World Wide Web (2021) 24:1–23
https://doi.org/10.1007/s11280-020-00825-8

Why current differential privacy schemes are inapplicable for correlated data publishing?

Hao Wang[1] · Zhengquan Xu[2] · Shan Jia[2] · Ying Xia[1] · Xu Zhang[1]

Received: 21 March 2019 / Revised: 19 May 2020 / Accepted: 22 May 2020 /
Published online: 8 June 2020

Abstract

Although data analysis and mining technologies can efficiently provide intelligent and personalized services to us, data owners may not always be willing to share their true data because of privacy concerns. Recently, differential privacy (DP) technology has achieved a good trade-off between data utility and privacy guarantee by publishing noisy outputs. Nonetheless, DP still has a risk of privacy leakage when handling correlated data directly. Current schemes attempt to extend DP to publish correlated data, but are faced with the challenge of violating DP or low-level data utility. In this paper, we try to explore the essential cause of this inapplicability. Specifically, we suppose that this inapplicability is caused by the different correlations between noise and original data. To verify our supposition, we propose the notion of Correlation-Distinguishability Attack (CDA) to separate IID (Independent and Identically Distributed) noise from correlated data. Furthermore, taking time series as an example, we design an optimum filter to realize CDA in practical applications. Experimental results support our supposition and show that, the privacy degree of current approaches has a degradation under CDA.

Keywords Data publishing · Correlated data · Privacy preserving · Differential privacy · Filtering attack

1 Introduction

As a common attribute of data, correlation can reflect the connections among data in real-world applications. Aggregating and mining the correlation attribute is beneficial to governments, businesses and individuals in lots of fields, such as travel routes recommendation [12, 18], road traffic dispatching [10], and environmental protection (e.g., air quality

✉ Zhengquan Xu
xuzq@whu.edu.cn

[1] Chongqing Engineering Research Center for Spatial Big Data Intelligent Technology,
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

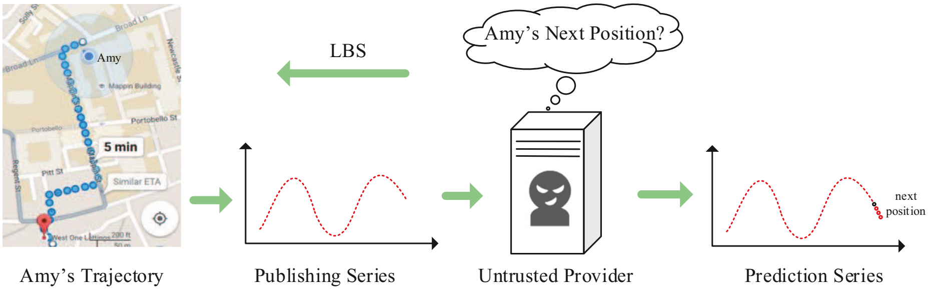
monitoring) [17, 26]. For example, correlated location data uploaded by the user to the service provider can be used to provide better navigation service. Moreover, trajectories aggregated and mined by the third party can support for business hotspots analysis.

As the above examples suggest, correlated data are significantly useful in knowledge discovery and acquisition. Nonetheless, correlated data publishing without special sanitization may violate individual's privacy. For example, by observing a user's historical locations, an adversary can infer the next position that the user wants to go to based on prediction techniques. Due to privacy leakage concerns [13, 14, 22], data owners may not be willing to publish their location data. A privacy disclosure instance of location data release is illustrated in Example 1 and Figure 1.

Example 1 Consider a trajectory publishing scenario, where a trajectory consists of correlated locations, as shown in Figure 1. To obtain better location-based services (LBSs), the user Amy needs to upload her location data sampled at different timestamps to the service provider. After collecting and curating Amy's historical positions, the provider is able to provide high-quality personalized LBSs for her, such as shopping recommendations and route planning. However, by analyzing Amy's historical positions, a malicious provider can predict her next position based on trajectory prediction algorithms, therefore, violating her privacy. It can be seen that publishing one's correlated locations without special sanitization may pose serious threats to individual location privacy.

The problem of private correlated data publishing has attracted attentions from researchers spanning multiple disciplines [4, 15]. In the advanced technologies, random perturbation induces uncertainty (e.g., random noise) about individual values, and the introduction of a small amount of noise can protect user's privacy while has little impact on data utility. Therefore, it has become a widely accepted and practical approach for private data publishing. Among the alternatives, differential privacy [7, 8] is a state-of-the-art standard privacy notion. By introducing IID (Independent and Identically Distributed) Laplace noise, which means that the distributions of the noise are the same and the noise are independent with each other, it provides privacy guarantee that can be mathematically proved. An obvious advantage of differential privacy is that it guarantees strong security regardless of the extent of background knowledge an adversary has of the data.

Because of the intrinsic limitation that standard differential privacy assumes that the data intended to be protected must be independent, it is not suitable for correlated data release. Literature on the study of this issue has explained this inapplicability using Conditional Probability Inference (CPI) [29]. They thought that priori knowledge about the correlation



Amy's Trajectory Publishing Series Untrusted Provider Prediction Series

Figure 1 Privacy disclosure instance of trajectory data release

may increase the success probability of an attack. Based on this deduction, they established specific correlation models, e.g., Markov [9], Bayesian [21, 29, 30] and correlation degree matrix model [24], to describe the correlation of original data. Then they calculated the correlation coefficients according to these models and regarded them as the weight of sensitivity function. The key idea of this kind of methods is that, increasing the noise size can offset the negative effect caused by the priori correlation knowledge. The other mechanism is the transform-based methods, which transformed the correlated data to another independent domain (e.g., Discrete Fourier Transform (DFT) [25] and Discrete Wavelet Transform (DWT) [28]), or extracted a set of independent properties to express the correlation features of the data (e.g., Principal Component Analysis (PCA) [27]). Then the perturbation was added into the transformed data or extracted properties. After inverse transformation, they got the data with perturbed noise.

Despite the improvement in privacy guarantee introduced by these two kinds of correlated data publishing schemes, current solutions are still faced with the following two challenges:

- **Low-level Data Utility.** By increasing noise size, model-based schemes can offset the negative effect caused by the correlation to a certain degree. But bigger noise size added into the outputs means worse data utility. Model-based methods sacrifice data utility to guarantee privacy.
- **Violating DP.** Although transform-based methods do not introduce more noise, the noise distribution after inverse transformation does not obey the Laplace form. Therefore, the noise form does not conform to the requirement of differential privacy.

These challenges imply that existing two kinds of mechanisms have not fundamentally solved the problem of differentially private publication for correlated data release. In this paper, we attempt to explore the essential reason of inapplicability of current methods for correlated data release. In consideration of the correlation, we find that the noise added into the output results in current methods is IID while the output results are correlated. According to the signal processing theory, IID noise can be filtered out from the correlated data. Thus, we suppose that the essential reason why current methods do not suit for correlated data release is that the correlation between noise and output results are different.

Based on this idea, we propose a mechanism, called “correlation distinguishability attack (CDA)”, to verify our supposition. CDA attempts to separate the IID noise from the correlated original outputs, which can validate the effectiveness of state-of-the-art protection schemes. Specifically, we first give the formal definition of CDA. Then based on this definition, we deduce the upper bound of privacy distortion in correlated data release. Furthermore, taking time series as an example, we design an optimum filter to conduct CDA in practice. To the best of our knowledge, it is the first work utilizing a correlation distinguishability attack to explore the privacy distortion of current methods for correlated data release. Our contributions are threefold:

- A notion of correlation distinguishability attack is proposed to explore the essential cause of privacy distortion. CDA attempts to filter out IID noise from the original correlated data utilizing the different correlations of them. Furthermore, an optimum filter is designed to conduct CDA in practice. Taking time series as an example, we verify the correctness of our supposition.
- Since existing methods are based on different principles and there is no general criterion to test the privacy guarantee, performance of them cannot be compared horizontally. As

a general attack model, CDA can be regarded as benchmark to test the performance of different methods.

- Theoretical and experimental analysis demonstrate the correctness of our supposition, which indicates that the reason why current methods can not apply to correlated data publishing is caused by the different correlation between noise and original data. It will lay a theoretical foundation for perfect differential privacy scheme design for correlated data release.

