International Journal of Database Theory and Application
Vol.10, No.1 (2017), pp.89-104
http://dx.doi.org/10.14257/ijdta.2017.10.1.09

An Attack Model on Differential Privacy Preserving Methods for Correlated Time Series

Wenjun Xiong[1,2], Zhengquan Xu[1,2,*] and Hao Wang[1,2]

[1]*State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, 430079, China*
[2] *Collaborative Innovation Center for Geospatial Technology, Wuhan, 430079, China*
**Corresponding author: xuzq@whu.edu.cn*
wendyxiong@whu.edu.cn, haowang354@whu.edu.cn

Abstract

Differential privacy has played a significant role in privacy preserving, and it has performed well in independent series. However, in real-world applications, most data are released in the form of correlated time series. Although a few differential privacy methods have focused on correlated time series, they are not designed by protecting against a specific attack model. Due to this drawback, the effectiveness of these methods cannot be verified and the privacy level of them cannot be measured. To address the problem, this paper presents an attack model based on the principle of filtering in signal processing theory. Since the distribution of the noise designed by current methods is independent and different from that of the original correlated series, a filter is designed as a unified attack model to sanitize the independent noise from the perturbed time series. Furthermore, the designed attack model can realize the function of measuring the effective privacy level of these methods and comparing the performance of them. Experimental results show that the attack model leads to degradation in privacy levels and can work as a unified measurement.

Keywords*: Privacy Preserving, Differential Privacy, Correlated Time Series, Attack Model*

1. Introduction

Time series is a collection of data recorded chronologically and generally correlated. As a significant form of data storage and publishing, time series is prevalent in various fields, such as process monitoring, financial predict and traffic dispatching.

As time series could carry abundant information and it is beneficial to our daily lives, many efforts have been focused on time series data mining. However, individuals' sensitive information may be revealed by the data mining results. Due to privacy considerations, an approach which could release personal data while ensuring the safety of the sensitive information is highly desired. In order to achieve this goal, Dwork proposed a definition referred as differential privacy [1]. As a novel privacy preserving mathematical framework, differential privacy provides a strict mathematical model, which is independent of the attackers' background knowledge. Therefore, it has become a widely accepted method for preserving data privacy.

Differential privacy mechanism is actually a noise perturbation mechanism, which is designed under the assumption that the datasets are independent. Global sensitivity is defined to measure the maximal effect a single record has on a dataset, and therefore the noise level added to the original dataset can be calculated. However, for preserving the privacy of correlated datasets, as the data correlation will raise the global sensitivity, for achieving the expected privacy level, adding noise according to the standard global

ISSN: 2005-4270 IJDTA

sensitivity function will introduce redundant noise and lead to degradation in data utility. Therefore, a differential privacy mechanism for correlated data, that can retain the data utility while preserving the data privacy, is in high demand. As a typical correlated series, time series has drawn considerable research interests. Existing differential privacy preserving methods for correlated time series can be categorized into model-based methods and transform-based methods. The model-based methods rebuilt the sensitivity function by applying correlation models, such as Markov [2] and Bayesian [3] correlation models, or by applying coefficient matrix models [4]. The transform-based methods can be divided into two types. The first type is to transform the correlated time series into independent series of another domain by transformation techniques, thus the series can be processed independently. Examples of such techniques include Discrete Fourier Transform (DFT) [5] and Wavelet Transform (WT) [6, 7]. Another type utilized data feature extraction methods, *e.g.*, Principal Component Analysis (PCA) [8], to extract the correlation properties of time series, and therefore they can be represented by a set of independent properties.

Although improvements of these methods are made, there are still some challenges: 1) the model-based methods and the transform-based methods are designed under the assumption that time series is correlated according to a certain rule. However, much of the research on privacy preserving has focused on building an attack-defense system, thus a specific attack model on the attacker-side is still in high demand; 2) the effectiveness of these methods cannot be verified and the privacy level of them cannot be measured.

To deal with the above challenges, we propose an attack model based on the principle of filtering in signal processing, under which the independent noise can be sanitized from the perturbed time series and attackers can obtain the original time series with a higher probability. The contributions in this paper are as follows:

1) As the distribution of the Laplace noise designed by the current methods is independent, and different from that of the original correlated series, it provides us with an opportunity to design an attack model in the view of signal processing. Based on this, an attack model is proposed in this paper.

2) To filter out the Laplace noise from the perturbed time series, we propose a practical and optimal filter as an attack model. Therefore, the perturbed time series can be sanitized by applying the filter.

3) The proposed attack model can be used as a unified measurement of the privacy level, thus the performance of the current methods can be compared.

The rest of this paper is organized as follows: we discuss related work in Section 2. Section 3 shows the preliminaries of differential privacy and provides the problem statement. We propose an attack model and analyze the effective privacy level under the attack model in Section 4. Section 5 presents the experimental evaluation, followed by conclusions in Section 6.

2. Related Work

Existing differential privacy preserving methods for correlated time series can be categorized into model-based methods and transform-based methods.

In the model-based methods, a major approach is to build a probability model for correlated series data releasing. Cao *et al.* [2] 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. Thus the detecting results were not accurate enough. To increase the accuracy of the detecting results, Yang *et al.* [3] proposed a privacy definition called Bayesian differential privacy, and then they constructed a Gaussian

correlation model, which assumed that the data released currently is related to all the data released before. Except for these probability models, Zhu *et al.* [4] 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. These model-based methods can preserve the privacy for correlated time series to some extent. However, there are still some challenges: 1) the correlations of time series are complicated, thus they cannot be represented by a single model; 2) these methods are proposed under the assumption that time series is correlated according to a certain rule, thus research on the effectiveness of these methods under a specific attack model is highly desired.

