Front.Comput.Sci.
DOI

RESEARCH ARTICLE

An Optimal Differentially Private Data Release Mechanism with Constrained Error

Hao WANG[1], Zhengquan XU[2], Xiaoshan ZHANG[1], Xiao PENG[1] and Kaiju LI(✉)[3]

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, Chongqing 400065, China
2 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
3 College of Computer Science, Chongqing University, Chongqing 400044, China

Abstract Privacy preserving methods supporting for data fusion 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. However, DP has no limit on the bound of introduced noise, leading to a low-level data utility. Recently, researchers explore how to achieve a good trade-off between privacy and utility, but the optimal result is still worth investigated. In this paper, we explore an optimal differential privacy mechanism for data release with constrained error. We first propose truncated noise mechanism to limit the noise introduced by DP to a fixed bound. Then we take advantage of particle filter to sanitize the perturbed results to obtain the optimal data utility. Experimental evaluation demonstrates that our mechanism outperforms current schemes in terms of security and utility for plenty of queries, while maintaining the privacy requirement of DP.

Keywords Privacy preserving, Data release, Differential privacy, Constrained error, Particle filter

1 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. For example, in location based applications, behaviors aggregating one's precise position to service provider can be used to get better shopping recommendations and route planning. 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 [1], [2].

Early privacy preserving schemes inherently rely on the security guarantee of designed algorithms, where the security is difficult to be theoretically proved and analyzed. To remedy this problem, DP proposed by Dwork [3], [4] 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.

DP controls protection strength by using a parameter ϵ. A small ϵ means a good protection. But a small ϵ will inject high-level noise into the data to be protected, leading to low-level data utility. 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

Received month dd, yyyy; accepted month dd, yyyy

E-mail: *dlikaiju@cqu.edu.cn*

have a great influence on the accuracy of mining results [5]. 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 the data utility is destroyed, as illustrated in Fig. 1 and Example 1.

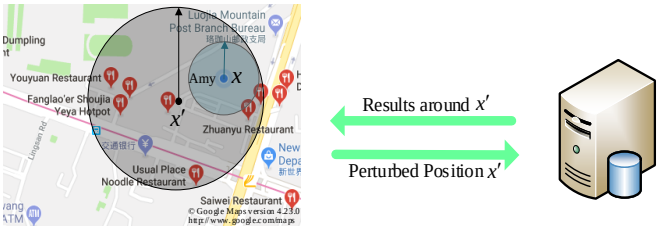


Fig. 1: Effect of true and perturbed positions on the mining results

Example 1. Consider a scenario that Amy 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 analyzing 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 explore the optimal noisy mechanisms while preserving DP.

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.
- **Data Utility:** The other privacy-first schemes try to improve the accuracy of publishing results by multiple schemes while meeting DP. But as far as we know, the release method to achieve optimal data availability is still worth investigated.

These challenges imply that an optimal mechanism with constrained error for differentially private data release is in high demand. With respect to the first challenge, to strictly meet the requirement of DP, we generate the needed noise in a truncated form. The truncated distribution can satisfy DP, while limiting the noisy error to a fixed bound. For the second challenge, we take advantage of the particle filter, which is the optimal filter to address non-Gaussian noise, to obtain the optimal availability of published data. Furthermore, we extend the particle filter to polar coordinate system, to deal with two dimensional data.

Based on these considerations, we propose an optimal error constrained differentially private data release solution. We first propose the truncated mechanisms in one and two dimensions respectively, to limit the error to a fixed bound while satisfying DP. Then we design optimal particle filters in one and two dimensions to obtain the best publishing results. To the best of our knowledge, our solution is the first technique that renders truncated mechanisms and filtering to obtain the optimal publishing accuracy while satisfying DP. Our contributions are threefold:

- **Error Constrained DP:** Standard DP has no limitation on error constraint, leading to an unexpected damage on data utility. We extend DP to the notions of AE-DP and AE-Geo-indistinguishability, which are the expansions of DP in one and two dimensional form with error constraint respectively. AE-DP and AE-Geo-indistinguishability limits the noise error to a fixed bound, while preserving ϵ-DP.
- **Truncated Mechanism:** We propose two truncated mechanisms, called "truncated Laplacian mechanism" and "truncated planar Laplacian mechanism", to realize AE-DP and AE-Geo-indistinguishability in practice. These two truncated mechanisms can limit the noisy error to a fixed bound while preserving ϵ-DP in one and

two dimensional form.

- **Optimal Filtering:** We propose two filtering schemes to sanitize perturbed publishing results. We take advantage of particle filter to deal with non-Gaussian noise and extend it to a polar form, to obtain optimal data utility in one and two dimensions.

The rest of this paper is arranged as follows. In Section 2, we introduce the mechanisms associated with our work. Then notations and preliminaries adopted in this work are described in Section 3. Section 4 demonstrates the truncated mechanisms in one and two dimensions respectively. Our designed optimal particle filters to sanitize the perturbed results are described in Section 5. Security and utility analysis of our solution are demonstrated in Section 6. The experimental evaluation was performed in Section 7, while conclusions and future work are in Section 8.

2 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.

2.1 Privacy-first

When publishing various sensitive data (such as search logs [6] [7], data mining results [8] [9], 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 [8]. Next, we will discuss some typical methods of privacy priorities.

Barak et al. [10] 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. Kasiviswanathan et al. [11] 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.

Blum et al. [12] presented a 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. Literature [13] proposed an method to improve the performance range of Blum et al.'s scheme. Furthermore, Xiao et al. [14] designed a method for multidimensional data publishing.

Li et al. [15] summarized the methods in [13] and [14]. 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.

2.2 Accuracy-first

Ligett et al. [16] proposed a framework 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.

The most relevant work is the truncated Gaussian mechanism proposed by Liu [17]. According to the definition of 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 pri-

vacy requirement of DP. But these methods have the risk of violating DP.

2.3 Summary

The privacy and accuracy are always a couple of contradiction in private data publishing. 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 constrained error. Although accuracy first methods can preserve good utility, the generated noise 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 error constrained DP in one and two dimensions respectively?
- How to guarantee the privacy requirement of DP with an error constrained noise?
- What is the upper bound of the utility loss caused by the noise and how to obtain the optimal sanitizing publishing results?

