(u^2)$ of random variable u is:

$$\begin{aligned} E(u^2) &= \int_{-\alpha s_z}^{\alpha s_z} u^2 f(u) du \\ &= \int_{-\alpha s_z}^{\alpha s_z} \frac{1}{2w\lambda(1 - e^{-\gamma/\lambda})} u^2 e^{-\frac{|u|}{\lambda w}} du. \end{aligned} \quad (17)$$

Let $\frac{u}{\lambda w} = y$, then we have

$$\begin{aligned} E(u^2) &= \int_{-\gamma/\lambda}^{\gamma/\lambda} \frac{1}{2w\lambda(1 - e^{-\gamma/\lambda})} \lambda^3 w^3 y^2 e^{-|y|} dy \\ &= \frac{\lambda^2 w^2}{2(1 - e^{-\gamma/\lambda})} \int_{-\gamma/\lambda}^{\gamma/\lambda} y^2 e^{-|y|} dy \\ &= \frac{\lambda^2 w^2}{1 - e^{-\gamma/\lambda}} \int_0^{\gamma/\lambda} y^2 e^{-y} dy \\ &= \frac{\lambda^2 w^2}{1 - e^{-\gamma/\lambda}} e^{-y} (-y^2 - 2y - 2) \Big|_0^{\gamma/\lambda} \\ &= \frac{\lambda^2 w^2}{1 - e^{-\gamma/\lambda}} [2 - e^{-\gamma/\lambda} ((\gamma/\lambda)^2 + 2\gamma/\lambda + 2)] \\ &= \frac{2\lambda^2 w^2}{1 - e^{-\gamma/\lambda}} \left[1 - \frac{1}{2} e^{-\gamma/\lambda} ((\gamma/\lambda)^2 + 2\gamma/\lambda + 2) \right]. \end{aligned} \quad (18)$$

If w, γ, α satisfy the relationship in Equation (14), then $E(u^2) = 2\lambda^2$. From Equation (14) we can solve that $w = \alpha s_z / \gamma$. After given a fixed relative bound αs_z and privacy degree λ , substitute it into the second equation, the bound value γ of truncated Laplacian distribution can be obtained. $\square$

At last, we investigate the MSE after applying our truncated mechanism and linear transformation. The method to obtain a random variable's MSE is:

$$MSE(z) = E(z^2) - E^2(z) \quad (19)$$

Since the mean $E(z)$ and variance $E(z^2)$ of the random variable z do not change compared to the standard Laplacian distribution after applying our truncated mechanism and linear transformation, the MSE can be regarded as an unbiased estimation to standard Laplacian noise. In other words, the

MSE of our mechanism is also unchanged compared to that of standard Laplacian mechanism. Algorithm 1 shows the complete working flow of our solution RE-DP.

Algorithm 1 RE-DP

Require: ϵ , query result s_z , Δf .

Ensure: z , perturbed query results s'_z .

1. Set $z/s_z = \alpha$;
 2. Compute noise scale parameter $\lambda = \Delta f / \epsilon$;
 3. Compute w and γ according to Theorem 3;
 4. Generate truncated Laplace noise z according to the PDF in Definition 7 with bound γ ;
 5. Compute the noisy response $s'_z = s_z + z$;
- return** s'_z .
-

RE-geo-indistinguishability

In last section, we have proposed the notion of RE-DP and its implement mechanism in one dimension. While the two dimensional Laplace noise is also commonly used in real-world applications, e.g., geographic data mining. In this section, we extend our notion and mechanism to two dimensional case. We first propose the notion of RE-Geo-indistinguishability, to define the two dimensional case with relative error. Then we give the truncated two dimensional noise form, to satisfy the notion of RE-Geo-indistinguishability. Finally, we give a linear transformation method to maintain the statistical properties (including mean, variance and MSE).

Definition of RE-geo-indistinguishability

In the mechanism of Geo-indistinguishability, their definition of privacy formalizes an intuitive concept, that is to protect the location of users (typical two-dimensional data) within the radius of r , and the privacy level depends on R , which satisfies DP. The noise conforms to Equation (6) can preserve the privacy of Geo-indistinguishability.

The data curators always have accuracy requirement and the r generated by Geo-indistinguishability mechanism has a infinite right bound, which destroys the accuracy of published data.