In the transform-based methods, a typical approach is to transform time series into independent series of another domain, thus the series can be processed independently. For example, Rastogi *et al.* [5] transformed time series into independent series of another domain by applying DFT, and then the noise was added to the Fourier coefficients. Thus, a perturbed series can be obtained by applying the inverse DFT transform. However, DFT is just a global transformation, which cannot describe the local features of the original time series accurately. As an improved algorithm, Xiao *et al.* [6, 7] expanded the range of applications by applying WT, which can preserve more features of the series in comparison with DFT. In dealing with high dimensionality time series, Jiang *et al.* [8] extracted the features of the correlated time series 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 to the model-based methods, these transform-based methods can ensure a high data utility. Whereas the transform-based methods may lose correlated features in the process of transforming, they cannot ensure the expected privacy level.

In privacy preserving for correlated time series, the existing methods are proposed under the assumption that time series is correlated according to a certain rule. However, due to the lack of a specific attack model, the effectiveness of these methods cannot be verified and the privacy level of these methods cannot be measured. To deal with the problems detailed above, an attack model on differential privacy preserving methods for correlated time series is proposed.

3. Preliminaries

In this section, we first illustrate the background of differential privacy; then we illustrate the problem faced by the existing differential privacy preserving methods for correlated time series under the attack model via filtering.

3.1. Differential Privacy

The main idea of differential privacy is to release aggregate information of dataset *D* without revealing the privacy of individuals. A formal definition of differential privacy is as follows:

Definition 1 (ε-Differential Privacy [9]**).** Denote $K(D)$ the output of a mechanism *K* on input dataset *D*. Then mechanism *K* is ε-differential privacy if for any dataset *D* and its neighbor dataset D' which differs in only one record, and for any output *S*, the following holds:

$$\Pr\left[K(D) \in S\right] \leq \exp(\varepsilon) \times \Pr\left[K(D') \in S\right] \quad (1)$$

where Pr is the output probability distribution of the mechanism *K*.

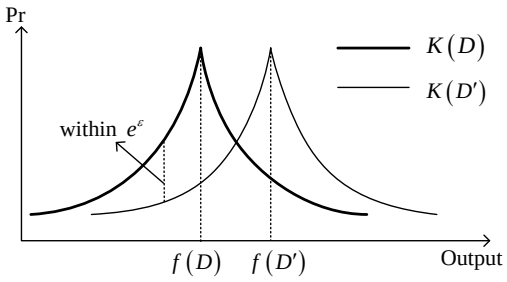

Figure 1. Output Probability Distribution of K on Neighboring Datasets

Figure 1 illustrates that mechanism *K* offers ε-differential privacy to dataset *D*. The privacy parameter ε ensures that the output of dataset *D* and its neighbor dataset D' cannot be distinguished within a certain probability. A lower ε implies a better privacy level.

To ensure ε-differential privacy, Laplace mechanism is applied to add a suitable noise level to the true query answer:

Definition 2 (Laplace Mechanism [10]**).** Let $f: D \to R$ be any query sequence. Denote $f(D)$ the query answer of a query f on input dataset *D*, and $Lap(\lambda)$ a random variable drawn from the Laplace distribution with scale λ. Then

$$K(D) = f(D) + Lap(\lambda) \tag{2}$$

satisfies ε-differential privacy. λ can be obtained through the following equation:

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

where Δf represents the global sensitivity, which measures the maximum change on the result of query f when removing one record from the dataset. The definition is as follows:

Definition 3 (Global Sensitivity [10]**).** For any query sequence $f: D \to R$, the global sensitivity of f is:

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

For many types of queries, Δf is quite small. In particular, the simple counting queries have $\Delta f = 1$.

Since the correlation of time series can raise the global sensitivity, for achieving the expected privacy level, adding noise according to the standard global sensitivity function will introduce redundant noise and lead to a low data utility. In order to achieve a better trade-off between privacy level and data utility, some differential privacy preserving methods for correlated time series improved the global sensitivity function, in order to introduce a smaller noise level. However, the noise added to the original time series is an independent and identically distributed (IID) series, which can be filtered out by applying an attack model. Therefore, these methods will not achieve the expected privacy level.

3.2. Problem Statement

This section illustrates the problem, which is faced by the existing differential privacy preserving methods for correlated time series, under the attack model via filtering.