3 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.

3.1 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 1 lists the main notations used in this paper.

3.2 Differential privacy

DP [3] is a state-of-the-art privacy preservation model which can guarantee the security of indistinguishability. Essentially,

Table 1: 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	Constrained error bound.
w	Parameter of the particle filter.
B	Coefficient of the truncated Planar Laplace mechanism.
r, θ	Perturbed radius and angle in two dimension.
u	Random noise generated in one dimension.

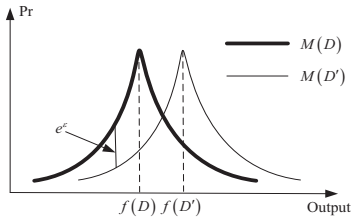


Fig. 2: Probability density function of random algorithm M on the statistical output of D and $D^{'}$

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) Considering two adjacent datasets, Dand $D^{'}$, 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 on Dand $D^{'}$, then M satisfies ϵ-DP.*

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

where $S \subseteq Range(M)$, $Range(M)$ 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. Fig. 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. In fact, Laplace mechanism is usually used to realize M. The Laplace mechanism is defined as follows.

Definition 2. *(Laplacian Mechanism) Let $f(\cdot)$ be the statistical function of the output result. The noisy samples $Z \sim Lap(\lambda)$ 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

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

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

$$\lambda = \frac{\Delta f}{\epsilon}, \tag{3}$$

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

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

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 $Lap(1/\epsilon)$ is enough to guarantee ϵ-DP.

Definition 3. *(Geo-indistinguishability [18]) For all observations S, a mechanism satisfies ϵ-Geo-indistinguishability if:*

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

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) 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:*

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

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.

4 AE-DP

In this section, we first propose the definition of AE-DP, which satisfies DP with constrained bound. 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 use a particle filter to sanitize the perturbed results to obtain the optimal results.

4.1 Definition of AE-DP

As we have discussed in "Introduction" section, the SP may always has requirement on the noisy error. In this case, the definition of DP is not appropriate. In this section, We propose the notion of AE-DP with error constraint. First of all, we give the definition of constrained bound.

Definition 5. *(Constrained Bound) Denote z is a random noise generated by privacy preserving method, z is limited by the bound α, i.e. $|z| \leq \alpha$, where z is generated with ϵ-AE-DP. Then we can say that the absolute error of z is α.*

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

Definition 6. *(AE-DP) Considering two adjacent datasets, Dand 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*

$$Pr[M(D) \in S \| z| \leq \alpha] \leq e^\epsilon Pr[M(D') \in S \| z| \leq \alpha], \tag{7}$$

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

4.2 Truncated Laplace Mechanism

Definition 6 gives the formal definition of AE-DP. Then we propose a truncated Laplace mechanism to realize AE-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 AE-DP.*

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

where $\lambda = \frac{\Delta f}{\epsilon}$.

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

The proof that AE-DP can maintain DP is in section 6. Algorithm 1 shows the working flow of AE-DP.

Algorithm 1 AE-DP

Require: ϵ, $Q = \{q_1, q_2, \cdots, q_n\}$, Δf.
Ensure: $Z = \{z_1, z_2, \cdots, z_n\}$, perturbed query results Q'.
for each round $k \leftarrow 1, \cdots, n$ **do**
1. Select a query q_k;
2. Compute noise scale parameter $\lambda = \Delta f/\epsilon$;
3. Generate truncated Laplace noise z_k according to the PDF in Definition 7 with bound α;
4. Compute the noisy response $q'_k = q_k + z_k$;
end for
return Q'.

4.3 Particle filtering for Truncated Laplace noise

After the truncated mechanism, we have limit the noise to a setting bound. In this section, we explore the optimal method to address the noisy results. According to the property of DP, the post processing of perturbed data still conforms to DP. We regard perturbed data as a time series, and filtering is a good method to obtain the optimal estimate result. Thus, we consider the process of sanitizing perturbed data as a filtering process to get the optimal estimate result. Then the problem is to explore the best filtering method in terms of truncated Laplace noise.

According to the theory of signal processing, the optimal practical filter to address non-Gaussian noise is the particle filter, whose process is easy to follow. Thus, in this paper, we utilize the particle filter to sanitize perturbed data.

In particle filtering technology, posterior probability is expressed by utilizing the weighted sum of some random samples, and integral operation is approximated by the. Theorem 1 shows the working flow of particle filtering.

Particle filtering takes advantage of the weighted sum of a series of random samples to express the posterior probability and the integral operation is approximated by this summation. Theorem 1 gives the process of particle filtering.

Theorem 1. *Let q'_k denote the observation value of the perturbed output corresponding to the query result q_k. Set $q(\hat{q}_k|q'_{1:k})$ as an important PDF, and its sample $\hat{q}_k$ is easy to produce, such as conforming to uniform or Gaussian distribution. Suppose that random variables $\hat{q}^i_k, i = 1, \ldots, M$ can be sampled from posterior probability $p(\hat{q}_k|q'_{1:k})$. Then, the estimate value $\hat{q}_k$ which meets LMS(Least Mean Square) criterion is:*

$$\hat{q}_k = \frac{1}{M}\sum_{i=1}^{M} \hat{q}^i_k w^i_k. \tag{9}$$

where w^i_k is a formalized weighted parameter.