We first give the formal definition of Geo-indistinguishability with an relative error constraint, as shown in Definition 8.

Definition 8 (RE-Geo-indistinguishability) For all observations S , the mechanism satisfies RE-Geo-indistinguishability if and only if:

$$\frac{Pr(S|p)}{Pr(S|p')} \leq e^{\epsilon r} \quad \forall r > 0 \forall p, p' : \frac{d(p, p')}{d(p, o)} \leq \alpha, \quad (20)$$

where $d(p, p')$ and $d(p, o)$ are Euclidean distances of p

between the perturbed location p' and coordinate origin o in Cartesian coordinate system.

We can see that the forms of Geo-indistinguishability and RE-Geo-indistinguishability are similar. The difference between them is that the relative error r in RE-Geo-indistinguishability has been limited by α .

Truncated planar Laplacian mechanism

In this section, we propose a truncated planar Laplacian mechanism to generate two dimensional Laplace noise with an relative error α , to satisfy the definition of RE-Geo-indistinguishability.

Definition 9 (Truncated Planar Laplacian Mechanism) Given the parameters $\epsilon \in \mathbb{R}^+$ and the actual location $p \in \mathbb{R}^2$, the PDF value of our noise mechanism at any other point $p' \in \mathbb{R}^2$, is

$$D_{\epsilon}(p_0)(p) = \frac{\epsilon^2}{2\pi[1 - (\epsilon\alpha d(p, o) + 1)e^{-\epsilon\alpha d(p, o)}]} e^{-\epsilon d(p, p')}, \quad (21)$$

where $\frac{d(p, p')}{d(p, o)} \leq \alpha$ and $\epsilon^2/2\pi$ is a normalization factor.

Next, we transform the noise form Cartesian coordinate system to polar coordinate system to conduct it in practice conveniently. Theorem 4 demonstrates the noise form in polar coordinate system.

Theorem 4 The PDF of the two dimension Laplace noise with a relative error α is:

$$D_{\epsilon}(r, \theta) = \frac{\epsilon^2}{2\pi[1 - (\epsilon\alpha d(p, o) + 1)e^{-\epsilon\alpha d(p, o)}]} r e^{-\epsilon r} \quad (22)$$

Proof We call it as a truncated two dimensional Laplace distribution. We proof that the integral of this PDF over the bound gives 1, which means that the PDF in Theorem 4 is the PDF of a probability distribution:

$$\begin{aligned} F_{\epsilon}(r, \theta) &= \int_0^{\alpha d(p, o)} \int_0^{2\pi} \frac{\epsilon^2}{2\pi[1 - (\epsilon\alpha d(p, o) + 1)e^{-\epsilon\alpha d(p, o)}]} r e^{-\epsilon r} dr d\theta \\ &= \int_0^{\alpha d(p, o)} \frac{\epsilon^2}{1 - (\epsilon\alpha d(p, o) + 1)e^{-\epsilon\alpha d(p, o)}} r e^{-\epsilon r} dr \\ &= \frac{\epsilon^2}{1 - (\epsilon\alpha d(p, o) + 1)e^{-\epsilon\alpha d(p, o)}} \frac{e^{-\epsilon r}}{\epsilon^2} (-\epsilon r - 1) \Big|_0^{\alpha d(p, o)} \\ &= 1. \end{aligned} \quad (23)$$

Now we notice that the truncated polar coordinate Laplace defined above has a very convenient feature to draw: the angle and radius are independent with each other. That is, we can use two margins to express the PDF. Indeed, if we use r (radius) and θ (angle) to represent these two random variables. The two margins are:

$$\begin{aligned} D_{\epsilon, R}(r) &= \int_0^{2\pi} D_{\epsilon}(r, \theta) d\theta \\ &= \frac{\epsilon^2}{2\pi[1 - (\epsilon\alpha d(p, o) + 1)e^{-\epsilon\alpha d(p, o)}]} r e^{-\epsilon r}. \end{aligned} \quad (24)$$

$$D_{\epsilon, \Theta}(\theta) = \int_0^{\alpha d(p, o)} D_{\epsilon}(r, \theta) dr = \frac{1}{2\pi}. \quad (25)$$

Hence we have $D_{\epsilon}(r, \theta) = D_{\epsilon, R}(r) D_{\epsilon, \Theta}(\theta)$. $D_{\epsilon, R}(r)$ indicates the PDF of truncated gamma distribution whose shape and scale are 2 and $1/\epsilon$ respectively with a boundary $\alpha d(p, o)$.