The remainder of this paper is organized as follows. In Section 2, we summarize related work on attack and differentially private publication methods over correlated data, and describe the limitations of existing methods. We then briefly introduce the notations and definitions adopted in this work in Section 3. Our proposal and experiments are described in Sections 4 and 5, respectively, followed by the conclusions and future work in Section 6.

2 Related work

Current methods demonstrate the incompatibility of differential privacy applied to correlated data from the aspect of Conditional Probability Inference (CPI). To overcome this limitation, differential privacy preserving methods for correlated data release have been developed, and can be categorized into model-based and transform-based mechanisms. The model-based methods establish specific models to describe the correlations of data and recalculate the noise according to these models. The transform-based mechanisms transform the correlated data to another independent domain or extract a set of independent properties to express the correlated ones, then IID noise is added in the independent domain or properties. In this section, we describe the attack methods first and then introduce these two kinds of preserving schemes in detail.

2.1 Attack methods

The pioneer study by Kifer et al. [5], confirmed that if correlated records are ignored, the released data will have a lower privacy guarantee than expected. They explained this idea based on examples from social network research as well as tabular data for which deterministic statistics have been previously released. The work in [23] used the example provided in [5] to see how correlation could enhance an adversary's ability in differentiating two neighboring databases. Kargupta et al. [16] developed a random matrix-based spectral filtering method to retrieve original data from perturbed distribution. Nevertheless, they do not focus on the attack method on correlated data publishing. Agrawal et al. [2] utilized Bayesian based method to infer the original data from perturbation. Their method is similar with the current attack on DP based methods. Agrawal et al. [1] discussed an Expectation Maximization (EM) algorithm for distribution reconstruction which is more effective than the currently available method in terms of the level of information loss. Domingo-Ferrer et al. [6] have shown that, for noise addition methods used in practice, it is possible for a user of the masked data to estimate the distribution of the original data. Literature [29] assumed the tuples are probabilistically correlated, and illustrated that the privacy leakage may not be bounded for weak adversaries. As an example to show the limitations of differential privacy under correlated data, a Bayesian attack on differentially private mechanisms proposed by Liu et al. [19] using real-world location datasets leverages the correlation between location information and social information of users. As the state-of-the-art study on DP's limitations for correlated data release, literature [31] also gave an example to consider the impact of tuple correlations on DP.

Current methods try to demonstrate DP's limitations for correlated data release by giving specific examples taking advantage of CPI. However, these methods do not provide a solid theoretical foundation and can not clarify the essential cause of this issue. Thus, a method is needed to give the essential reason of privacy leakage and its upper bound in theory.

2.2 Model-based mechanisms

In the model-based methods, Cao et al. [3] proposed a correlated Hidden Markov detection model to deal with the problem that abnormal data may raise the global sensitivity. They detected and removed the abnormal data by applying the one-step transition probability, which can decrease the noise level added to the original data. However, this model assumed that the releasing probability of the current data is only relevant to its former data, leading to the decline of the detecting results. To increase the accuracy of the detecting results, Yang et al. [29] proposed a privacy definition called Bayesian differential privacy. They constructed a Gaussian correlation model, which assumed that the data to be released conform to the Gaussian distribution. Except for these probability models, Zhu et al. [24] built a correlated degree matrix to measure the whole relationship between records. The coefficients of the correlated degree matrix were used as weights to rebuild the sensitivity function, in place of the traditional global sensitivity. Therefore, the correlated sensitivity can be used to decrease the redundant noise introduced by the global sensitivity.

They can preserve the privacy of correlated data under their assumption. The idea of these model-based methods is to increase the noise size to offset the privacy leakage caused by the correlation, but the behavior of increasing noise size will destroy the data utility.

2.3 Transform-based mechanisms

In the transform-based methods, a typical approach is to transform correlated data into independent series in another domain, thus the correlated data can be processed independently. For example, Rastogi et al. [25] transformed correlated data into independent series in another domain by applying DFT, and then the noise was added to the Fourier coefficients. Thus, perturbed data can be obtained by applying the inverse DFT transform. However, DFT is just a global transformation, which can not describe the local features of the original data accurately. As an improved algorithm, Xiao et al. [28] expanded the range of applications by applying DWT, which can preserve more features of the data in comparison with DFT. In dealing with high dimensionality data, Jiang et al. [27] extracted the features of the data using the properties of PCA, and then these correlated features were classified into several groups of independent features by applying Singular Value Decomposition (SVD).

Compared with the model-based methods, the transform-based methods can ensure a high data utility. However, the noise after inverse transform does not confirm to Laplace distribution, which will lead to a risk of violating DP.

2.4 Summary

In terms of extending differential privacy technology for correlated data publishing, existing schemes illustrate the privacy distortion problem using intuitive examples based on CPI. Unfortunately, they do not give the essential cause of this distortion. Intuitively, current methods try to solve this problem by establishing correlation models or transforming correlated data to independent ones. However, model-based methods are faced with the problem of excessive noise while transform-based methods can not preserve differential privacy. Therefore, in this paper, we aim to address the following issues:

- What is the essential reason of the privacy distortion caused by correlated data publishing using standard differential privacy technology?
- How much is the upper bound of this privacy distortion in theory?
- Are the current schemes effective for correlated data publishing and how much is the privacy distortion if they do not achieve the privacy level as they claim?

3 Preliminaries

In this section, we first define the notations associated with our work and then we review the theory of differential privacy. Next, we demonstrate the problem of privacy distortion using specific example taking advantage of conditional probability inference. Finally, we explain the principle of current methods and indicate the shortcomings of them.

3.1 Notations

As differential privacy aims to guarantee the highest level of privacy, and it assumes that even when an attacker can obtain the entire background information, differential privacy should still preserve the privacy for the target individual. Thus, we first give the definition of correlated and sensitive dataset to be protected, as formalized in Definitions 1 and 2.

Definition 1 (Correlated Dataset) A correlated dataset $D = \{D_1, \cdots, D_i, \cdots, D_n\}$ means that the value change of a variable D_i has an effect on the value change of another variable D_j, where $D_i, D_j \in D$. In this case, we will say D_i and D_j are correlated.

Usually, different data types have different correlation representations. For example, time series usually uses auto-correlation function to represent its correlation, while tuple data use correlation coefficients. For the sake of clarity, in this paper, we take time series as an example and analyze the correlation and attack method of time series.

Definition 2 (Sensitive Dataset) A correlated dataset $D = \{D_1, \cdots, D_i, \cdots, D_n\}$ that the user wants to protect is defined as a sensitive dataset. Suppose two arbitrary records D_i and D_j are correlated in dataset D, whose correlation relationship is expressed by notation δ_{ij}. Then a sensitive dataset D is a set of data with correlation matrix:

$$\Delta = \begin{pmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1n} \\ \delta_{21} & \delta_{21} & \cdots & \delta_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \delta_{n1} & \delta_{n2} & \cdots & \delta_{nn} \end{pmatrix},$$

where $\delta \in \Delta$.

Note that the ways of computing correlation matrix Δ are different in various data types. In time interval analysis, it can be expressed by an auto-correlation function while in attribute analysis, Pearson correlation coefficient is an efficient way to express correlated records.

Differential privacy preserves user's privacy by defining a neighboring dataset which differs the record that the user wants to protect. In Definition 3, we formalize the neighboring dataset that lacks this record.

Definition 3 (Neighboring Dataset) If a sensitive data D_i is removed from the dataset D, denoted as D_{-i}. Then we called the dataset D_{-i} as the neighboring dataset of D.

The above definitions give the formal definitions of correlated data and the dataset consisting of sensitive data associated with differential privacy technology. Next, we demonstrate how differential privacy protects sensitive data using the notion of neighboring dataset.

3.2 Differential privacy

Differential privacy is a currently recognized preservation model that can guarantee stricter security. It is essentially a kind of noise perturbed mechanism. By adding noise to the raw data or statistical results, differential privacy can guarantee that the value changing of a single record has a minimal effect on the statistical output results. Thus, differential privacy can not only preserve the privacy of sensitive data, but also support data mining technologies on statistical results. Its formal definition is shown in Definition 4.