Proof. We can consider $p(\hat{q}_k|q'_{1:k})$ as a approximation value of sampling, then

$$p(\hat{q}_k|q'_{1:k}) = \sum_{i=1}^{M} w^i_k \delta(\hat{q}_k - \hat{q}^i_k),$$

where $\delta(\cdot)$ indicates the Dirac functionunit(unit impulse function), i.e., $\delta(\hat{q}_k - \hat{q}^i_k) = 0$ only if $\hat{q}_k \neq \hat{q}^i_k$ and $\int \delta(x)dx = 1$. w^i_k is calculated by

$$w^i_k \propto \frac{p(\hat{q}_k|q'_{1:k})}{q(\hat{q}_k|q'_{1:k})}.$$

If there are plenty of sampling particles, $p(\hat{q}_k|q'_{1:k})$ can indeed approximate the real posterior PDF. Then the expectation estimate value of $p(\hat{q}_k|q'_{1:k})$ is

$$\begin{aligned}\hat{q}_k &= E[\hat{q}_k|q'_{1:k}] \\ &= \frac{1}{M}\hat{q}^i_k \frac{p(\hat{q}_k|q'_{1:k})}{q(\hat{q}_k|q'_{1:k})} = \frac{1}{M}\sum_{i=1}^{M} \hat{q}^i_k w^i_k.\end{aligned}$$

Algorithm 2 shows the working flow of the particle filtering, where $PF(\cdot)$ indicates the particle filtering.

5 AE-Geo-Indistinguishability

In last section, we have proposed the notion of AE-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 AE-Geo-indistinguishability, to define the two dimensional case with constrained error. Then we give the truncated two dimensional noise form, to satisfy the notion of AE-Geo-indistinguishability. Finally, we give a particle filter in polar coordinate to obtain the optimal result.

Algorithm 2 $\hat{Q}$=$PF(Q')$

Require: Perturbed query result Q'.
Ensure: Optimal estimate $\hat{Q}$.
for each round $k \leftarrow 1, \ldots, N$ **do**
1. Sample $\hat{q}_k^i$ from important distribution $p(\hat{q}_k|q'_{1:k})$ with $i = 1, 2, \ldots, M$;
2. Compute weight $w_k^i \leftarrow w_{k-1}^i \frac{p(q'_k|\hat{q}_k^i)p(\hat{q}_k^i|\hat{q}_{k-1}^i)}{q(\hat{q}_k^i|\hat{q}_{1:k-1}^i, q'_{1:k})}$;
3. Normalize weight as $w_k^i \leftarrow \frac{w_k^i}{\sum_{j=1}^M w_k^i}$;
4. Compute estimate $\hat{q}_k = \frac{1}{M}\sum_{i=1}^M \hat{q}_k^i w_k^i$;
5. Record $\hat{q}_k$;
end for
return $\hat{Q}$.

5.1 Definition of AE-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.

In fact, 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 AE-Geo-indistinguishability with an relative error constraint, as shown in Definition 8.

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

$$\frac{Pr(S|p)}{Pr(S|p')} \leq e^{\epsilon r} \; \forall r > 0 \forall p, p' : d(p, p') \leq r, r \leq \alpha. \tag{10}$$

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

5.2 Truncated planar Laplacian mechanism

In this section, we propose a truncated planar Laplacian mechanism to generate two dimensional Laplace noise with a constrained error α, to satisfy the definition of AE-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*

$$D_\epsilon(p_0)(p) = \frac{\epsilon^2}{2\pi[1 - (\epsilon\alpha + 1)e^{-\epsilon\alpha}]} e^{-\epsilon d(p,p')}, d(p, p') \leq \alpha. \tag{11}$$

where $\epsilon^2/2\pi$ is a normalization factor. Next, we transform the noise form Cartesian coordinate system to polar coordinate system to conduct it in practice conveniently. Theorem 2 demonstrates the noise form in polar coordinate system.

Theorem 2. *The PDF of the two dimensional Laplace noise with constrained error α is:*

$$D_\epsilon(r, \theta) = \frac{\epsilon^2}{2\pi[1 - (\epsilon\alpha + 1)e^{-\epsilon\alpha}]} r e^{-\epsilon r} \tag{12}$$

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 2 is in fact the PDF of a probability distribution:

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

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:

$$D_{\epsilon,R}(r) = \int_0^{2\pi} D_\epsilon(r, \theta) d\theta = \frac{\epsilon^2}{2\pi[1 - (\epsilon\alpha + 1)e^{-\epsilon\alpha}]} r e^{-\epsilon r}. \tag{13}$$

$$D_{\epsilon,\Theta}(\theta) = \int_0^\alpha D_\epsilon(r, \theta) dr = \frac{1}{2\pi}. \tag{14}$$

Hence we have $D_\epsilon(r, \theta) = D_{\epsilon,R}(r) D_{\epsilon,\Theta}(\theta)$. In fact, $D_{\epsilon,R}(r)$ 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 $D_{\epsilon,R}(r)$ 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 $D_{\epsilon,R}(r)$:

$$C_\epsilon(r) = \int_0^r \frac{\epsilon^2}{1-(\epsilon\alpha+1)e^{-\epsilon\alpha}}\rho e^{-\epsilon\rho}d\rho = \frac{1-(1+\epsilon r)e^{-\epsilon r}}{1-(\epsilon\alpha+1)e^{-\epsilon\alpha}}. \tag{15}$$

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 $D_{\epsilon,R}(r)$.

Algorithm 2 shows the working flow of our solution AE-Geo-indistinghuishability.

Algorithm 3 AE-Geo-indistinghuishability

Require: ϵ, α, p.
Ensure: p'.
1. Set $r \leftarrow C_\epsilon^{-1}(\rho)$ according to Equation (15);
2. Draw p' around p with radius r and a random angle from $1/2\pi$;
return p'.

5.3 Particle filtering in Polar Coordinate System

Based on the truncated mechanism, we have limit the noise to a setting bound. In this section, we explore the optimal method to address the noisy results in polar coordinate system. From the last section 5.2, we know that the posterior probability is expressed by utilizing the weighted sum of some random samples, which can obtain the optimal result in theory. Similar with that case in one dimension, we utilize particle filter to address the radius and angle in polar coordinate system respectively. Theorem 3 gives the process of particle filtering in polar coordinate system.