Since Θ and R are independent with each other, it is necessary to generate r and θ from $D_{\epsilon, R}(r)$ and $D_{\epsilon, \Theta}(\theta)$ respectively if we want to generate a point (r, θ) from $D_{\epsilon}(r, \theta)$.

The drawing method of truncated gamma distribution is different with that of truncated Laplace distribution. We first consider the cumulative function $C_{\epsilon}(r)$ of $D_{\epsilon, R}(r)$:

$$\begin{aligned} C_{\epsilon}(r) &= \int_0^r \frac{\epsilon^2}{1 - (\epsilon\alpha d(p, o) + 1)e^{-\epsilon\alpha d(p, o)}} \rho e^{-\epsilon \rho} d\rho \\ &= \frac{1 - (1 + \epsilon r)e^{-\epsilon r}}{1 - (\epsilon\alpha d(p, o) + 1)e^{-\epsilon\alpha d(p, o)}}. \end{aligned} \quad (26)$$

We draw a random variable ρ that conforms to uniform distribution in the interval $[0, 1)$, and set $r = C_{\epsilon}^{-1}(\rho)$. Then the random number ρ that exceeds α will be throw out and the number falls in the range $(0, \alpha d(p, o)]$ conforms to the distribution of $D_{\epsilon, R}(r)$. $\square$

Statistical properties

The $d(p, p')$ in Equation (21) indicates the distance between p and p' , then the PDF of truncated planar Laplacian distribution in Equation (21) equals to the following form:

$$f(x, y) = \frac{\epsilon^2}{2\pi[1 - (\epsilon\alpha d(p, o) + 1)e^{-\epsilon\alpha d(p, o)}]} e^{-\epsilon \sqrt{x^2 + y^2}}, \quad (27)$$

where $\sqrt{x^2 + y^2} \in (0, \alpha d(p, o)]$.

Since x and y are symmetrical around the true position. Thus, the mean of the two dimensional PDF does not change,

which is the same as that of Geo-indistinguishability. Next we investigate the change of variance and MSE in terms of our truncated planar Laplacian mechanism.

Let $B = \frac{1}{2\pi[1-(\epsilon\alpha d(p,o)+1)e^{-\epsilon\alpha d(p,o)}]}$, then the Equation (21) can be transformed to

$$D_\epsilon(r, \theta) = B\epsilon^2 r e^{-\epsilon r}, r \in (0, \alpha d(p, o)]. \quad (28)$$

Then the variance of $D_\epsilon(r, \theta)$ is

$$\begin{aligned} \sigma^2(r, \theta) &= \int_0^{\alpha d(p,o)} B\epsilon^2 r^3 e^{-\epsilon r} e^{tr} dr \\ &= \frac{B}{\epsilon^2} (\epsilon r)^3 e^{-\epsilon r} d(\epsilon r) \\ &= \frac{B}{\epsilon^2} \gamma(4, \epsilon\alpha d(p, o)) \end{aligned} \quad (29)$$

where $\gamma(\cdot)$ is the incomplete gamma function.

Equation (29) indicates that the variance of truncated gamma random variable is changed. According to the properties of gamma distribution, a linear transformation of gamma distribution still conforms to gamma distribution. We utilize this property to generate truncated noise to make the variance not changed.

Theorem 5 If $\omega = \sqrt{\frac{2}{B\gamma(4, \epsilon\alpha d(p, o))}}$, $R = \omega r$, then the variance of R is

$$\sigma_R^2 = \frac{2}{\epsilon^2}, \quad (30)$$

equals to the variance of random variable r in Geo-indistinguishability.