Definition 4 (ε-Differential Privacy [7]) We give the dataset D and its neighboring dataset D_{-i}, which have the same cardinality but differ in only one record. A random perturbation mechanism, M, ensures ε-differential privacy if M makes every set of outcomes, S, for any pair of D and D_{-i} satisfy:

$$Pr[M(D) \in S] \leq exp(\varepsilon) \times Pr[M(D_{-i}) \in S], \tag{1}$$

where $S \subseteq Range(M)$ and $Range(M)$ is the value range of M. $Pr[\cdot]$ and ε denote probability distribution and privacy budget parameter, respectively. A smaller ε means better privacy. Figure 2 depicts the output probability distribution of randomized algorithm M satisfying ε-differential privacy on D and D_{-i}.

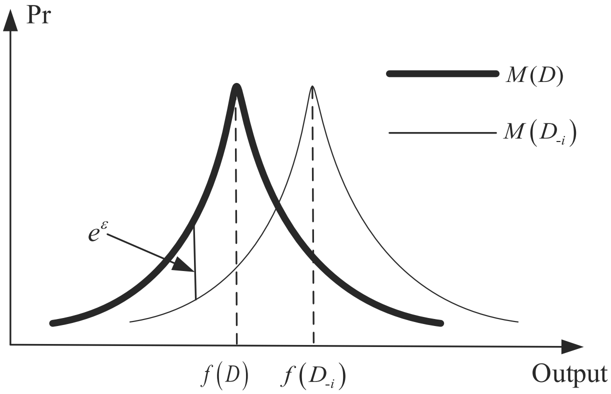


Figure 2 Output probability density of random algorithm M on D and D_{-i}

Table 1 Joint Distribution (Independent)

	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	0.1	0.15
$x_2 = 1$	0.01	0.49

In practical applications, M is generally realized by a Laplace mechanism. By adding IID noise conformed to Laplace distribution into the correlated data series, Laplace mechanism achieves the privacy requirement of differential privacy. Its formal definiton is shown in Definition 5.

Definition 5 (Laplace Mechanism [8]) Assuming that $f(\cdot)$ is the statistical output function, then a noise sequence $Y \sim Lap(\lambda)$, which obeys Laplace distribution, can make randomized algorithm $M(D) = f(D) + Y$ satisfy ε-differential privacy. λ is the scale parameter of the Laplace distribution, and the PDF of the Laplace distribution is

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

The scale parameter λ is determined by sensitivity function Δf and privacy preserving intensity ε:

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

where Δf is the maximum effect of the statistical output function that a single record has on:

$$\Delta f = \max_{D, D_{-i}} \|f(D) - f(D_{-i})\|_1 \tag{4}$$

As an example, consider a dataset whose sensitivity of a query is 1. According to the differential privacy, the noise added to the true answer, which is distributed according to $Lap(1/\varepsilon)$, suffices to guarantee ε-differential privacy.

3.3 Conditional probability inference

Current methods explain the issue of privacy distortion utilizing CPI and specific examples. Here we give an example of CPI to demonstrate the principle of their idea.

Suppose the dataset, $D = \{x_1, x_2\}$, ($x_1, x_2 \in 0, 1$), with the probability shown in Table 1. Consider a query function, $f(x_1, x_2) = x_1 + x_2$. The Laplacian mechanism, M, adds the Laplacian noise, $Lap(1/\varepsilon)$, to the query result. Since x_1 and x_2 are independent, we have $Pr(x_2|x_1) \equiv Pr(x_2)$. For M satisfying ε-DP, we have $e^{-\varepsilon} \leq \frac{Pr(s|x_1=0)}{Pr(s|x_1=1)} \leq e^{\varepsilon}$, $s \in S$, as shown in Figure 3a.

If x_1 and x_2 are correlated, for instance, with the probabilities in Table 2, according to the Bayes' theorem, $Pr(s|x_1) = \sum_{x_2} Pr(r|x_1, x_2) \cdot Pr(x_2|x_1)$. Since $x_1 = x_2$ with high

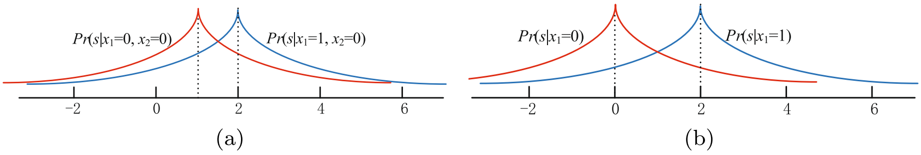


Figure 3 Conditional probability distribution. **a** Independent data. **b** Correlated data

Table 2 Joint Distribution (Correlated)

	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	0.49	0.01
$x_2 = 1$	0.01	0.49

probability, $Pr(s|x_1 = 0) \approx Pr(s|x_1 = 0, x_2 = 0)$ and $Pr(s|x_1 = 1) \approx Pr(s|x_1 = 1, x_2 = 1)$. As the curves $Pr(s|x_1 = 0)$ and $Pr(s|x_1 = 1)$ separate, the distinguishability is no longer bounded in $[e^{-\varepsilon}, e^{\varepsilon}]$, as shown in Figure 3b, although the mechanism M still satisfies ε-DP.

3.4 Analysis of current methods

The differential privacy theory itself requires that the data to be protected are independent of each other. The existing two types of methods are model-based or transform-based, and the purpose is to meet this requirement. We analyze the existing two types of methods and try to explore the essential reason of inapplicability for correlated data release.

3.4.1 Analysis of model-based mechanisms

Model-based mechanisms first use correlation models (such as Markov chain model [3], Bayesian model [29], correlation coefficient matrix [24], etc.) to describe the correlation between data. Then they regard the correlation coefficient as the weight of the differential privacy sensitivity function to recalculate the noise size. In terms of these methods, deleting a single record D_i has an effect on the sensitivity function:

$$\Delta f_i = \sum_{j=0}^{n} |\delta_{ij}| \left(\left\| f\left(D_j\right) - f\left(D_{-j}\right) \right\|_1 \right) \tag{5}$$

where D_{-j} is the dataset that deleting the jth element from D

$$\Delta f = \max_{i \in n} \left(\Delta f_i\right) \tag{6}$$

Figure 4 is the diagram of model-based methods. Combaning Figure 4 and (6), we can conclude that model-based approaches essentially compensates for the decline of privacy degree by increasing the size of the IID noise, but excessive noise leads to a sharp drop in data availability.

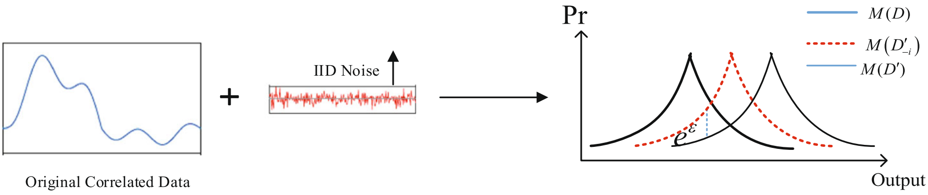


Figure 4 Diagram of Model-based Mechanisms

3.4.2 Analysis of transform-based mechanisms

As shown in Figure 5, the transform-based methods first transforms the correlated query results into independent domains by using transformation methods (e.g., DFT, DWT, etc.):

$$f_T(D) = T(f(D)) \tag{7}$$

where $f_T(D)$ indicates the mutual independent query results after transform and $T(\cdot)$ is the transform function. Then we add IID Laplace noise into the indenpendent domain and obtain:

$$f'_T(D) = T(f(D)) + Y \tag{8}$$

where $f'_T(D)$ is the perturbed results in the independent domain.

Finally, inverse transform is used to generate the perturbed query results:

$$M(D) = T^{-1}(f'_T(D)) \tag{9}$$

where $T^{-1}(\cdot)$ is the inverse transform function.

Although the transform-based methods can guarantee high data availability, the noise does not conform to Laplace distribution after inverse transformation, so it no longer satisfies the definition of differential privacy.

4 Methodology

In this section, we first formalize the problem definition. Then we illustrate the diagrammatic sketch of CDA and give its formal definition. Next, based on our CDA, we derive the relationship between CPI and CDA. Finally, aiming at verifying our supposition, we design an optimum filter to implement CDA in time series.

4.1 Problem definition

Section 3.3 demonstrates that standard differential privacy has a privacy distortion in correlated data release. Our goal is to explore the essential reason of this inapplicability. Next we formalize the problem of privacy distortion and give some necessary definitions associated with the problem.