Theorem 3. *Let r_k and θ_k denote the observation value of the perturbed radius and angle respectively. Set $q(\hat{r}_k|r_{1:k})$ and $q(\hat{\theta}_k|\theta_{1:k})$ as important PDFs, and their samples $\hat{r}_k$ are generated by uniform or Gaussian distributions. Suppose that random variables $\hat{r}_k^i$ and $\hat{\theta}_k^i$, $i = 1, \ldots, M$ can be sampled from posterior probability $p(\hat{r}_k|r_{1:k})$ and $p(\hat{\theta}_k|\theta_{1:k})$ respectively. Then the estimates of radius and angle $\hat{r}_{LMS}(k)$, $\hat{\theta}_{kLMS}(k)$ that conform to the LMS(Least Mean Square) criterion are:*

$$\hat{r}_{LMS}(k) = \frac{1}{M}\sum_{i=1}^{M}\hat{r}_k^i u_k^i, \tag{16}$$

$$\hat{\theta}_{kLMS}(k) = \frac{1}{M}\sum_{i=1}^{M}\hat{\theta}_k^i v_k^i. \tag{17}$$

where u_k^i and v_k^i are normalized weight parameters.

Proof. The proof process is similar with that in Theorem 1.

Algorithm 4 describes the working flow of the particle filtering in polar coordinate system, where $PPF(\cdot)$ indicates the particle filtering in polar coordinate.

Algorithm 4 $<\hat{R}, \hat{\Theta}>=PPF(<R, \Theta>)$

Require: Perturbed radius r_k and angle θ_k.
Ensure: Optimal sanitized radius $\hat{r}_k$ and angle $\hat{\theta}_k$.
for each round $k \leftarrow 1, \ldots, N$ **do**
1. Sample $\hat{r}_k^i$ and $\hat{\theta}_k^i$ from important distribution $p(\hat{r}_k|r_{1:k})$ and $p(\hat{\theta}_k|\theta_{1:k})$ respectively with $i = 1, 2, \ldots, M$;
2. Compute weights $u_k^i \leftarrow u_{k-1}^i \frac{p(r'_k|\hat{r}_k^i)p(\hat{r}_k^i|\hat{r}_{k-1}^i)}{r(\hat{r}_k^i|\hat{r}_{1:k-1}^i, r'_{1:k})}$, $v_k^i \leftarrow v_{k-1}^i \frac{p(\theta'_k|\hat{\theta}_k^i)p(\hat{\theta}_k^i|\hat{\theta}_{k-1}^i)}{\theta(\hat{\theta}_k^i|\hat{\theta}_{1:k-1}^i, \theta'_{1:k})}$;
3. Normalize weights as $u_k^i \leftarrow \frac{u_k^i}{\sum_{j=1}^M u_k^i}$, $v_k^i \leftarrow \frac{v_k^i}{\sum_{j=1}^M v_k^i}$;
4. Compute estimate $\hat{r}_k = \frac{1}{M}\sum_{i=1}^M \hat{r}_k^i w_k^i$, $\hat{\theta}_k = \frac{1}{M}\sum_{i=1}^M \hat{\theta}_k^i w_k^i$;
5. Record $\hat{r}_k$, $\hat{\theta}_k$.
end for
return $\hat{R}, \hat{\Theta}$.

6 Security and Utility Analysis

In section 4 and 5, we have proposed the truncated mechanisms to limit the noisy error to a fixed bound, and we also propose the particle filtering in one and two dimensions to obtain the optimal results. In this section, we analyze the security and utility of our mechanisms. Specifically, in terms of security, we prove that AE-DP and AE-Geo-indistinguishability also meet the privacy definition of baseline DP and Geo-indistinguishability. For utility analysis, we deduce the change of noisy variance, which is an base index to measure the performance of utility.

6.1 Security Analysis

In this section, we prove that AE-DP and AE-Geo-indistinguishability satisfy the requirement of DP and Geo-indistinguishability.

6.1.1 Security of AE-DP

Indeed, the definition of AE-DP is a hard version of DP, i.e., it can also satisfy the definition of DP. We first prove that AE-DP also meets DP, as shown in Theorem 4.

Theorem 4. *AE-DP can also preserve ϵ-DP.*

Proof. The proof is similar with that of DP:

$$\begin{aligned}\frac{p_D(z)}{p_{D'}(z)} &= \prod_{i=1}^{n}\left(\frac{\exp(-\frac{\epsilon|q_i(D)-z_i|}{\triangle f})}{\exp(-\frac{\epsilon|q_i(D')-z_i|}{\triangle f})}\right)\\ &= \prod_{i=1}^{n}\exp\left(\frac{\epsilon(|q_i(D')-z_i|-|q_i(D)-z_i|)}{\triangle f}\right)\\ &\le \prod_{i=1}^{n}\exp\left(\frac{\epsilon|q_i(D')-q_i(D)|}{\triangle f}\right)\\ &= \exp\left(\frac{\epsilon\|Q(D)-Q(D')\|_1}{\triangle f}\right)\\ &\le \exp(\epsilon).\end{aligned}$$

That $\frac{p_D(z)}{p_{D'}(z)} \ge \exp(-\epsilon)$ follows by symmetry.

According to the theory of [4], the post processing of the DP still preserves DP. Thus, the sanitized results $\hat{Q}$ after particle filtering still preserves DP. The proof process of AE-Geo-indistinguishability that preserves Geo-indistinguishability is similar with that of AE-DP.

6.2 Utility Analysis

In this section, we analyze the foundational statistical properties mean and variance to measure the utility loss of our mechanisms.

6.2.1 Utility of AE-DP

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 5.

Theorem 5. *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:

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

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

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

Theorem 5 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 6 demonstrates the variety of variance using our proposed mechanism.

Theorem 6. *Given a random variable z that conforms to the truncated Laplacian distribution, the variance of z is:*

$$\sigma_z^2 = \frac{\lambda^2}{1-e^{-\alpha/\lambda}}[2-e^{-\alpha/\lambda}((\frac{\alpha}{\lambda})^2+2\frac{\alpha}{\lambda}+2)]. \tag{18}$$

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

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

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

$$\begin{aligned}\sigma_z^2 &= \int_{-\alpha/\lambda}^{\alpha/\lambda}\frac{\lambda^2}{2(1-e^{-\alpha/\lambda})}y^2e^{-|y|}dy\\ &= \int_{0}^{\alpha/\lambda}\frac{\lambda^2}{1-e^{-\alpha/\lambda}}y^2e^{-y}dy\\ &= \frac{\lambda^2}{1-e^{-\alpha/\lambda}}e^{-y}(-y^2-2y-2)\Big|_0^{\alpha/\lambda}\\ &= \frac{\lambda^2}{1-e^{-\alpha/\lambda}}[2-e^{-\alpha/\lambda}((\frac{\alpha}{\lambda})^2+2\frac{\alpha}{\lambda}+2)]\\ &= \frac{2\lambda^2}{1-e^{-\alpha/\lambda}}\left[1-e^{-\alpha/\lambda}\left(\frac{1}{2}\left(\frac{\alpha}{\lambda}\right)^2+\frac{\alpha}{\lambda}+1\right)\right]\\ &< 2\lambda^2.\end{aligned}$$

Theorem 6 indicates that the variance of truncated Laplace random variable is smaller than that of traditional Laplace mechanism, which may affect quantity of applications.