Proof If $R = \omega r$, then the PDF of R is

$$f_R(r) = f\left(\frac{r}{\omega}\right) \frac{1}{\omega} = B\epsilon^2 \frac{r}{\omega} e^{-\epsilon \frac{r}{\omega}} \frac{1}{\omega}, \quad (31)$$

then the variance of R is

$$\begin{aligned} \sigma_R^2 &= \int_0^{\omega\alpha d(p,o)} \frac{\omega^2}{\epsilon^2} B \left(\epsilon \frac{r}{\omega}\right)^3 e^{-\epsilon \frac{r}{\omega}} d\left(\epsilon \frac{r}{\omega}\right) \\ &= \frac{\omega^2}{\epsilon^2} B\gamma(4, \epsilon\alpha d(p, o)) \\ &= \frac{2}{\epsilon^2}. \end{aligned} \quad (32)$$

Similar with that case in one dimension, since the mean of truncated planar Laplacian mechanism does not have deviation, and the variance of the random variable R do not change after applying our truncated mechanism and linear transformation, the MSE can be regarded as an unbiased estimation to standard Laplacian noise. In other words, the MSE of

our mechanism is also unchanged compared to that of Geo-indistinguishability mechanism. Algorithm 2 shows the working flow of our solution RE-Geo-indistinguishability.

Algorithm 2 RE-Geo-indistinguishability

Require: ϵ, α, p .

Ensure: p' .

1. Set $r \leftarrow C_\epsilon^{-1}(\rho)$ according to Equation(28);
 2. Set $R = \omega r$, where $\omega = \sqrt{\frac{2}{B\gamma(4, \epsilon\alpha d(p, o))}}$;
 3. Draw p' around p with radius R and a random angle;
- return** p' .
-

Experimental evaluation

In this section, we evaluate the performance of the proposed solution on multiple real datasets. Specifically, we aim to explore the answers of the following questions:

- *How does our solution performance on the statistical properties, including mean, variance and MSE on different datasets?*

Since our proposed RE-DP and RE-Geo-indistinguishability mechanisms can keep the statistical properties not changed, while limiting relative error in the fixed range. In sections “Mean”, “Variance” and “MSE”, we will evaluate the mean, variance and MSE of our proposal and compare the results with current representative approaches.

- *How much is the utility superiority of our solution on the data in terms of polytype queries?*

We use the index (α, β) -accuracy to evaluate the performance of data utility on multiple applications. In section “ (α, β) -accuracy”, we evaluate the performance in different types of applications.

Datasets and configuration

We evaluate our proposed solution by experimenting with real datasets. We plan to prove the effectiveness of our solution through four real world datasets in machine learning, social networking and transportation applications. Each experiment was performed 1000 times.

Adult: Thanks to UC Irvine Machine Learning Repository¹, this dataset predicts whether income exceeds 50K/yr based on census data, which is also known as “Census Income” dataset. Its total number of instances and attributes are 48,842 and 14 respectively.

¹ <https://archive.ics.uci.edu/ml/index.php>.

Fig. 4 Mean of noise on different datasets. **a** Adult. **b** Social Network. **c** Check-in. **d** Trajectory

Social Network [28]: The dataset collected friendship relationships among 32,768 students from an online social networking site. Each student has up to 1678 friends.

Check-in [29]: The Check-in dataset collects the timestamp, user ID, location and location type from 31,000 social users online and 49,000 Americans offline in Los Angeles, New York.

Trajectory: In terms of Geolife project [30], this dataset is consisted of 17,621 trajectories. Its distance is 1,292,951 km and duration is 50,176 h. Tracks contain longitude, latitude, altitude coordinates, and time stamps.

Among the four datasets, the adult is a machine learning dataset to test the performance of our solution on machine learning results. Social network is a typical one dimension dataset to evaluate the performance of our RE-DP. Check-in and trajectory are two-dimensional datasets used to measure the performance of our RE-Geo-indistinguishability solution.

To evaluate the effectiveness of our solution, we compare our solution with standard DP and 3 state-of-the-art mechanisms, which are Kasra [26], Arinjita [24], and Kasra [25]. Besides the statistical properties, mean, variance and MSE, the data utility was measured by the index, (α, β) -accuracy.

Mean

The mean value of the noise can most intuitively reflect the extent to which the added noise affects the original data. To evaluate the effect of the introduced noise, we measure the mean of the noise introduced by state-of-the-art schemes, including standard DP Geo-indistinguishability. Figure 4 shows the box plot results on four test datasets. We set the privacy budget $\epsilon = 1$ and conduct the experiments on all datasets.