Definition 6 (Noisy Dataset) Suppose that the data curator wants to know the true values of elements in D, i.e., if $f(\cdot)$ indicates the query function, then $f(D) = D$. The noisy dataset D' can be explained as an original sensitive dataset D plus a corresponding noise series, i.e.,

$$D' := D + Y, \tag{10}$$

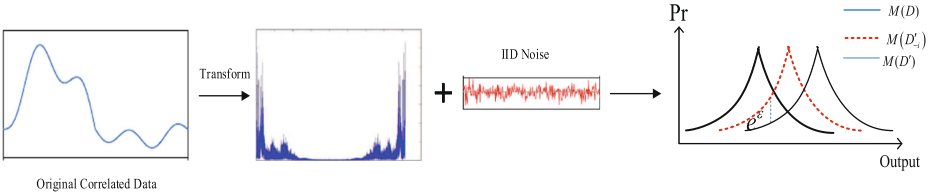


Figure 5 Diagram of Transform-based Mechanisms

In the differential privacy protection technology concerned in this paper, Y obeys the Laplace distribution with a scale parameter λ.

Example in Section 3.3 shows that data correlation can be regarded as an auxiliary information to infer the true data value. Here, we consider the inferring procedure as a posterior estimate processing. It means that the attacker wants to know the true data values on the condition of the perturbed publishing data and correlation. The estimate processing is formalized in Definition 7.

Definition 7 (Posterior Estimate) The dataset after inferring can be regarded as a series that contains the posterior estimate of the sensitive data. The posterior estimate, denoted by $\hat{D}$, can be given by the following conditional estimate:

$$\hat{D} := ES(D|D', \Delta), \tag{11}$$

where D' is the noisy dataset and denotes the set of observations obtained by the data collector, and $ES(\cdot)$ is the estimate function.

Actually, no matter what kind of attack, the essence of CPI and our proposed CDA is a process of estimation, which can be expressed by Definition 7.

4.2 CDA

In this section, we first illustrate the diagrammatic sketch of CDA to demonstrate the principle of our solution. Then we give the formal definition of CDA. Finally, we analyze the relationship of CPI and CDA in theory.

4.2.1 Diagrammatic sketch of CDA

Figure 6 illustrates the diagrammatic sketch of our solution CDA. As shown in Figure 6, to preserve differential privacy of the sensitive dataset D, mechanism M adds an IID noise sequence Y to D and obtains a perturbed dataset D'. ε-differential privacy makes the dif-

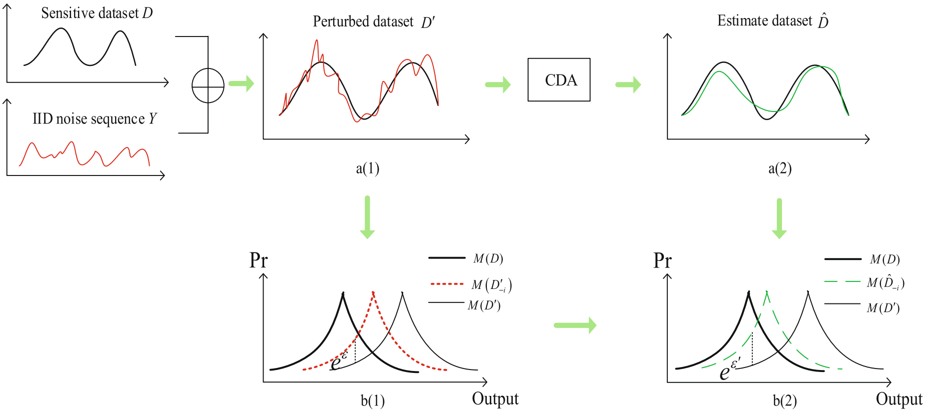


Figure 6 Illustration of CDA for correlated data release

ference of the probability distribution of the outputs between the perturbed dataset D' and its neighboring dataset D'_{-i} be bounded by $[e^{-\varepsilon}, e^{\varepsilon}]$, as shown in Figure 6b(1). However, the original sensitive data in D are correlated but the noise introduced by mechanism M is an IID series. Intuitively, the IID noise series Y can be sanitized from the perturbed sensitive dataset D' by applying CDA (e.g., a filter) and the adversary obtains a estimate dataset $\hat{D}$, as shown in Figure 6a(2). Compared with the probability distribution of $M\left(D'_{-i}\right)$, the probability distribution of $M(\hat{D}_{-i})$ is closer to $M\left(D\right)$, which means that the probability distribution of estimate dataset $\hat{D}$ is closer to that of D, as shown in Figure 6b(2). As a result, the privacy parameter ε' in Figure 6b(2) is larger than ε in Figure 6b(1), indicating a reduction in privacy degree.

4.2.2 Correlation-distinguishability attack

Because of the correlation of data, there is an inherent disadvantage when using IID noise to preserve differential privacy. Although differential privacy has achieved complete privacy within its defined strength, and large quantities of methods claim that they can effectively protect the privacy of correlated data, IID noise can still be sanitized to a certain extent and these schemes still have the potential risk of a privacy distortion as illustrated in Section 4.2.1. Inspired by this observation, we propose a notion called "Correlation-Distinguishability Attack (CDA)". Based on this notion, we also explore the relationship between CPI and CDA. The formal definition of CDA is as follows:

Definition 8 (CDA) Correlation-Distinguishability Attack is defined as a mechanism $\mathcal{F}$ to obtain the posterior estimates of the true data values in D based on the noisy dataset D' and the correlation matrix $\varDelta$:

$$\mathcal{F}(\hat{D}, D) := ES(\hat{D} \rightsquigarrow D|D', \varDelta), \tag{12}$$

where $\rightsquigarrow$ means the process of the posterior estimate.

In practice, a mechanism satisfies CDA can be realized under the least mean square error criterion, which makes the probability distribution of D and $\hat{D}$ minimum, i.e.,

$$\begin{aligned} &\mathcal{F}(\hat{D}, D) \\ s.t. \quad &min\{E[\hat{D} - D]^2\} \end{aligned} \tag{13}$$

4.2.3 CDA VS CPI

From the example of CPI given in Section 3.3, we can see that CPI is indeed a estimation process under the posteriori maximum probability condition. While CDA is also a estimation process under the least mean square error condition. Theorem 1 proof the equivalence of these two attack methods.

Theorem 1 *If the noise added into the original dataset obeys Laplace distribution, then the CPI and CDA is equal to the attacker.*

Proof According to the related knowledge of signal processing, if the following condition can be met, maximum posterior probability estimation and minimum mean square error are equal to the attacker:

If the cost function is a symmetric convex U function, and the posterior probability density function $\rho(D|D')$ is symmetric with the posterior mean value, i.e.,

- $C(\hat{D}) = C(-\hat{D})$, where $C(\cdot)$ is the cost function.
- $C(b\hat{D}_1 + (1+b)\hat{D}_2) \leq bC(\hat{D}_1) + (1-b)C(\hat{D}_2)$ (Convexity), where $0 \leq b \leq 1$.
- $\rho(\varphi|D') = \rho(-\varphi|D')$ (Symmetry), where $\varphi \overset{def}{=} D - \hat{D}_{MMSE} = D - E(D|D')$.

In this paper, we use the square error as the cost function. Obviously, it is easy to meet the first two conditions above. Next, we prove that the third condition is satisfied.

According to the knowledge of estimate theory, Bayes criterion transforms to MMSE as the square error is the cost function. Then the estimate under Bayes criterion is equal to that under MMSE's, i.e.,

$$Risk(\hat{D}_{MMSE}|D') = \int_{-\infty}^{+\infty} (D - \hat{D})^2 f(D|D')dD,$$

where $Risk(\cdot)$ is the average risk function, indicating the expected value of square error.