Indeed, we can design a particle filter by various ways. In this paper, we take the Gaussian distribution as an example, to analyze the least mean square error, as decribed in Theorem 7.

Theorem 7. *Denote the original and perturbed query result as $Q, \hat{Q} \in R$. If an important distribution $q(\cdot)$ whose length is M has a variance σ_q^2, a m length series sampled from this important distribution has the following posterior errors:*

$$\sigma_{Q,\hat{Q}}^2 = \frac{m}{M}\left[\left(\frac{\sigma_q^2}{2\sigma_q^2-1}\right)^{m/2}-1\right], \tag{19}$$

where $\sigma_q^2 > 1$.

Proof. Consider the Monte Carlo method without resampling, we have:

$$\left\{w_k^i\right\}_{i=1}^M \left\{q_k^i\right\}_{i=1}^M = \prod_{i=1}^M w_k^i(q_k^i).$$

Besides,

$$p\left\{q_k^i\right\}_{i=1}^M = \prod_{i=1}^M (2\pi)^{m/2} \exp\left(-\frac{(q_k^i)^2}{2}\right).$$

If the specific distribution conform to

$$q\left\{q_k^i\right\}_{i=1}^M = \prod_{i=1}^M q(q_k^i)$$

and $\sigma_q^2 > 1$, we have $var(\hat{q}_k) < \infty$ and

$$var(\hat{q}_k - q_k) = \frac{1}{M}\left[\left(\frac{\sigma_q^2}{2\sigma_q^2 - 1}\right)^{m/2} - 1\right].$$

If m samples are generated from the defined distribution, their asymptotic variances are

$$\frac{1}{M}\Big(\int \frac{w_m^2(q_1)}{q(q_1)} dq_1 - 1 + \sum_{k=2}^{m} \int \frac{w_m^2(q_1^k)}{w_{k-1}(q_1^{k-1} q(q_k|q_1^{k-1}))} dq_{k-1}^k - 1\Big).$$

Then, the asymptotic variance is finite iif $\sigma_q^2 > \frac{1}{2}$ and

$$\begin{aligned}\sigma^2_{Q,\hat{Q}} &\approx \frac{1}{M}\sum_k^m \Big[\int \frac{w_m^2(q_1)}{q(q_1)} dq_1 - 1 \\ &+ \sum_{k=2}^{m} \int \frac{w_m^2(q_k)}{q(q_k)} dq_k - 1\Big] \\ &= \frac{m}{M}\Big[\Big(\frac{\sigma_q^2}{2\sigma_q^2 - 1}\Big)^{m/2} - 1\Big].\end{aligned}$$

Theorem 7 gives the error after particle filtering. Observing the error in Theorem 7, when the samples m is big enough, the limit of the error $\sigma^2_{Q,\hat{Q}}$ convergence to $\left(\frac{1}{2}\right)^{m/2}$, which means a small variance error. It indicates that our particle filtering is effective in practice to reduce the variance error.

6.2.2 Utility of AE-Geo-indistinghuishability

The $d(p, p')$ in Equation (11) indicates the distance between p and p', then the PDF of truncated planar Laplacian distribution in Equation (11) equals to the following form:

$$f(x, y) = \frac{\epsilon^2}{2\pi[1 - (\epsilon\alpha + 1)e^{-\epsilon\alpha}]} e^{-\epsilon\sqrt{x^2+y^2}}, \tag{20}$$

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

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 $B = \frac{1}{2\pi[1-(\epsilon\alpha+1)e^{-\epsilon\alpha}]}$, then the Equation (14) can be transformed to

$$D_\epsilon(r, \theta) = B\epsilon^2 r e^{-\epsilon r}, r \in (0, \alpha]. \tag{21}$$

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

$$\begin{aligned}\sigma^2(r, \theta) &= \int_0^\alpha B\epsilon^2 r^3 e^{-\epsilon r} e^{tr} dr \\ &= \frac{B}{\epsilon^2}(\epsilon r)^3 e^{-\epsilon r} d(\epsilon r) \\ &= \frac{B}{\epsilon^2}\gamma(4, \epsilon\alpha)\end{aligned} \tag{22}$$

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

Equation (22) indicates that the variance of truncated gamma random variable is changed. The variance error after our polar particle filter is similar with that in Theorem 7.

7 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 AE-DP and AE-Geo-indistinguishability mechanisms can keep the statistical properties not changed, while limiting absolute error in the fixed range. In sub-section 7.2, 3 and 4, we will evaluate the mean, variance and MSE of our proposal and compare the results of our solution 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 subsection 7.5, we investigate this performance on different applications.

7.1 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 Repository [1)], this dataset predicts whether income exceeds 50K/yr based on census data, which is also known as "Census Income" dataset. It total number of instances and attributes are 48,842 and 14 respectively.

Social Network [13]: The dataset collected friendship relationships among 32,768 students from an online social networking site. Each student has up to 1,678 friends.

Check-in [19]: 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 [20], 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.