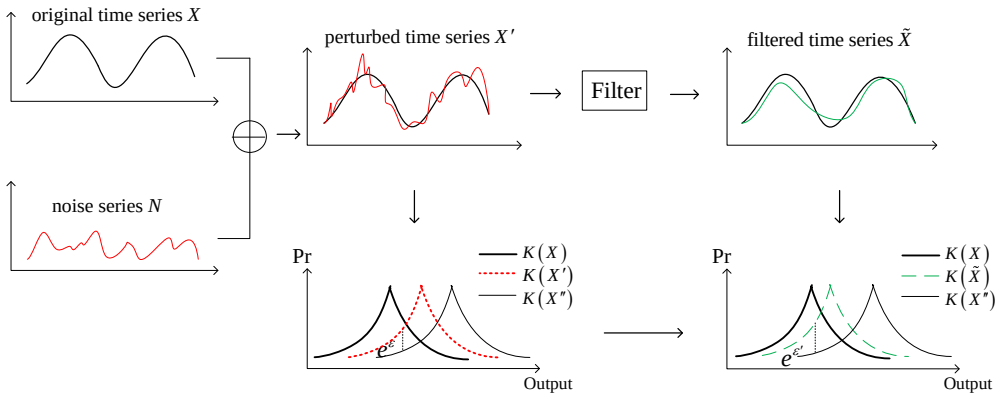

Figure 2. Impact of Correlation on Privacy Level

As is shown in Figure 2, to preserve differential privacy of the original time series X, mechanism K add an IID noise series N to X, and therefore obtain a perturbed time series X'. We query X, as well as the neighboring series X' and X''. The probability distribution function (PDF) of the query answers are $K(X)$, $K(X')$ and $K(X'')$, respectively. However, the original time series X is correlated and the noise introduced by mechanism K is an IID series. Thus, the IID noise series can be sanitized from the perturbed time series by applying a filter. Therefore, attackers can obtain the filtered time series $\tilde{X}$. Compared to $K(X')$, $K(\tilde{X})$ is closer to $K(X)$, which means that the filtered time series $\tilde{X}$ is closer to the original time series X. As a result, the privacy parameter ε increases to ε', which means the privacy level decreases, and attackers can obtain the original time series with a higher probability.

To address the correlation problem, a specific attack model on differential privacy preserving methods for correlated time series is proposed in the next section.

4. Attack Model

In this section, we first provide the illustration of the filtering attack; then we calculate the impulse response of the filter; finally, the effective privacy level of the standard differential privacy mechanism is measured under the proposed attack model.

4.1. Illustration of Filtering Attack

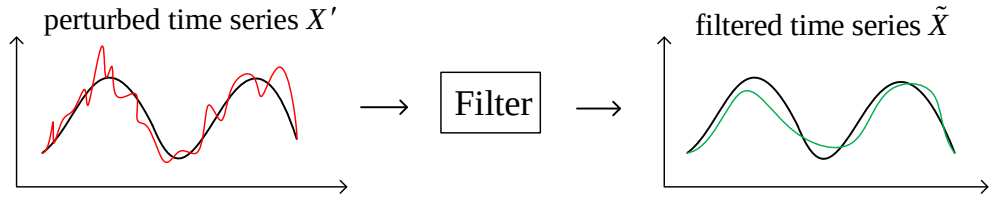

Figure 3. Illustration of Filtering Attack

Figure 3 shows a specific attack model, which is a practical filter designed in the view of signal processing. Since the noise introduced by the Laplace mechanism is not large, the correlation of the time series changes little under the attack model. Assume that we already know the correlation of the original time series, and the noise introduced by the Laplace mechanism is an IID series, then the noise can be filtered out by applying the practical and optimal filter. Therefore attackers can obtain the original time series with a higher probability.

Algorithm 1 $\tilde{X} = Filter(X')$

Input:
Original time series *X*, perturbed time series X'
Output:
Filtered time series $\tilde{X}$

1: Calculate the auto-correlation function *R* of X', as well as the cross-correlation function *P* of *X* and X'
2: Design a proper impulse response $h(k)$ of the filter according to *R* and *P*, in order to filter out the noise from X' as much as possible
3: Obtain $\tilde{X}$ by applying the optimal filter to filter out the IID noise series from X'
4: Return $\tilde{X}$

Algorithm 1 illustrates the working processes of the attack model, the most important part of which is the calculation of the impulse response function $h(k)$. Its calculation process will be given in the next section.

4.2. Impulse Response Calculation

We already know that correlated time series can be seen as short-time stationary processes, and the noise introduced by the Laplace mechanism is an IID series. Since the Wiener filter is usually applied to filter out the IID series from stationary processes, we take the classic Wiener filter as an example to illustrate the calculation of the impulse response function.

The perturbed time series X' is obtained by adding noise series *N* to the original time series *X* according to the Laplace mechanism:

$$X' = X + N \tag{5}$$

The Wiener filter takes X' as the input series. Assume that the impulse response of the filter is $h(k)$, we obtain the output series $\tilde{X}$ based on the principle of filtering in signal processing:

$$\tilde{x}(k) = \sum_{k=-\infty}^{\infty} h(k)x'(j-k) \tag{6}$$

where $x'(j) \in X'$.

According to the Wiener-Hopf equation, we obtain the impulse response of the Wiener filter from:

$$P^T = h^T R \tag{7}$$

where *R* is the auto-correlation function of X', and *P* is the cross-correlation function of *X* and X'. Accordingly, the impulse response function $h(k)$ of the Wiener filter is:

$$h(k) = R^{-1}P \tag{8}$$

Since the noise introduced by the Laplace mechanism is a white noise series, then the auto-correlation function of *N* is:

$$R_n = \delta(k) \tag{9}$$

Accordingly, the auto-correlation function *R* of X', the cross-correlation function *P* of *X* and X' are:

$$R = E\left[x'(k)x'^T(k)\right] \tag{10}$$

$$P = E\left[x(k)x'(k)\right] \tag{11}$$

According to Equation 8, we obtain the impulse response function $h(k)$ of the filter.

4.3. Privacy Level Evaluation

This section evaluates the effective privacy level of the standard differential privacy under the attack model proposed by this paper.

$$\frac{\Pr\left[K(X)\in S\right]}{\Pr\left[K(X')\in S\right]}=\frac{\Pr\left[K(N)\in S-X\right]}{\Pr\left[K(N')\in S-X'\right]} \tag{12}$$

According to Equation 12, the effective privacy level can be calculated by analyzing the output series of the filter when we input the Laplace noise series *N*. The output series of the filter is analyzed in the following theorem:

Theorem 1. A noise series *N* with *m* points introduced by the Laplace mechanism passes through the filter, and the impulse response function of which is $h(k)$, then the output series $\tilde{N}$ is approximate subject to Gaussian distribution with a variance of $\frac{2m\lambda^2}{h^2(k)}$, *i.e.*, $\tilde{N}\sim N\left(0,\frac{2m\lambda^2}{h^2(k)}\right)$, where λ is the magnitude of the noise.