Obtain the first order partial derivative of $\hat{D}_{MMSE}$, and make the formula to 0, then we have

$$\frac{dRisk(\hat{D}_{MMSE}|D')}{d\hat{D}_{MMSE}} = \int_{-\infty}^{+\infty} 2(\hat{D}_{MMSE} - D)f(D|D')dD = 0.$$

Then

$$\hat{D}_{MMSE} = \frac{\int_{-\infty}^{+\infty} Df(D|D')dD}{\int_{-\infty}^{+\infty} f(D|D')dD}.$$

Since

$$\int_{-\infty}^{+\infty} f(D|D')dD = 1,$$

Then we have

$$\hat{D}_{MMSE} = \int_{-\infty}^{+\infty} Df(D|D')dD = E(D|D'),$$

i.e.,

$$\varphi = D - \hat{D}_{MMSE} = D - E(D|D').$$

In addition, φ obeys the Laplace distribution with mean value $\mu = 0$, then the probability density function is symmetric when $\mu = 0$. Therefore, the symmetry in condition 3 is proved. □

Theorem 1 has proved the equivalence of CPI and CDA. Thus, CDA can be regarded as a benchmark to test the performance of current protect methods. Next, we give the upper bound of privacy distortion under CDA.

Although differential privacy uses a mathematical tool ε to represent the privacy degree, the noise must strictly obey a Laplace distribution. However, this is a rigorous restriction and the noise after sanitizing may not obey this specific distribution. Thus, we need a new privacy metric to measure the retained privacy strength. Since entropy can reflect the uncertainty of the true data values for an adversary regardless of the form of noise [20], in this paper, we use the notion of entropy to quantify the privacy distortion. A privacy distortion based on entropy is intuitively defined as Definition 9.

Definition 9 (Privacy Distortion) Given the perturbed and estimate dataset D' and $\hat{D}$, the privacy distortion $PD(\cdot)$ is given by the following formula:

$$PD(\hat{D}, D') := -\sum_{i=1}^{n} Pr(D'_i - \hat{D}_i) \ln Pr(D'_i - \hat{D}_i), \tag{14}$$

where $Pr(D'_i - \hat{D}_i)$ denotes the probability of the difference value between D'_i and $\hat{D}_i$.

In our work, we define privacy distortion using a general index. Then the attack mechanisms including CPI and CDA can be analyzed compared with a unified index. Theorem 2 gives the upper bound of privacy distortion after CDA.

Theorem 2 *The upper bound of privacy distortion after CDA, $PD_{CDA}(\hat{D}, D')$, is*

$$\max[PD_{CDA}(\hat{D}, D')] = \ln(2\lambda) + 1, \tag{15}$$

where λ is the scale parameter of the noise introduced by Laplace mechanism.

Proof

$$\begin{aligned} PD_{CDA}(\hat{D}, D') : &= PD(\mathcal{F}(\hat{D}, D), D') \\ &= PD(ES(\hat{D} \rightsquigarrow D|D', \Delta), D'). \end{aligned}$$

If the correlation of original data is strong enough or the designed CDA mechanism is optimal, the IID noise Y can be sanitized completely from the perturbed dataset D'. Then we have

$$\begin{aligned} PD(ES(\hat{D} \rightsquigarrow D|D', \Delta), D') &\leq PD(D, D') \\ &= -\sum_{i=1}^{n} Pr(D'_i - D_i) \ln Pr(D'_i - D_i) \\ &= -\sum_{i=1}^{n} Pr(Y_i) \ln Pr(Y_i) \\ &= \sum_{i=1}^{n} [\ln(2\lambda) + \frac{|x|}{\lambda}] \frac{1}{2\lambda} \exp^{-|x|/\lambda} dx \\ &= \ln(2\lambda) + E(Y)/\lambda \\ &= \ln(2\lambda) + 1. \end{aligned}$$ □

Since $\ln(\cdot)$ is a monotone increasing function and $\sigma'^2 < \sigma^2$ always holds, we have $Dis_{\hat{D},D'}(\hat{D}, D') < 1$. In addition, σ'^2 and σ^2 are usually bigger than $\sqrt{\frac{1}{2\pi e}}$, leading to $Dis_{\hat{D},D'}(\hat{D}, D') > 0$. Thus, Theorem 2 demonstrates the existence of privacy distortion after filtering.

4.3 Optimum filtering against time series

Section 4.2 gives the proposed attack model CDA in this paper, and theoretically proves that the existing CPI attack is equivalent to CDA. In this section, taking time series as an example, the optimal filter is designed to implement CDA, and our supposition is verified by experiments, which lays a foundation for correlated data publishing using differential privacy technology.

4.3.1 Principle of filtering attack

As shown in Figure 7, taking time series as an example, we present a filtering attack model for correlated time series release via differential privacy. Since the noise added by DP is small, the correlation of time series before and after filtering does not change much. It is assumed that the correlation of the original time series is public, since the added noise is an IID Laplace sequence, and the associated time series can be regarded as a short-term stationary process.

In Figure 7, $X = \{X(1, t_1), \ldots, X(k, t_k), \ldots, X(n, t_n)\}$ is the original time series contains timestamps. $R_{XX}(\tau)$ is the auto-correlation function of X, with a range of $(-\infty, +\infty)$. τ is the time interval. $R_Y(\tau)$ is the auto-correlation function of Laplace noise Y. Since Y is IID in standard DP, it can be considered as a white noise series. Then the auto-correlation of Y is:

$$R_Y(\tau) = (N_0/2) * \delta(\tau), \tag{16}$$

where $N_0/2$ is the power spectrum of Y, $\delta(\tau)$ is impulse response and $\delta(\tau) = 0, \ \tau \neq 0$. Then $R_Y(\tau)$ has a value only when $\tau \neq 0$. The filter can use the different correlation characteristics of the original data and noise to filter out the noise, leading the increase of probability of successful attack.

4.3.2 Design of optimum filter

From the knowledge of signal processing, the optimum estimate of the stationary time series can be obtained by Wiener filtering under the minimum mean square error criterion. As shown in Figure 7, the perturbed sequence X' is obtained by adding Laplace noise series Y to the original time series X, i.e.,

$$X' = X + Y. \tag{17}$$

If X' passes through a filter with an impulse response $h(\tau)$, the series $\hat{X}$ after filtering is

$$\hat{X}(k) = \sum_{k=-\infty}^{\infty} h(k) x'(j-k), \tag{18}$$

where $x'(j) \in X'$. Thus, the noise series filtered is

$$y'(k) = x'(k) - \sum_{k=-\infty}^{\infty} h(k) x'(j-k). \tag{19}$$

Since the Wiener filter is optimum to filter out IID noise from stationary time series, we use the Wiener filter to conduct the filtering process. The solution process of the impulse response $h(\tau)$ is described below.

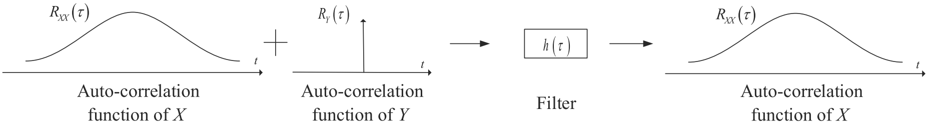

$R_{XX}(\tau)$
t
+
$R_Y(\tau)$
t
$h(\tau)$
$R_{XX}(\tau)$
t
Auto-correlation function of X
Auto-correlation function of Y
Filter
Auto-correlation function of X

Figure 7 Illustration of Filtering Attack

According to the Wiener-Hoff equation, the formula to solve the impulse response of the Wiener filter is

$$P^T = h^T R, \tag{20}$$

where R is the auto-correlation function of X', P is the cross-correlation function of X and X'. Then $h(\tau)$ should be

$$h(\tau) = R^{-1} P. \tag{21}$$

Since Y is a white noise series, its auto-correlation function is

$$R_Y(\tau) = (N_0/2) * \delta(\tau). \tag{22}$$

In addition, the auto-correlation of perturbed series X', R, and the cross-correlation function of X and X', P, can be calculated by

$$\begin{aligned} R &= E\left[x'(k)\, x'^T(k)\right], \\ P &= E\left[x(k)\, x'(k)\right]. \end{aligned} \tag{23}$$

Since the noise added by DP is small, the correlation of the time series before and after filtering does not change much, and the formula of P can be equivalent to

$$P = E\left[x'(k)\, x'(k)\right], \tag{24}$$

Substitute formula (24) and (23) into formula (21), we will obtain the impulse response $h(\tau)$ of Wiener filter.

After the impulse response of the Wiener filter is obtained, a filtering attack can be initiated. The implementation steps of the filtering attack are as shown in Algorithm 1.