As shown in Fig. 4, from the top to the bottom of the box plot is the maximum, upper quartile, median, lower quartile and minimum value respectively. In addition, circle is the abnormal value. From Fig. 4, we observe that the maximum

Fig. 5 Variance of noise on different datasets. **a** Adult. **b** Social Network. **c** Check-in. **d** Trajectory

and minimum values of our solution have minimum fluctuation except on the Social Network dataset. The reason of the big fluctuation on Social Network dataset is that the node and edge are strongly connected. Change of one node or edge degree will affect others' value. Besides, means of noise of our solution is approximately 0 on all of the testing datasets, indicating that our solution has not changed the mean of original data intensely. Moreover, all of the abnormal values are in the box, indicating that current methods do not generate outliers. We also observe that the upper and lower quartiles of our solution have no big difference with the maximum and minimum values, demonstrating that our solution does not change the data distribution obviously. The experimental results show that our solution has optimal performance on the mean value of noise compared with current methods.

Variance

This subsection compares the variance of noise on different datasets with current mechanisms. According to Dwork [9], ϵ less than 1 will be appropriate for privacy intensity setting, then we will follow this heuristic in the following experiments. To investigate our solution's performance comprehensively, we set the range of privacy budget from 0.1 to 0.9 and vary it with a 0.2 step on four datasets, holding various privacy preserving levels.

Figure 5 also shows the results of variance of noise on different datasets corresponding to standard one and two dimensional DP and other mechanisms. As shown in Fig. 5, we find that our solution has a lower variance of noise than the other schemes on four datasets. Specially, in one dimensional dataset, such as Fig. 5a, when $\epsilon = 0.1$, RE-DP has a variance of 85 while the current optimal mechanism Kasra [2020] achieves 124, with an improvement of 31.5%. When

Fig. 6 MSE of noise on different datasets. **a** Adult. **b** Social Network. **c** Check-in. **d** Trajectory

$\epsilon = 0.9$, RE-DP achieves a variance of 1.6 while Kasra [2020] achieves a variance of 1.7, an improvement of 5.9%. This improvement can also be observed in two dimensional dataset. For example, in Fig. 5c, when $\epsilon = 0.1$, RE-Geo-indistinguishability has a variance of 98 while the current optimal mechanism Kasra [2020] is 160, increases 38.8%. When $\epsilon = 0.9$, RE-Geo-indistinguishability achieves a variance of 0.7 while Kasra [2020] achieves a variance of 2.1, an improvement of 66.7%. Besides, compared with current schemes, the variance of our solution is closest to standard DP and Geo-indistinguishability, indicating that our schemes maintain the variance of the original data well. The reason is that we use a linear transformation of Laplace and Gamma distribution to generate the required noise while keep the variance unchanged.

MSE

As shown in Fig. 6, we observe that our solutions are optimal whether in one-dimensional or two-dimensional data. Specif-

ically, in one dimension, for example Fig. 7a, when $\epsilon = 0.1$, our proposal RE-DP achieves a MSE of 215 while suboptimal mechanism except DP achieves 324, with an improvement by 33.6%. When $\epsilon = 0.9$, RE-DP's MSE is 2.6 and outperforms Kasra[2020] by 62.9%. The improvement of our solution can also be observed in two dimension. In Fig. 6c, our mechanism RE-Geo-indistinguishability's MSE outperforms the suboptimal mechanism except for Geo-indistinguishability by 49.8% when $\epsilon = 0.1$ and 20.0% when $\epsilon = 0.9$. These results support the conclusion that our solution can keep the MSE at a small level either in one or two dimensional data. Our proposed mechanisms perform better because they keep the statistical properties of the noise, including the mean, variance not changed and limit the noise error to a fixed bound.

Figure 6 also shows the effect of our solution compared with standard DP and Geo-indistinguishability. For example, in adult dataset, the MSE of standard DP is when $\epsilon = 0.1$, which is close to the MSE, 215 of RE-DP. This case is similar in two dimensional dataset. In Check-in dataset, the MSE of

Fig. 7 (α, β) -accuracy on different datasets. **a** Adult. **b** Social Network. **c** Check-in. **d** Trajectory

Geo-indistinguishability is 298 when $\epsilon = 0.1$, while that of our mechanism is 213. The same cases can also be found in Fig. 6b, d. These results demonstrate that our solution has closest statistical error to standard one and two dimensional DP, compared with current methods. This indicates that our solution can preserve a better data utility in multiple real-world applications.