Proof. According to the principle of filtering in signal processing, if the noise series *N* passes through a linear system, and the impulse response function of the system is $h(k)$, then the output series is

$$\tilde{n}(k)=\sum_{k=-\infty}^{\infty}h(k)n(j-k) \tag{13}$$

where $n(j)\in N$. According to Equation 13, the impulse response function $h(k)$ can be seen as the weight coefficient of $n(k)$, thus $\tilde{n}(k)$ is the weighted linear combination of $n(k)$. According to the features of the Laplace probability distribution, $\tilde{n}(k)$ is an IID Laplace series, the weight coefficient of which is $\tilde{\lambda}=\frac{\lambda}{h(k)}$.

According to the central-limit theorem (CLT), if a series *N* of random variables is an IID series with mean □ and variance σ^2, then the sum of the first *m* terms of *N* is approximate subject to Gaussian distribution with mean m□ and variance $m\sigma^2$, *i.e.*,

$$\sum_{k=1}^{m}n(k)\sim N\left(m\mu,m\sigma^2\right) \tag{14}$$

As the filter consists of many adders, when the Laplace noise series *N* with variance $D[N]=2\lambda^2$ passes through the filter, the output series $\tilde{N}$ is approximate subject to Gaussian distribution, and the variance of the output series $\tilde{N}$ is

$$D\left[\tilde{N}\right]=2m\tilde{\lambda}^2=\frac{2m\lambda^2}{h^2(k)} \tag{15}$$

Since the Laplace noise introduced by the differential privacy preserving methods has mean 0, *i.e.*, □ =0, we obtain

$$\tilde{N}\sim N\left(0,\frac{2m\lambda^2}{h^2(k)}\right) \tag{16}$$

Accordingly, when the noise series introduced by the Laplace mechanism passes through the Wiener filter, the output series is approximate subject to Gaussian distribution. Thus we can obtain the mean and variance of the output series. The effective privacy level under the attack model is illustrated in the following theorem:

Theorem 2. A correlated time series X' with m points perturbed by the Laplace mechanism passes through the attack model, then the effective privacy level is $\varepsilon' = \frac{\left(R^{-1}P\right)^2}{2m}\varepsilon^2$, where $R = E\left[x'(k)x'^T(k)\right]$, and $P = E\left[x(k)x'(k)\right]$.

Proof. According to a technical report [11], Gaussian noise can provide δ - approximate ε -differential privacy. Specifically when $\varepsilon > \left[\log(\frac{1}{\delta})/\sigma^2\right]^{\frac{1}{2}}$, Gaussian noise can provide $\frac{1}{\sigma^2}$ - indistinguishable, where σ^2 is the variance. Since $\varepsilon > 0$ and the value of δ is quite small, the inequality can be established under general conditions.

Take counting queries as an example, it has $\Delta f = 1$. Combined with Equation 3, the effective privacy level under the attack model is:

$$\varepsilon' = \frac{1}{\sigma^2} = \frac{h^2(k)}{2m\lambda^2} = \frac{h^2(k)}{2m}\varepsilon^2 \tag{17}$$

According to Equation 8, we obtain:

$$\varepsilon' = \frac{\left(R^{-1}P\right)^2}{2m}\varepsilon^2 \tag{18}$$

where $R = E\left[x'(k)x'^T(k)\right]$, and $P = E\left[x(k)x'(k)\right]$.

5. Experiments and Evaluation

In this section we first introduce datasets and configuration; then we evaluate the impact of the correlation on the privacy level; finally, we measure the performance of the state-of-the-art differential privacy preserving methods for correlated time series by evaluating the effective privacy level and data utility of them.

5.1. Datasets and Configuration

The experiments are running on an Intel Core 2 Quad 2.93 GHz Windows 7 machine equipped with 4 GB memory. Each experiment runs 1, 000 times. The experiments involve four time series datasets, including the fields of transportation, medical, network and economic.

(a) Trajectory [12]: This GPS trajectory dataset contains 17, 621 trajectories, which is represented by a sequence of time-stamped points. Each point contains the information of latitude, longitude, height, speed and heading direction.

(b) Diabetes [13]: This dataset contains the information of outpatient care on 70 patients, each record of which represents the physical condition of these patients, including date, time, code and value.

(c) NetTrace [14]: This dataset contains the IP-level network trace sampled from a border gateway of a university. There are 65, 536 records in total with the connection number ranging from 1 to 1, 423. Each record contains the number and time of external hosts connected to an internal host.

(d) Amazon Access Samples [13]: This dataset contains the assigned access of users, each file of which contains four categories, including person, resource, group and system-support.

Among the four datasets, Trajectory has the strongest correlation and Amazon Access Samples has the weakest correlation. We generate a query set F with 1, 000 random linear queries. The number of queries is represented by $|F|$. On Trajectory: the query returns the number of points whose attribute value is greater than a fixed value. On

Diabetes: the query returns the mean value of each indicator. On NetTrace: the query returns the number of connected internal and external hosts. On Amazon Access Samples: the query returns the number of possibly supported users. The probability of each query answer fell into [0, 1].

5.2. Experimental Methods

According to Dwork [15], $\varepsilon \le 1$ is suitable for privacy preserving purposes. Therefore, we conduct the experiments with a fixed privacy parameter ε varied from 0.1 to 0.9 with a 0.2 step on four datasets.