Algorithm 1 $\hat{X} = Filter\left(X'\right)$

```
Require: $X'$.
Ensure: $\hat{X}$.
  1. $R \leftarrow E\left[x'(k)\, x'^T(k)\right]$, $P \leftarrow E\left[x'(k)\, x'(k)\right]$;
  2. $h(\tau) = R^{-1} P$;
  3. $X' \overset{h(\tau)}{\rightsquigarrow} \hat{X}$;
  return $\hat{X}$.
```

5 Experimental evaluation

In this section, we evaluate the performance of current schemes under our proposed attack model to verify our supposition. Specifically, we first analyze the influence of correlation on privacy protection strength, and explore whether our solution is effective. Then the resistance performance of current schemes are evaluated on four real-world datasets. Finally, we evaluate the performance of current schemes in terms of data utility.

5.1 Datasets and configuration

The experiments were performed on an Intel Core 2 Quad 3.06-Hz Windows 7 machine equipped with 16 GB main memory. Each experiment was run 1000 times. In order to evaluate the effectiveness of attack model and the existing methods, we select four real-world datasets from four areas, including traffic, medical, network and finance. Datasets are described as follows:

Trajectory : Owing to the Geolife project [32], this dataset contains 17,621 trajectories with a total distance of 1,292,951 km and a total duration of 50,176 h. These trajectory datasets were collected by Microsoft Research Asia from 182 users over five years (from April 2007 to August 2012). A trajectory in this dataset contains the latitude, longitude, altitude coordinates and timestamp.

Netrace [11] : The dataset contains forwarding records of IP layer in a university , which records the timestamps and the number of external devices accessing a device on the intra net. The dataset contains a total of 65,536 records from 1,423 connections.

Flu :[1] Flu is the weekly surveillance data of Influenza-like illness provided by the Influenza Division of the Centers for Disease Control and Prevention4. We collected the weekly outpatient count of the age group [5-24] from 2006 to 2010. This time-series consists of 209 data points.

Unemployment :[2] Unemployment is the monthly unemployment level of African American women of age group [16-19] from ST. Louis Federal Reserve Bank6. This data set contains observations from January 1972 to October 2011 with 478 data points.

In the four datasets, the data in the Traffic dataset have the strongest correlations since the cars can only travel on the road, i.e., the direction and velocity vary slowly. On the contrary, the data in Netrace have the weakest correlations. This paper mainly tests the proposed filtering attack model from four aspects, including the impact of correlation, practical, effective privacy degree and privacy degree before and after filtering.

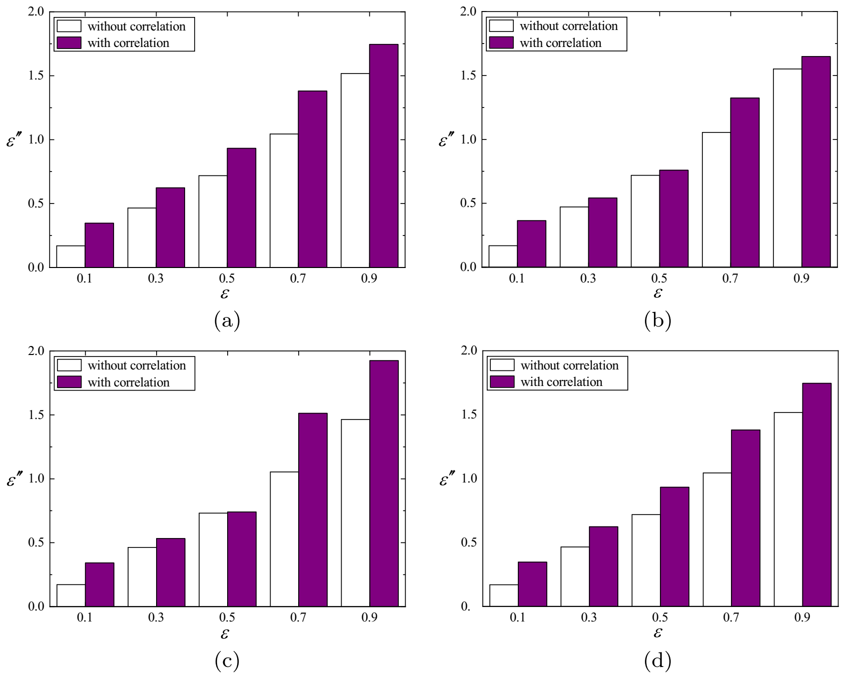
5.2 Impact of correlation

In order to evaluate the impact of correlation knowledge possessed by the attacker on the privacy protection intensity, it is assumed that the attacker has all and no correlation background knowledge. We use the attack model proposed in this paper to attack the four time series protected by ε-differential privacy. According to calculation method of privacy degree, the effective privacy protection strength ε'' before and after attack is obtained respectively. The experimental results are shown in Figure 8.

As shown in Figure 8, on the four datasets, attacks with correlation knowledge filter out more noise than attacks without background knowledge. That is to say, correlation knowledge can be used as auxiliary information for attackers to get original data values more easily. The experimental results in Figure 8 show that, a malicious attack with correlation knowledge on the four time series protected by ε-differential privacy, achieves a lower privacy degree ε'' than that without correlation knowledge. For example, attacking the Trajectory sequence using our attack model, when $\varepsilon = 0.7$, the attacker with correlation knowledge gets a privacy degree $\varepsilon'' = 1.380$, while the privacy degree without background knowledge is 1.044.

[1] http://www.cdc.gov/flu/

[2] http://research.stlouisfed.org/



(a)

(b)

(c)

(d)

Figure 8 Comparison results of impact of correlation on privacy degree. **a** Trajectory. **b** Netrace. **c** Flu. **d** Unemployment

The experimental results show that, the attacker with correlation knowledge using the optimal filter designed by this paper has higher probability of successful attack than that without correlation background knowledge, which proves the effectiveness of the filtering attack method.

5.3 Practical privacy degree

In this section, the four time series and their neighboring series with ε-differential privacy are queried to calculate the practical privacy degree of current methods, which is denoted by ε'.

As shown in Figure 9, the practical privacy preserving strength of each method is different when protecting the same dataset: For Trajectory, when $\varepsilon = 0.5$, the practical privacy degree of the MCMC method is 0.573, while that of DWT is 0.952. The results on the other three datasets have the same trend. For example, for Unemployment, when $\varepsilon = 0.5$, the practical privacy degree of Bayesian is 0.579 while that of CIM is 0.621. At the same time, we observe that the practical privacy degrees on four datasets are different even if we use the same method: when $\varepsilon = 0.1$, for Trajectory, the practical privacy degree ε' of Bayesian is 0.165, while that of Unemployment is 0.135.

In addition, it can be observed that, the practical privacy degrees of MCMC, Bayesian, and CIM are lower than that of DWT and FPA on the same time series. It indicates that the model-based approaches (MCMC, Bayesian, and CIM) have higher privacy degree than the transform-based approaches (DWT and FPA).