(α, β) -accuracy

Figure 7 shows the utility performance, (α, β) -accuracy, under privacy preserving setting $\epsilon = 1$. We conclude that our solution has much lower β in experimental one and two dimensional dataset, i.e., with higher probability $1 - \beta$ under the same α and ϵ , than current mechanisms. Therefore, as privacy preserving level ϵ increases, the accuracy of our solution will increase. When $\alpha = 10$, $\epsilon = 1$, the value of $1 - \beta$ for our solution drops to nearly 0 while that values for other methods are not 0 in terms of this probability. Similarly, from Fig. 7,

we can see that under the same utility constraint, our solution can also provide significantly better privacy guarantee than the current mechanisms. Therefore, our solution shows significant utility and privacy superiority over state-of-the-art approaches.

Summary for the experimental analysis

Experimental evaluation in sections “Mean”, “Variance”, “MSE” and “ (α, β) -accuracy” supports the following conclusions:

- Our solution can keep the statistical properties, including mean, variance and MSE, close to the standard one dimension DP and two dimension Geo-indistinguishability mechanisms, indicating that our solution has a good performance on the statistical properties while limit the noise error to a fixed range.

- Compared with state-of-the-art schemes, our solution performs best on statistical properties, demonstrating the effectiveness of our solution.
- Our solution retains significant utility gains with the same privacy budget compared with the existing approaches. Therefore, our solution can obtain a good trade-off between privacy protecting and data utility by selecting a appropriate privacy budget ϵ .

Conclusions and future works

Differential privacy provides a good trade-off between privacy preserving and data utility. For this reason, an emerging consensus around its applications and possible extensions in the academic institution and privacy community is being pursued. However, DP has no constraint on the relative error of the noise, leading to disruption of data availability. State-of-the-art schemes attempt to improve data utility while preserving DP or design noise that satisfies DP while limiting the relative error to a fixed bound. But they do not consider the effect of their methods on the statistical properties.

In this paper, we first give two definitions of DP with limited relative error of noise in one and two dimension respectively. Then we propose the corresponding mechanisms, to realize the definitions in practice. Furthermore, we utilize linear transformation of Laplace and Gamma distributions to maintain the statistical properties while limiting the relative error to a fixed bound. Experimental evaluation demonstrates that our mechanism outperforms current schemes in terms of security and utility for numbers of frequent items.

The limitation of our work mainly includes two parts. The first is that our work can only adapt count query or histogram scenarios. One of our future works is to explore the possibility of extending our method to complex intelligent algorithms. Other interesting extension of our work would be the need of mechanisms to design corresponding improved algorithms according to the needs of practical applications.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China (42001398), Chongqing Natural Science Foundation (cstc2020jcyj-msxmX0635), Science and Technology Research Project of Chongqing Education Commission (KJQN201900612), Open Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University (20S02), China Postdoctoral Science Foundation (2021M693929), the PhD Starts Fund Project of Chongqing University of Posts and Telecommunications (A2020-46) and SRTF of CQUPT(A2020-106).