We calculate the PDF of the queries on four original datasets and their neighboring datasets with the fixed privacy parameters. The practical privacy level ε' and the effective privacy level ε'' can be calculated according to the following equation:

$$\frac{\Pr\left[K(X) \in S\right]}{\Pr\left[K(X') \in S\right]} \le \exp(\varepsilon) \tag{19}$$

A lower privacy parameter value implies a higher privacy level.

To measure the data utility of the current methods under the attack model, we calculate the PDF of the queries on four datasets and their neighboring datasets. The accuracy of results can be measured by Mean Square Error (MSE):

$$\text{MSE} = \frac{1}{|F|}\sum_{F_i \in F}\left(\tilde{F}_i(X) - F_i(X)\right)^2 \tag{20}$$

A lower MSE implies a better data utility.

5.3. Evaluation of Privacy Level

We first evaluate the impact of correlation on the privacy level; then we evaluate the practical privacy level of the state-of-the-art differential privacy preserving methods for correlated time series, as well as the effective privacy level of them under the attack model.

5.3.1. Impact of Correlation

We evaluate the impact of correlation by calculating the practical privacy level under various privacy parameters on four datasets, which are protected by the standard differential privacy mechanism.

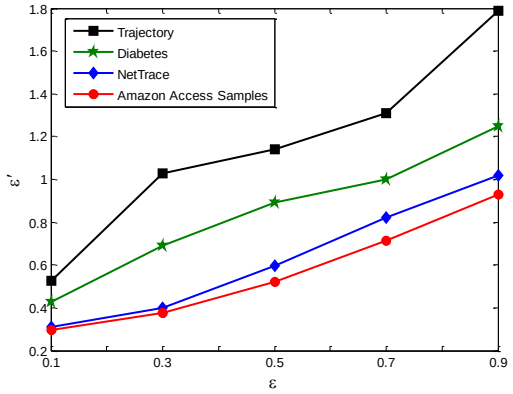

Figure 4. Impact of Correlation on Privacy Level

Figure 4 shows that, for Trajectory, ε' is higher than other datasets at each privacy parameter ε. Specifically, when $\varepsilon = 0.9$, the ε' is 1.7950, while for Diabetes the ε' is 1.2470, for NetTrace is 1.0240 and for Amazon Access Samples is 0.9320.

The experimental result shows that, the stronger correlation of a dataset, the less privacy level it achieves.

5.3.2. Calculation of Practical Privacy Level

We calculate the practical privacy level of the current methods on four datasets, and the practical privacy level can be measured by ε'.

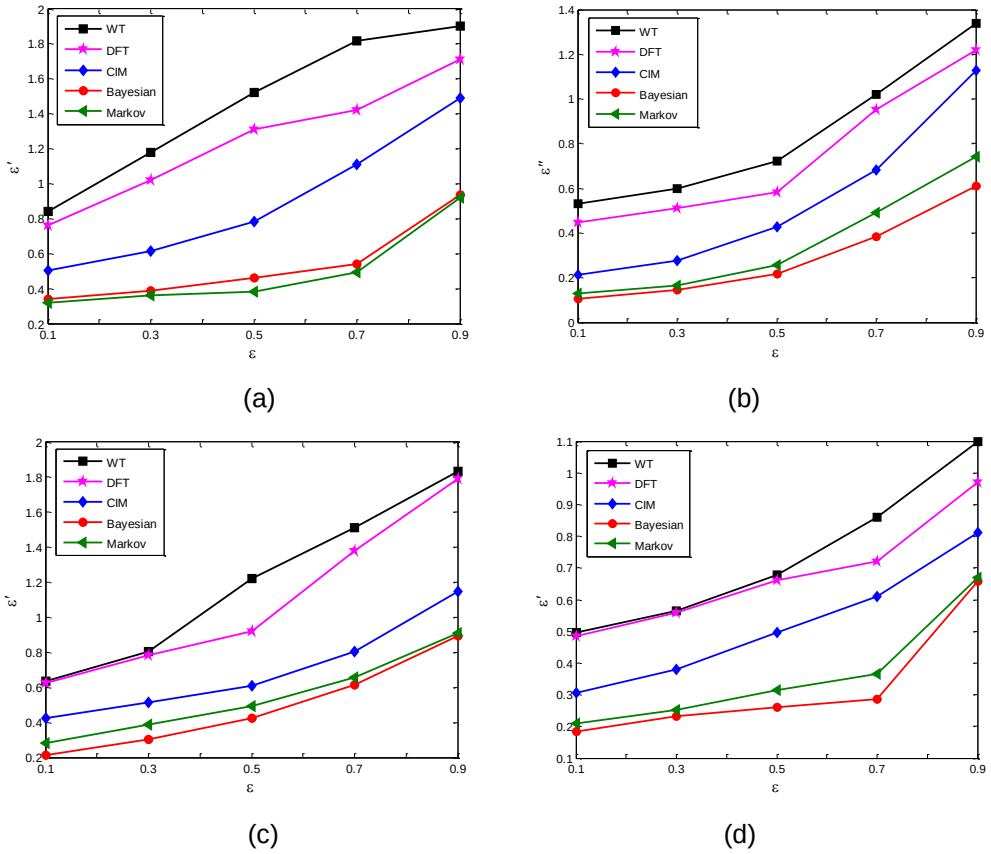

Figure 5. Comparison of Practical Privacy Level (a) Trajectory (b) Diabetes (c) NetTrace (d) Amazon Access Samples

From Figure 5 we observe that each method performs differently when protecting the same dataset. Specifically, for Trajectory, when $\varepsilon = 0.1$, Markov achieves a ε' at 0.3210, while WT achieves 0.8420. Other datasets have similar trends. For example, for Diabetes, when $\varepsilon = 0.5$, Bayesian achieves a ε' at 0.2155, while CIM achieves 0.4296. We also observe that the same method performs differently when protecting different datasets. Specifically, when $\varepsilon = 0.1$, Bayesian achieves a ε' at 0.3422 for Trajectory, while Bayesian achieves a ε' at 0.1842 for Amazon Access Samples.