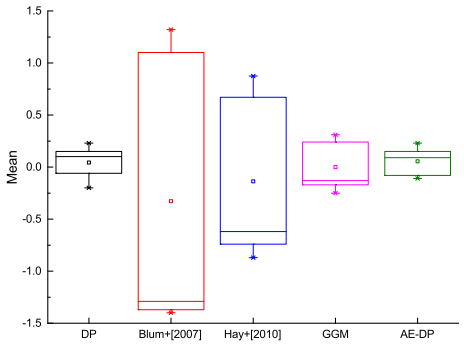
Among the four datasets, the adult is a machine learning dataset to test the performance of our solution on machine learning results. Social network is a typical one dimension dataset to evaluate the performance of our RE-DP. Check-in and trajectory are two-dimensional datasets used to measure the performance 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 Blum+[2008] [12], Hay+[2010] [13], and GGM [17]. Besides the statistical properties, mean, variance and MSE, the data utility was measured by the index, (α,β)-accuracy.

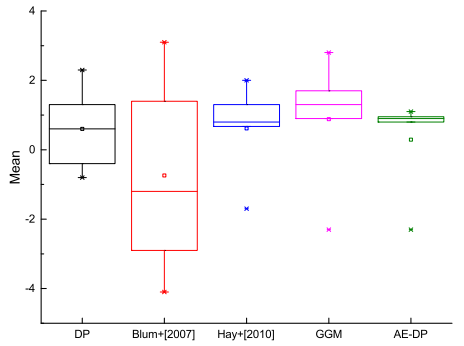
7.2 Mean

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. Fig. 3 shows the box plot results on four test datasets. We set the privacy budget $\epsilon = 1$ and conduct the experiments on all datasets.

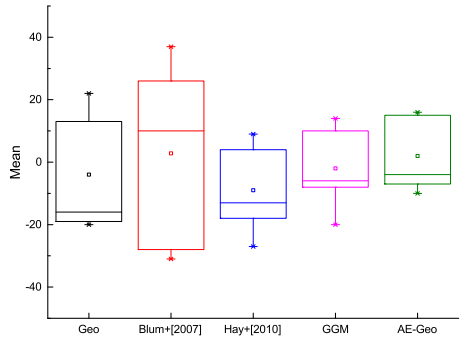
As shown in Fig. 3, the box plot includes the maximum, minimum and median value, upper and lower quartile. From Fig. 3, we observe that the means of noise in current methods, including our solution, are all approximately 0, indicating that the current methods have not change the mean of original data intensely. However, the maximum and minimum values of our solution have minimum fluctuation, i.e.,



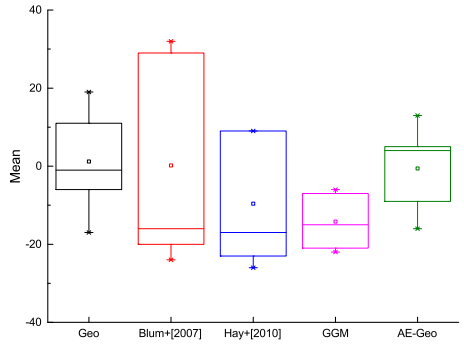
(a) Adult



(b) Social Network



(c) Check-in



(d) Trajectory

Fig. 3: Mean of noise on different datasets

[1)] https://archive.ics.uci.edu/ml/index.php

the mean of noise generated by our solution is closer to 0 compared with current methods. In addition, the horizontal line denotes the median value. From Fig. 3, we observe that the median of our solution is closest to 0. This case indicates that our solution has best stability on the performance of noise mean. Besides, the mean of our solution is closest to standard DP and Geo-indistinguishability, indicating that our solution has not change the mean of noise obviously. The experimental results show that our solution has optimal performance on the mean value of noise compared with current methods.

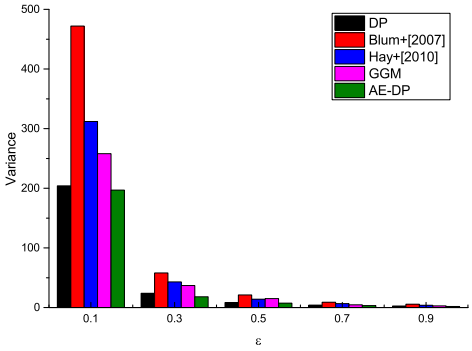
7.3 Variance

This subsection compares the variance of noise on different datasets with current mechanisms. According to Dwork [4], ϵ less than 1 will be appropriate for privacy preserving 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.

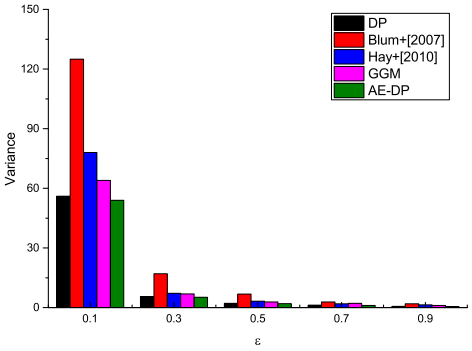
Fig. 4 shows the results of variance of noise on different datasets corresponding to standard one and two dimensional DP and other mechanisms. As shown in Fig. 4, we find that our solution has a lower variance of noise than the other schemes on four datasets. Specially, in one dimensional dataset, such as Fig. 4 (a), when $\epsilon = 0.1$, AE-DP has a variance of 197 while the current optimal mechanism GGM achieves 258, improving 23.6%. When $\epsilon = 0.9$, AE-DP achieves a variance of 1.9 while GGM achieves a variance of 2.8, an improvement of 32.1%. This improvement can also be observed in two dimensional dataset. For example, in Fig. 4 (c), when $\epsilon = 0.1$, AE-Geo-indistinguishability has a variance of 201 while the current optimal mechanism GGM obtains 329, improving of 38.9%. When $\epsilon = 0.9$, AE-Geo-indistinguishability achieves a variance of 3 while GGM achieves a variance of 4.6, an improvement of 34.8%. Besides, compared with current methods, 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.

7.4 MSE

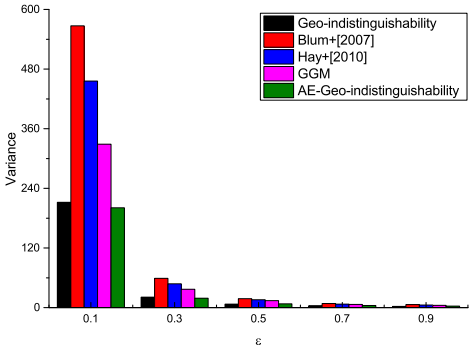
As shown in Fig. 5, we observe that our solutions are optimal Whether in one-dimensional or two-dimensional data.