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Wang W, Huang H, Zhang L, Su C (2020) Secure and efficient mutual authentication protocol for smart grid under blockchain. *Peer Peer Netw Appl* 14:1–13
2. Wang W, Xu H, Alazab M, Gadekallu TR, Han Z, Su C (2021) Blockchain-Based Reliable and Efficient Certificateless Signature for IIoT Devices. *IEEE Trans Ind Inf* 10.1109/TII.2021.3084753
3. Wang H, Z Xu (2017) CTS-DP: publishing correlated time-series data via differential privacy. *Knowl-Based Syst* 122:167–179
4. Cynthia D (2011) A firm foundation for private data analysis. *Commun ACM* 54(1):86–95
5. Jing L, Xiaohui K, Shujie L, Xu M, Yi T (2020) Privacy preservation for machine learning training and classification based on homomorphic encryption schemes. *Inf Sci* 526:166–179
6. Zhihong T, Yuhang W, Yanbin S, Jing Q (2020) Location privacy challenges in mobile edge computing: classification and exploration. *IEEE Net* 34(2):52–56
7. Zhang L, Zhang Z, Wang W, Jin Z, Su Y, Chen H. Research on a covert communication model realized by using smart contracts in blockchain environment. *IEEE Syst J* 10.1109/JSYST.2021.3057333
8. Dwork C (2006) Differential privacy. In *Proc 33rd Int Conf Autom Lang Programm (ICALP)*, pp:1–12
9. Dwork C (2008) Differential privacy: a survey of results. In *Proc 5th Int Conf Theory Appl Mod Comp (TAMC)* 4978:1–19
10. Xiao X, Bender G, Hay M, Johannes G (2011) ireduct: Differential privacy with reduced relative errors. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pp:229–240
11. Gautz M, Machanavajjhala A, Wang G, Xiao X, Gehrke J (2011) Publishing search logs - a comparative study of privacy guarantees. *IEEE Trans Knowl Data Eng* 99:520–532
12. Korolova A, Kenthapadi K, Mishra N, Ntoulas A (2009) Releasing search queries and clicks privately. In *International World Wide Web Conference (WWW2009)*, pp:172–180
13. Danish V, Mamoun A, Sobia W, Babak S, Qin Z (2020) Image-based malware classification using ensemble of cnn architectures (imcec). *Comp & Secur* 92:101748
14. Muhammad N, Fazli S, Wazir ZK, Saqib H, Sajjad H, Thippa RG, Alireza J, Mamoun A (2020) A systematic review on clone node detection in static wireless sensor networks. *IEEE Access* 8:65450–65461
15. Quoc VM, Swarna RM, Thien H (2021) Fusion of federated learning and industrial internet of things: A survey. [arXiv:2101.00798](https://arxiv.org/abs/2101.00798)
16. Wang H, Huan W (2021) Correlated tuple data release via differential privacy. *Inf Sci* 560:347–369
17. Barak B, Chaudhuri K, Dwork C, Kale S, McSherry F, Talwar K (2007) Privacy, accuracy, and consistency too: a holistic solution

- to con-tingency table release. In Proc. of ACM Symposium on Principles of Database Systems, pp:273-282
18. Kasiviswanathan SP, Rudelson M, Smith A, Ullman J (2010) The price of privately releasing contingency tables and the spectra of random matrices with correlated rows. In Proc. of ACM Symposium on Theory of Computing, pp:775-784
 19. Blum A, Ligett K, Roth A (2008) A learning theory approach to non-interactive database privacy. In Proc. of ACM Symposium on Theory of Computing, pp:609-618
 20. Michael H, Vibhor R, Gerome M, Dan S (2010) Boosting the accuracy of differentially private histograms through consistency. Proc VLDB Endow 3(1-2):1021–1032
 21. Xiao X, Wang G, Gehrke J (2010) Differential privacy via wavelet transforms. In Proc. of International Conference on Data Engineering, pp:225-236
 22. Li C, Hay M, Rastogi V, Miklau G, McGregor A (2010) Optimizing linear counting queries under differential privacy. In Proc. of ACM Symposium on Principles of Database Systems, pp:123-134
 23. Ligett K, Neel S, Roth A, Waggoner B, Wu S (2017) Accuracy first: Selecting a differential privacy level for accuracy constrained erm. In 31st Conference on Neural Information Processing Systems(NIPS 2017), pp:11-27
 24. Arinjita P, Vorapong S, Mitali B, PanduRangan C (2020) Improving accuracy of differentially private kronecker social networks via graph clustering. In 2020 International Symposium on Networks, Computers and Communications (ISNCC), pp:26-35
 25. Kasra Y, Matthew H (2020) Error bounds and guidelines for privacy calibration in differentially private Kalman filtering. In 2020 American Control Conference (ACC), pp:89-95
 26. Fang L (2019) Generalized Gaussian mechanism for differential privacy. IEEE Trans Knowl Data Eng 31:747–756
 27. Chatzikokolakis K, Palamidessi C, Stronati M (2015) Geo-indistinguishability: A principled approach to location privacy. In International Conference on Distributed Computing and Internet Technology, pp:49–72
 28. Hay M, Rastogi V, Miklau G, Suciu D (2010) Boosting the accuracy of differentially private histograms through consistency. Proc. VLDB Endowment, 3(2):1021-1032
 29. Elahe G, Mahdi A, Fatemeh D (2016). Pptd: Preserving personalized privacy in trajectory data publishing by sensitive attribute generalization and trajectory local suppression. Knowl-Based Syst 94:43–59
 30. Zheng Y, Xie X, Ma WY (2010) Geolife: A collaborative social networking service among user, location and trajectory. Bulletin Tech Comm Data Eng 33(2):32–39

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.