Moreover, we observe that the practical privacy level of Markov, Bayesian and CIM are lower than that of WT and DFT, which means the model-based methods (Markov, Bayesian and CIM) perform better in privacy preserving than the transform-based methods (WT and DFT).

5.3.3. Calculation of Effective Privacy Level

We calculate the effective privacy level of the current methods on four datasets under the attack model, and the effective privacy level can be measured by ε''.

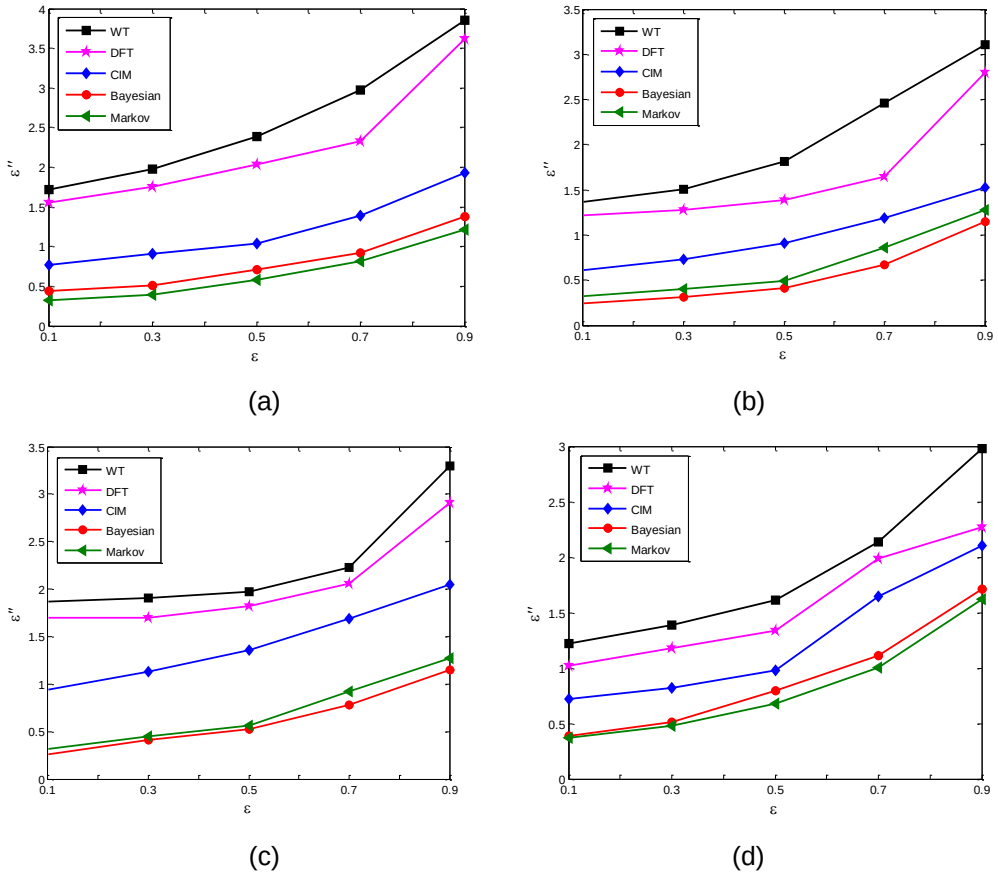

Figure 6. Comparison of Effective Privacy Level (a) Trajectory (b) Diabetes (c) NetTrace (d) Amazon Access Samples

Figure 6 shows the effective privacy level of the current methods on four datasets. Compared to Figure 5, the privacy parameters are higher, which means the privacy level are lower under the attack model. For example, for Trajectory, when $\varepsilon = 0.5$, CIM achieves a ε'' at 1.0360, while it achieves a ε' at 0.7832 in Figure 5a. Similarly for NetTrace, when $\varepsilon = 0.3$, the privacy level of Markov increases from 0.3852 to 0.4598, and the privacy level of WT increases from 0.8035 to 1.9210. We can infer from the experimental results that, the privacy level of these methods are lower under the attack model, and the changes are related to the correlation of the time series and the impulse response function of the filter.

5.3.4. Comparison of Privacy Level

More specifically, this section measures the privacy parameter ε' and ε'' of the current methods on four datasets under the fixed privacy parameter $\varepsilon = 0.7$.