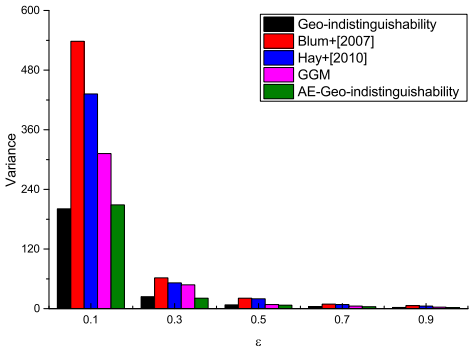
(a) Adult



(b) Social Network

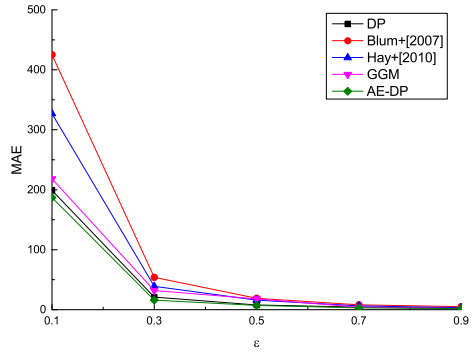


(c) Check-in

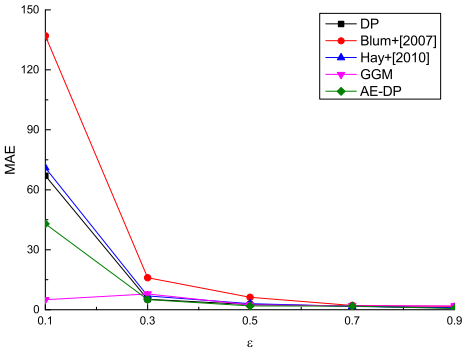


(d) Trajectory

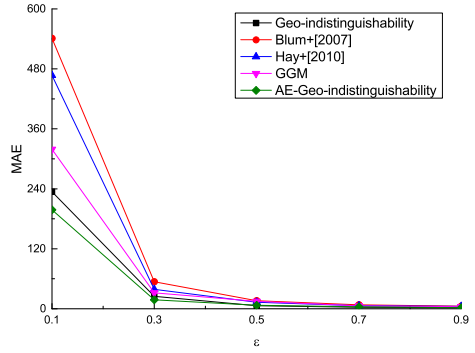
Fig. 4: Variance of noise on different datasets



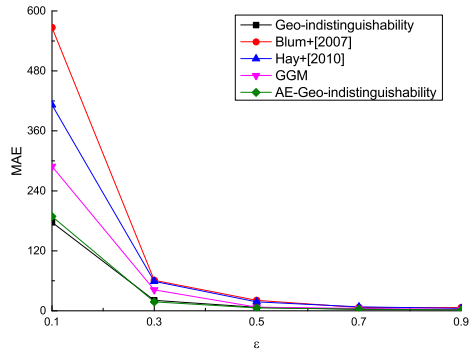
(a) Adult



(b) Social Network



(c) Check-in



(d) Trajectory

Fig. 5: MSE of noise on different datasets

Specifically, in one dimension, for example Fig. 6 (a), when $\epsilon = 0.1$, our proposal AE-DP achieves a MSE of 187 while suboptimal mechanism except DP achieves 218, with an improvement by 14.2%. When $\epsilon = 0.9$, AE-DP's MSE is 1.7 and outperforms GGM by 19.0%. The improvement using our solution can also be observed in two dimension. In Fig. 5 (c), our mechanism AE-Geo-indistinguishability's MSE outperforms the suboptimal mechanism except Geo-indistinguishability by 37.9% when $\epsilon = 0.1$ and 51.2% 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.

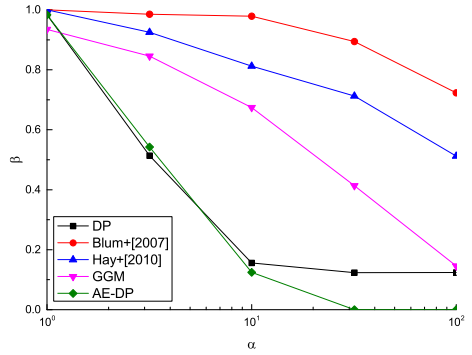
Fig. 5 shows the effect of our solution compared with standard DP and Geo-indistinguishability. For example, in adult dataset, the MSE of standard DP is 199 when $\epsilon = 0.1$, which is close to the MSE, 187 of AE-DP. This case is similar in two dimensional dataset. In Check-in dataset, the MSE of Geo-indistinguishability is 234 when $\epsilon = 0.1$, while that of our mechanism is 198. The same cases can also be found in Fig. 5 (b) and (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.

7.5 (α,β)-accuracy

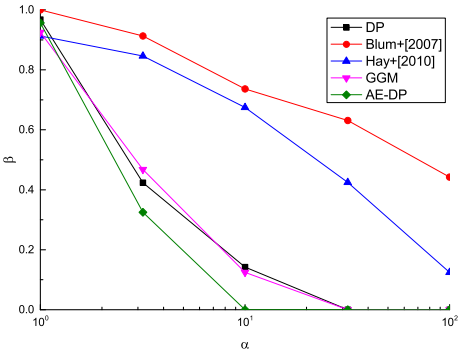
Fig. 6 shows the utility performance, α,β-accuracy, under privacy preserving setting $\epsilon = 1$. We conclude that our solution has much lower β in experimental one and two dimensional dataset, i.e., with higher probability $1 - \beta$ under the same α and ϵ, than the current mechanisms. Therefore, 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. 6, 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.