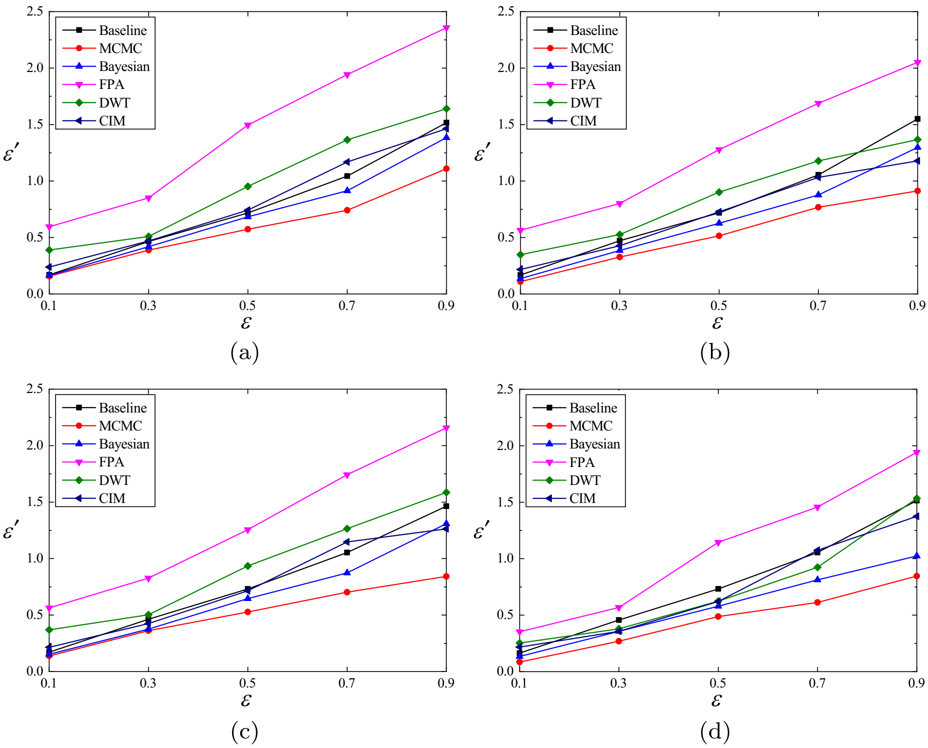


Figure 9 Comparison results of practical privacy degree of current methods. **a** Trajectory. **b** Netrace. **c** Flu. **d** Unemployment

5.4 Effective privacy degree

This section calculates the effective privacy protection strength ε'' of each method under our attack model. Experimental comparison results are shown in Figure 10.

Compared with Figure 9, the privacy degrees of current methods in Figure 10 are all higher than that in Figure 9. For Trajectory, when $\varepsilon = 0.5$, the effective privacy degree ε'' of CIM is 1.647, while the practical privacy degree ε' is 0.742. Similarly, for Unemployment, when $\varepsilon = 0.3$, the privacy degree reduces from 0.964 to 0.357. It can be inferred from the experimental results that, the privacy protection strength of each method under the attack model is reduced, indicating that the filtering attack does filter out part of the noise.

5.5 Privacy degree before and after filtering

In order to make the impact on the privacy degree more intuitive, this section verifies the change of privacy protection intensity of each method before and after filtering attack when $\varepsilon = 0.7$.

The experimental results in Figure 11 are comparisons of the practical and effective privacy degree of current methods under the attack model proposed in this paper. For example, for Trajectory, the practical privacy degree of the CIM method is 1.169, while the effective privacy degree is 2.151. Similarly, for Unemployment, the practical privacy protection

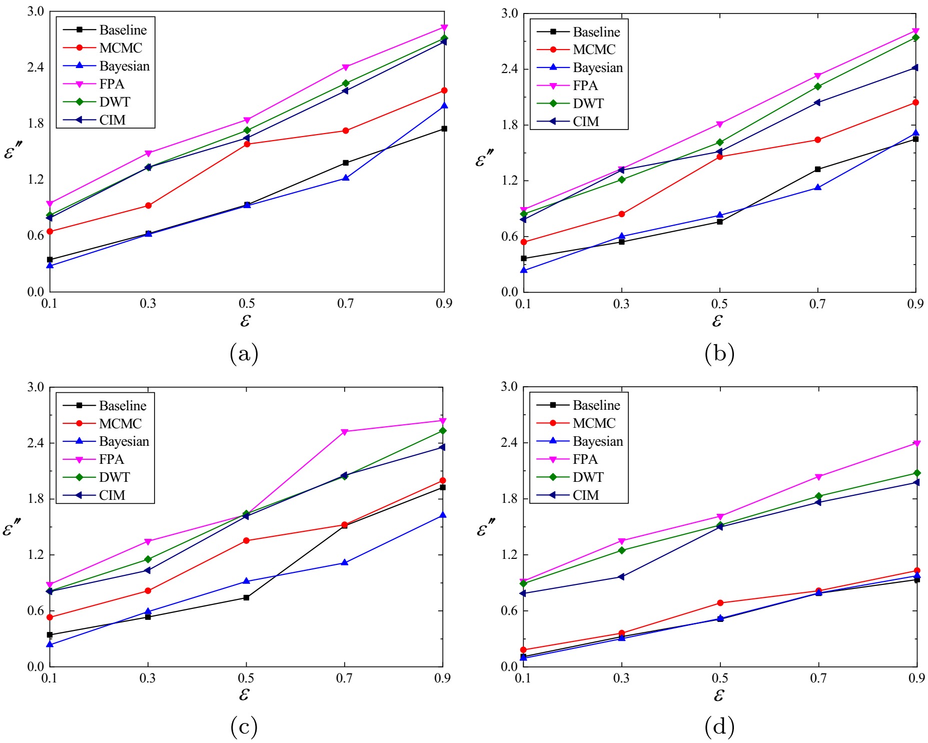


Figure 10 Comparison results of effective privacy degree of current methods. **a** Trajectory. **b** Netrace. **c** Flu. **d** Unemployment

strength of FPA is 1.457, while the effective privacy degree is 2.039. The experimental results show that, under our attack model, less privacy budget is needed.

In summary, the above evaluation demonstrates the following aspects:

- The attacker with correlation knowledge has a higher probability of successful attack than that without correlation knowledge, and the stronger the correlation is, the bigger the privacy distortion will be;
- The effective privacy degrees of existing methods under our attack model are lower than the practical ones, indicating that current approaches do not achieve the desired privacy degree;
- Model-based methods have a less degree of privacy distortion than that of transform-based methods, while transform-based methods have a higher data utility than that of the other.

6 Discussion

We have demonstrate the decline of privacy degree when differential privacy handles correlated data using our proposed CDA model. Furthermore, we have proved the equivalence of CDA and current attack model CPI. The purpose of CDA is to provide a simple and convenient benchmark instead of CPI, which is hard to imply in practice, to test the resistance of

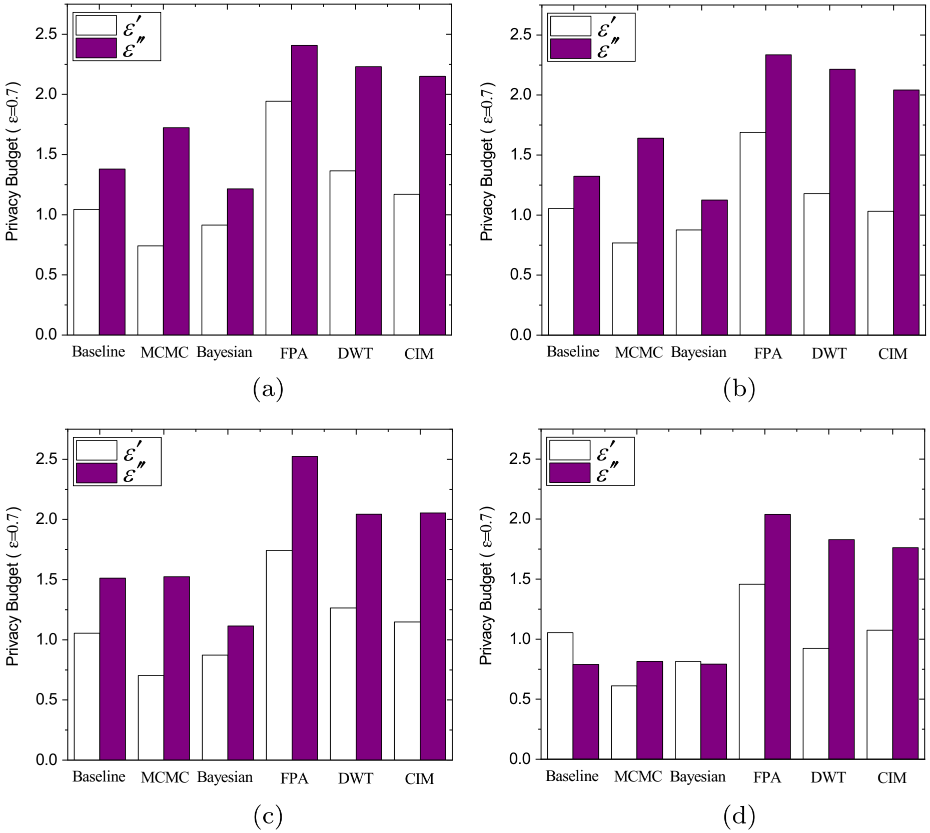


Figure 11 Comparison results of privacy degree of before and after filtering. **a** Trajectory. **b** Netrace. **c** Flu. **d** Unemployment

current protect methods. In fact, considering the opposite of CDA provides an inspiration for us to design differential privacy protection methods for related data release. In our prior work, we have propose a differential privacy release mechanism called "CTS-DP [13]" to address correlated time series. CTS-DP uses auto-correlation function to express the dependence of data and makes the correlation of noise and original data be the same. In addition, except for time series, other correlated data type (e.g., tuple) can also be handled based on the supposition in this paper.