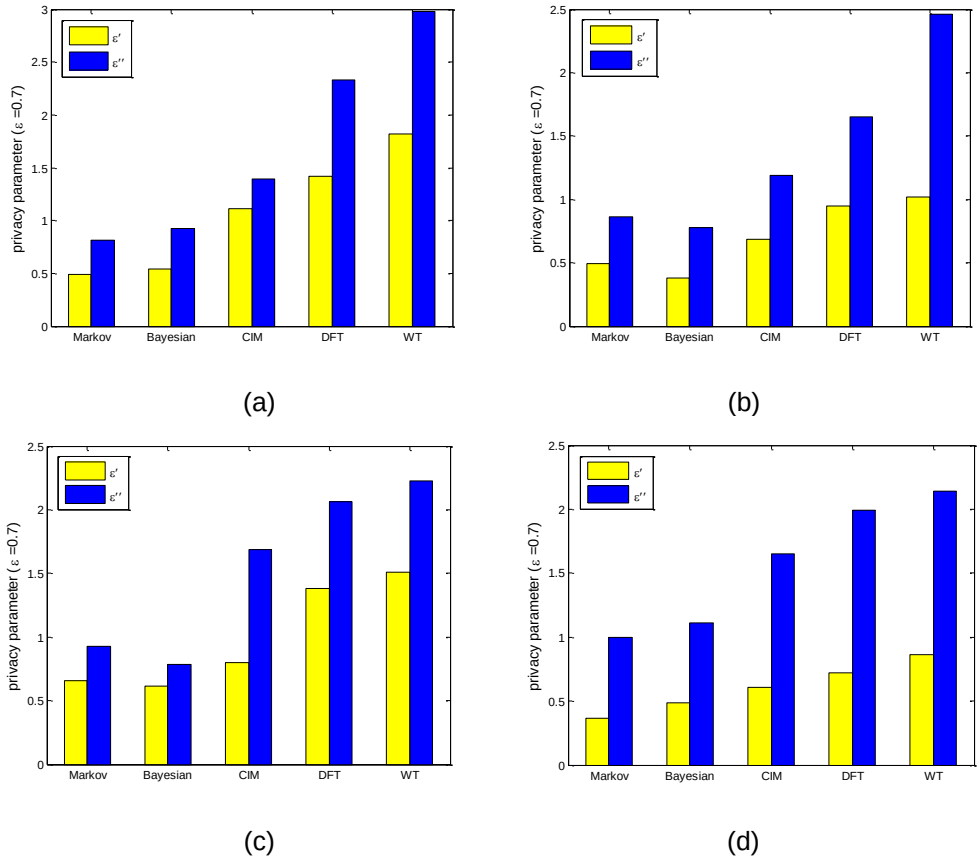

Figure 7. Comparison of Practical Privacy Level and Effective Privacy Level under the Attack Model (a) Trajectory (b) Diabetes (c) NetTrace (d) Amazon Access Samples

Figure 7 shows the practical privacy level and the effective privacy level under the attack model. The results show that, the proposed attack model obtains less effective privacy level than expected. The privacy level of different methods changes differently under the attack model. Specifically, for Trajectory, the practical privacy level of CIM is 1.1120, and the effective privacy level under the attack model is 1.3910, which means we obtain less privacy budget according to the effective privacy under the attack model. The similar trends can also be observed on other methods and other datasets. For example, for Diabetes, the practical privacy level of DFT is 0.9510, and the effective privacy level under the attack model is 1.6550. The experimental result shows that, under the proposed attack model, we can obtain less privacy budgets, and therefore the performance of the current methods can be compared.

5.4. Evaluation of Data Utility

We evaluate the data utility of the current methods under the attack model by calculating the MSE of each method on four datasets.

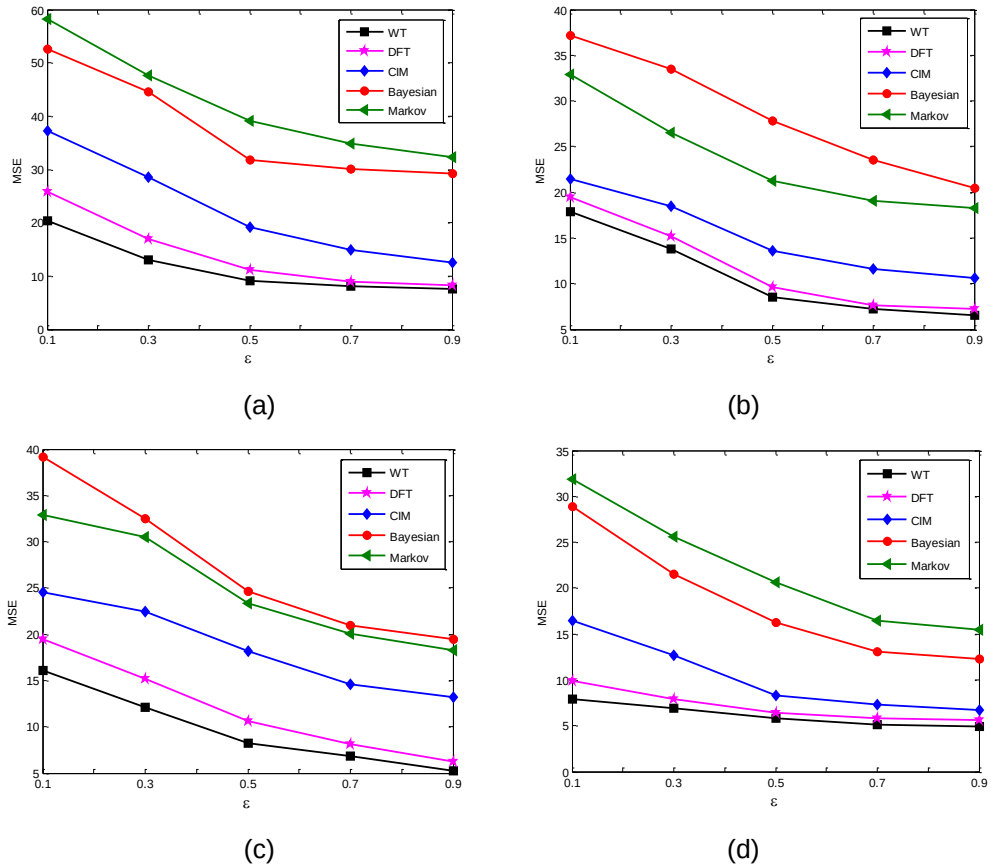

Figure 8. Comparison of Data Utility (a) Trajectory (b) Diabetes (c) NetTrace (d) Amazon Access Samples

Figure 8 shows that, for Trajectory, when $\varepsilon = 0.1$, DFT achieves a MSE of 25.7720, so the query answer is quite inaccurate; when $\varepsilon = 0.7$, MSE drops to 9.0380, retaining an acceptable utility of the result. Other datasets show similar trends. For example, when $\varepsilon = 0.7$, CIM achieves a MSE of 11.6690 for Diabetes, a MSE of 14.6120 for NetTrace, a MSE of 7.2930 for Amazon Access Samples, respectively. The experimental result confirms that the data utility is enhanced as the privacy parameter increases.