7.6 Summary for the Experimental Analysis

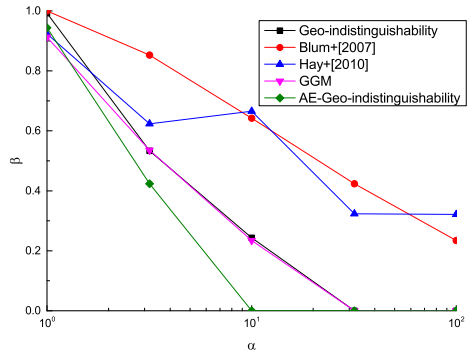
Experimental evaluation in sub-sections 7.2, 3, 4 and 5 supports the following conclusions:



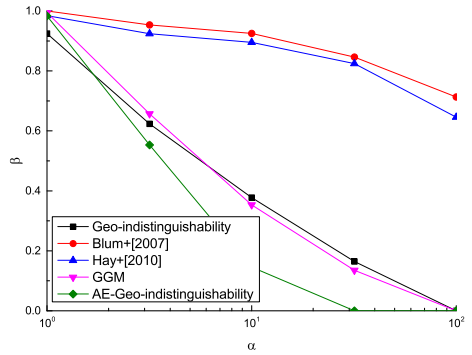
(a) Adult



(b) Social Network



(c) Check-in



(d) Trajectory

Fig. 6: (α,β)-accuracy on different datasets

- 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, CDP can achieve a good trade-off between privacy and utility by selecting a appropriate privacy budget ϵ.

8 Conclusions and future works

DP provides a good trade-off between privacy preserving and data utility. For this reason, an emerging consensus around its application and possible extensions in the academic institution and privacy community is being pursued. However, DP has no constraint on the noise error, 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 noise error to a fixed bound. But the optimal result of data utility is still worth investigated.

In this paper, we first give two definitions of DP with limited noise error in one and two dimensions respectively. Then we propose the corresponding mechanisms to realize the definitions in practice. Furthermore, we utilize particle filter to obtain the optimal sanitized perturbed results. Experimental evaluation demonstrates that our mechanism outperforms current schemes in terms of security and utility for numbers of queries.

Our work in the future is mainly divided into two parts. Future work includes exploring an improved method to guarantee the same effect of this paper with another important relative error. 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 NSFC(42001398), Natural Science Foundation of Chongqing(cstc2020jcyj-msxmX0635), China Postdoctoral Science Foundation funded project(2021M693929), Science and Technology Research Project of CEC(KJQN201900612), Open Fund of LIESMARS(20S02), PhD Starts Fund of CQUPT(A2019-302) and SRTP of CQUPT(A2020-106).

References

1. H Wang and Z Xu. CTS-DP: publishing correlated time-series data via differential privacy. *Knowl.-Based Syst.*, 122:167–179, 2017.
2. Cynthia Dwork. A firm foundation for private data analysis. *Commun. ACM*, 54(1):86–95, January 2011.
3. C. Dwork. Differential privacy. In *Proc. 33rd Int. Conf. Autom., Lang. Programm. (ICALP)*, pages 1–12, 2006.
4. C Dwork. Differential privacy: a survey of results. *Proc. 5th Int. Conf. Theory Appl. Mod. Comp. (TAMC)*, 4978:1–19, 2008.
5. Xiao X, Bender G, Hay M, and Johannes G. ireduct: Differential privacy with reduced relative errors. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 229–240, 2011.
6. M. GÃűtz, A. Machanavajjhala, G. Wang, X. Xiao, and J. Gehrke. Publishing search logs - a comparative study of privacy guarantees. *IEEE Transactions on Knowledge and Data Engineering*, 99:520–532, 2011.
7. A. Korolova, K. Kenthapadi, N. Mishra, and A. Ntoulas. Releasing search queries and clicks privately. In *International World Wide Web Conference (WWW2009)*, pages 172–180, 2009.
8. A. Friedman and A. Schuster. Data mining with differential privacy. In *Proc. of ACM Knowledge Discovery and Data Mining (SIGKDD)*, page 493âĂŞ502, 2010.
9. F. McSherry and I. Mironov. Differentially private recommender systems: Building privacy into the netflix prize contenders. In *Proc. of ACM Knowledge Discovery and Data Mining*, page 627âĂŞ636, 2009.
10. B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *Proc. of ACM Symposium on Principles of Database Systems*, pages 273–282, 2007.
11. S. P. Kasiviswanathan, M. Rudelson, A. Smith, and J. Ullman. 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*, page 775âĂŞ784, 2010.
12. A. Blum, K. Ligett, and A. Roth. A learning theory approach to non-interactive database privacy. In *Proc. of ACM Symposium on Theory of Computing*, page 609âĂŞ618, 2008.
13. Michael Hay, Vibhor Rastogi, Gerome Miklau, and Dan Suciu. Boosting the accuracy of differentially private histograms through consistency. *Proc. VLDB Endow.*, 3(1-2):1021–1032, 2010. 1920970.
14. X. Xiao, G. Wang, and J. Gehrke. Differential privacy via wavelet transforms. In *Proc. of International Conference on Data Engineering*, page 225âĂŞ236, 2010.
15. C. Li, M. Hay, V. Rastogi, G. Miklau, and A. McGregor. Optimizing linear counting queries under differential privacy. In *Proc. of ACM Symposium on Principles of Database Systems*, page 123âĂŞ134, 2010.
16. K. Ligett, S. Neel, A. Roth, B. Waggoner, and S. Wu. Accuracy first: Selecting a differential privacy level for accuracy constrained erm. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*, page 1âĂŞ27, 2017.
17. L. Fang. Generalized gaussian mechanism for differential privacy. *IEEE Transactions on Knowledge and Data Engineering*, 31:747–756, 2019.
18. K. Chatzikokolakis, C. Palamidessi, and M. Stronati. Geo-indistinguishability: A principled approach to location privacy. In *International Conference on Distributed Computing and Internet Technology*, pages 49–72, 2015.
19. Elahe Ghasemi Komishani, Mahdi Abadi, and Fatemeh Deldar. Pptd: Preserving personalized privacy in trajectory data publishing by sensitive attribute generalization and trajectory local suppression. *Knowledge-Based Systems*, 94:43–59, 2016.
20. Y Zheng, X Xie, and WY Ma. Geolife: A collaborative social networking service among user, location and trajectory. *Bulletin Tech. Commi. Data Eng.*, 33(2):32–39, 2010.