7 Conclusions and future work

In this paper, we explore the essential cause of the inapplicability of current DP schemes for correlated data release. We suppose that the inapplicability is caused by the difference of correlations between noise and data. To verify our supposition, we propose a notion called correlation-distinguishability attack (CDA), which can separate the IID noise used by current schemes from the correlated data. We prove that the privacy distortion after CDA is equal to that of current CPI, and give the upper bound of the privacy distortion. Compared with CPI, CDA is simple and convenient to conduct in practice. Furthermore, taking time series as an example, we design a optimum filter to conduct CDA in practice. As far as we know, this is the first work that attempts to explore the inapplicability of current schemes taking advantage of the different correlations between noise and correlated data. Extensive experiments on real-life datasets support our supposition. CDA can significantly reduce the privacy strength of state-of-the-art approaches, and provide a benchmark for researchers to design more effective privacy preserving methods for correlated data release.

Future work includes investigating more robust differentially private correlated data publishing mechanisms that can resist the attack model proposed in this paper.

Acknowledgments This work was supported in part by the Science and Technology Research Project of Chongqing Education Commission (KJQN201900612) and the PhD Starts Fund Project of Chongqing University of Posts and Telecommunications (A2020-46). The authors are grateful for the anonymous reviewers who made constructive comments and improvements.

References

1. Agrawal, D., Aggarwal, C.: On the design and quantification of privacy preserving data mining algorithms. In: Proceedings of the Twentieth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, SIGACT 2001, Jun 1-4, pp. 247–255, Santa Barbara, California, USA (2001)
2. Agrawal, R., Srikant, R.: Privacy-preserving data mining. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2000, May 11-15, **9**, pp. 439–450, New York, USA (2000)
3. Cao, L., Ou, Y., PS, Y.: Coupled behavior analysis with applications. IEEE Trans. Knowl. Data Eng. **24**(8), 1378–1392 (2011)
4. Chen, R., Fung, B.C.M., Mohammed, N., Desai, B.C., Wang, K.: Privacy-preserving trajectory data publishing by local suppression. Inf. Sci. **231**(1), 83–97 (2013)
5. Daniel, K., Machanavajjhala, A.: No free lunch in data privacy. In: ACM SIGMOD International Conference on Management of Data, SIGMOD 2011, June 12-16, pp. 193–204, Athens, Greece (2011)
6. Domingo-Ferrer, J., Sebe, F., Castella-Roca, J.: On the security of noise addition for privacy in statistical databases. In: Lecture Notes in Computer Science, PSD 2004, **3050**, pp. 149–161 (2004)
7. Dwork, C.: Differential Privacy. In: International Colloquium on Automata, Languages & Programming, ICALP 2006, July 10-14, pp. 1–12, Venice, Italy (2006)
8. Dwork, C.: Differential privacy: a survey of results. In: International Conference on Theory & Applications of Models of Computation, TAMC 2008, April 25-29, **4978**, pp. 1–19. Xi'an, China (2008)
9. Entong, S., Ting, Y.: Mining frequent graph patterns with differential privacy. In: The 19Th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, SIGKDD 2013, Aug 12-16, pp. 545–553, New York, USA (2013)
10. Falvi, G., Pedersen, TB.: Mining long, sharable patterns in trajectories of moving objects. Geoinformatica **13**(1), 27–55 (2009)
11. Fan, L., Xiong, L., Sunderam, V.: Fast:differentially private real-time aggregate monitor with filtering and adaptive sampling. In: ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Jun 6-9, pp. 1065–1068, New York, USA (2013)
12. Hao, W., Kaiju, Li.: Resistance of iid noise in differentially private schemes for trajectory publishing. Compt. J. https://doi.org/10.1093/comjnl/bxz097 (2019)
13. Hao, W., Zhengquan, X.: CTS-DP: Publishing correlated time-series data via differential privacy. Knowl.-Based Syst. **122**, 167–179 (2017)
14. Hao, W., Zhengquan, X.U., Jia, S.: Cluster-indistinguishability: a practical differential privacy mechanism for trajectory clustering. Intel. Data Ana. **21**(6), 1305–1326 (2017)
15. Hao, W., Zhengquan, X.U., Xiong, L., Wang, T.: Conducting correlated laplace mechanism for differential privacy. In: International Conference on Cloud Computing and Security, ICCCS 2017, June 16-18, pp. 72–85, Nanjing, China (2017)

16. Kargupta, H., Datta, S., Wang, Q., Sivakumar, K.: On the privacy preserving properties of random data perturbation techniques. In: Proceedings of the Third IEEE International Conference on Data mining, ICDM 2003, Nov 22-22, pp. 1–8, Melbourne, FL, USA (2003)
17. Kitamoto, A.: Spatio-temporal data mining for typhoon image collection. J Intell. Inf. Syst. **19**(1), 25–41 (2002)
18. Lee, W.H., Tseng, S.S., Tsai, S.H.: A knowledge based real-time travel time prediction system for urban network. Expert. Syst. Appl. **36**(3), 4239–4247 (2009)
19. Liu, C., Chakraborty, S., Mittal, P.: Dependence makes you vulnerable: differential privacy under dependent tuples. In: Proc. 24Th Netw. Distrib. Syst. Security, Symp NDSS 2016, May 6-9, pp. 1–15, Barbara, California, USA (2016)
20. Liu, X.: Entropy, distance measure and similarity measure of fuzzy sets and their relations. Fuzzy Set. Syst. **52**(3), 305–318 (1992)
21. Reza, S.: Privacy games: optimal protection mechanism design for bayesian and differential privacy. arXiv:1402.3426 (2014)
22. Rokach, L., Choo, K.K.R., Bettini, C.: Mobile security and privacy: advances, challenges and future research directions. Pervasive Mob. Comput. **32**, 1–2 (2016)
23. Rui, C., Benjamin, C., Fung, M., Philip, S., Bipin, C.: Correlated network data publication via differential privacy. VLDB J. **23**(4), 653–676 (2014)
24. Tianqing, Z., Ping, X., Gang, L., Wanlei, Z.: Correlated differential privacy: hiding information in non-IID data set. IEEE Trans. Inf. Forens. Security **10**(2), 229–242 (2015)
25. Vibhor, R., Suman, N.: Differentially private aggregation of distributed time-series with transformation and encryption. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2010, June 6-10, pp.735–746, Indianapolis, Indiana, USA (2010)
26. Wang, H., Kaiju, Li.: SRS-LM: differentially private publication for infinite streaming data. J. Amb. Intel. Hum. Comp. **10**(6), 2453–2466 (2019)
27. Wuxuan, J., Cong, X., Zhihua, Z.: Wishart mechanism for differentially private principal components analysis. Comput. Sci. **9285**, 458–473 (2015)
28. Xiao, X., Wang, G., Gehrke, J.: Differential privacy via wavelet transforms. IEEE Trans. Knowl. Data Eng. **23**(8), 1200–1214 (2011)
29. Yang, B., Sato, I., Nakagawa, H.: Bayesian Differential Privacy on Correlated Data. In: The 36Th ACM SIGMOD International Conference on Management of Data, SIGMOD 2015, May 12-16, pp. 747–762. Melbourne Victoria, Australia (2015)
30. Yonghui, X., Li, X.: Dynamic differential privacy for location based applications. arXiv:1410.5919 (2014)
31. Zhao, J., Zhang, J., Poor, H.: Dependent differential privacy for correlated data. In: IEEE GLOBECOM Workshops, GLOBECOM 2017, May 2-6, pp. 1–7, New York, USA (2017)
32. Zheng, Y., Xie, X., WY, M.: Geolife: a collaborative social networking service among user, location and trajectory. Bulletin Tech. Commi. Data Eng. **33**(2), 32–39 (2010)