Moreover, Figure 8 shows that MSE decreases faster when ε increases from 0.1 to 0.5, than when ε increases from 0.5 to 0.9, which means that, for achieving a higher privacy level, there will be a larger utility cost. We also observe that, when $\varepsilon \geq 0.7$, CIM, DFT and WT perform stable, which indicates that they are capable of retaining the data utility while preserving a proper data privacy.

6. Conclusions

In this paper, we proposed an attack model on differential privacy preserving methods for correlated time series. The proposed attack model can verify the effectiveness of the current methods and measure the privacy level of them. The experimental results show that, the privacy level of the current methods degraded by approximately 50% under the attack model, and its function of working as a unified measurement was verified. Our future work will focus on the design of a differentially private time series data release mechanism, which can achieve a better trade-off between privacy level and data utility.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant No. 41671443), the Applied Basic Research Project of Wuhan Science and Technology Bureau (Grant No. 2016010101010024) and the National Natural Science Foundation of China (Grant No. 41371402).

References

[1] C. Dwork, "Differential Privacy", Automata, Languages and Programming. Springer Berlin Heidelberg, vol. 2, no. 26, **(2006)**, pp. 1-12.
[2] L. Cao, Y. Ou and P. S. Yu, "Coupled Behavior Analysis with Applications", IEEE Transactions on Knowledge & Data Engineering, vol. 8, no. 24, **(2012)**, pp. 1378-1392.
[3] B. Yang, I. Sato and H. Nakagawa, "Bayesian Differential Privacy on Correlated Data", SIGMOD/PODS, Melbourne, Victoria, Australia, **(2015)**, May 31-June 4.
[4] T. Zhu, P. Xiong, G. Li and W. Zhou, "Correlated Differential Privacy: Hiding Information in Non-IID Data Set", IEEE Transactions on Information Forensics & Security, vol. 2, no. 10, **(2015)**, pp. 229-242.
[5] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption", ACM SIGMOD International Conference on Management of Data, SIGMOD 2010, Indianapolis, Indiana, USA, **(2010)** June 6-11.
[6] X. Xiao, G. Wang and J. Gehrke, "Differential privacy via wavelet transforms", Knowledge & Data Engineering IEEE Transactions, vol. 8, no. 23, **(2010)**, pp. 1200-1214.
[7] X. Xiao, "Differentially Private Data Release: Improving Utility with Wavelets and Bayesian Networks", Web Technologies and Applications, Changsha, China, vol. 8709, **(2014)**, pp. 25-35.
[8] W. Jiang, C. Xie and Z. Zhang, "Wishart Mechanism for Differentially Private Principal Components Analysis," Computer Science, vol. 9285, **(2015)**, pp. 458-473
[9] C. Dwork, "A firm foundation for private data analysis", Communications of the ACM, vol. 1, no. 54, **(2011)**, pp. 86-95.
[10] C. Dwork, F. Mcsherry and K. Nissim, "Calibrating Noise to Sensitivity in Private Data Analysis", Theory of Cryptography. Springer Berlin Heidelberg, vol. 3, no. 3876, **(2006)**, pp. 265-284.
[11] S. P. Kasiviswanathan, "A Note on Differential Privacy: Defining Resistance to Arbitrary Side Information", Journal of Privacy & Confidentiality, **(2008)**.
[12] Y. Zheng, X. Xie, W.Y. Ma, "GeoLife: A Collaborative Social Networking Service among User, Location and Trajectory", Bulletin of the Technical Committee on Data Engineering, vol. 33, no. 2, **(2010)**, pp. 32-39.
[13] C. Blake, "UCI Repository of Machine Learning Databases", http://www.ics.uci.edu/~mlearn/MLRepository.html, **(1998)**.
[14] M. Hay, V. Rastogi, G. Miklau and D. Suciu, "Boosting the accuracy of differentially private histograms through consistency", Proceedings of the VLDB Endowment, vol. 3, no. 1, **(2009)**, pp. 66-69.
[15] C. Dwork, "Differential Privacy: A Survey of Results", Theory and Applications of Models of Computation. Springer Berlin Heidelberg, vol. 4978, **(2008)**, pp. 1-19.

Authors

Wenjun Xiong, is currently studying for her M.Sc. degree of Communication and Information Systems in Wuhan University. She has received the B.Sc. degree from Huanggang Normal University in 2014. Her research interests include cyber security and privacy preserving.

Zhengquan Xu, is currently a professor of LIESMARS in Wuhan University. He has received the B.Sc. and M.Sc. degree of Communication and Electronic System in Tsinghua University, and the Ph.D. degree of Biomedicine Engineering in Hong Kong Polytechnic University. His research interests include digital multimedia security and cloud computing security.

Hao Wang, is currently a Ph.D. candidate of Computer Science and Application in Wuhan University. He has received the M.Sc. degree from South-Central University for Nationalities in 2014. His research interests include data mining and privacy preserving.