JUSTC

Article

http://justc.ustc.edu.cn

Received: May 25, 2023; Accepted: September 11, 2023

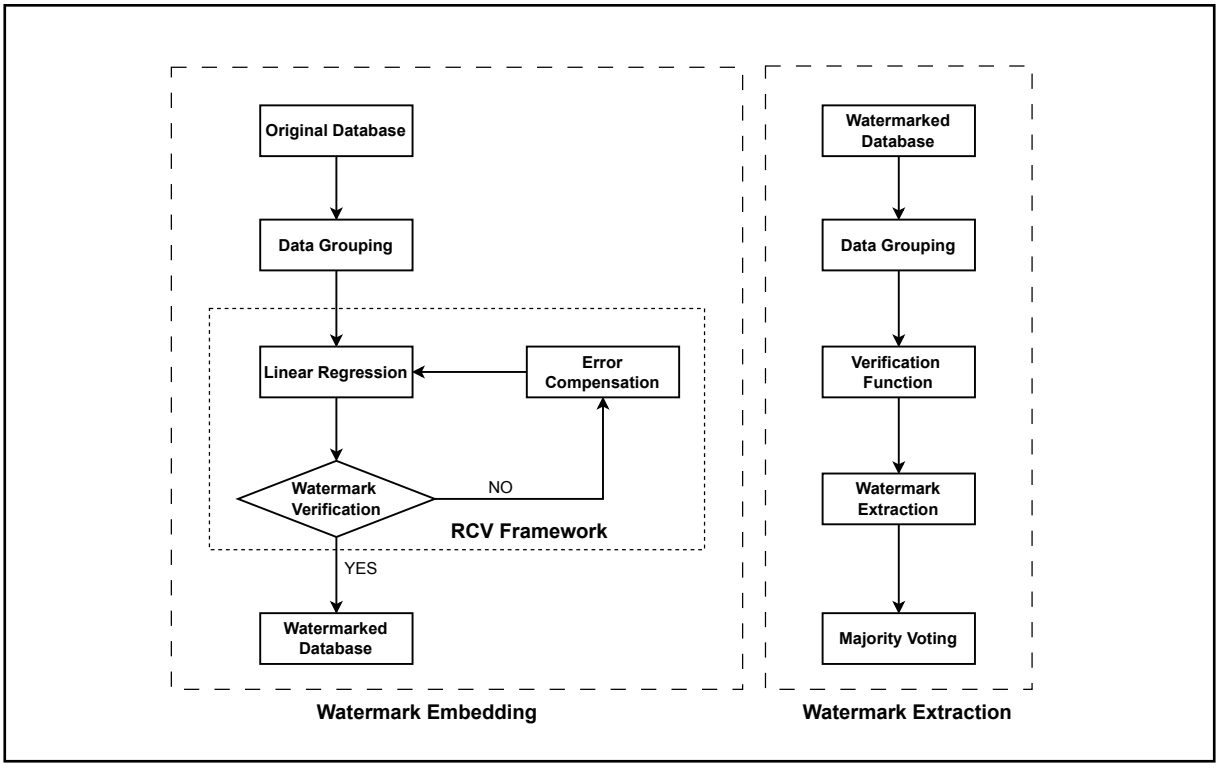
A statistical characteristics preserving watermarking scheme for time series databases

Yelu Yu[1], Zehua Ma[1], Jie Zhang[2], Han Fang[3] ✉, Weiming Zhang[1] ✉, and Nenghai Yu[1]

[1]*CAS Key Laboratory of Electromagnetic Space Information, University of Science and Technology of China, Hefei 230027, China;*
[2]*School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore;*
[3]*School of Computing, National University of Singapore, 117417, Singapore*

✉Correspondence: Han Fang, E-mail: fanghan@nus.edu.sg; Weiming Zhang, E-mail: zhangwm@ustc.edu.cn

Graphical abstract



A robust time series database watermarking method that can keep statistical characteristics unchanged.

Public summary

- This paper proposes a robust database watermarking scheme for time series databases, which can effectively ensure the consistency of statistical characteristics before and after watermark embedding.
- Based on the time-group characteristics of TSDBs, we propose a three-step watermarking method, which is based on linear regression, error compensation, and watermark verification, named RCV.
- The effectiveness of our scheme in keeping the statistical characteristics unchanged is verified both theoretically and practically. The experimental results show that our scheme has good robustness against database malicious attacks.

Citation: Yu Y L, Ma Z H, Zhang J, et al. A statistical characteristics preserving watermarking scheme for time series databases. *JUSTC*, **2024**, 54(4): 0401. DOI: 10.52396/JUSTC-2023-0091

JUSTC

Article

http://justc.ustc.edu.cn

Received: May 25, 2023; Accepted: September 11, 2023

A statistical characteristics preserving watermarking scheme for time series databases

Yelu Yu[1], Zehua Ma[1], Jie Zhang[2], Han Fang[3] ✉, Weiming Zhang[1] ✉, and Nenghai Yu[1]

[1]*CAS Key Laboratory of Electromagnetic Space Information, University of Science and Technology of China, Hefei 230027, China;*
[2]*School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore;*
[3]*School of Computing, National University of Singapore, 117417, Singapore*

✉Correspondence: Han Fang, E-mail: fanghan@nus.edu.sg; Weiming Zhang, E-mail: zhangwm@ustc.edu.cn

Cite This: *JUSTC*, **2024**, 54(4): 0401 (10pp)

Read Online

Abstract: Database watermarking is one of the most effective methods to protect the copyright of databases. However, traditional database watermarking has a potential drawback: watermark embedding will change the distribution of data, which may affect the use and analysis of databases. Considering that most analyses are based on the statistical characteristics of the target database, keeping the consistency of the statistical characteristics is the key to ensuring analyzability. Since statistical characteristics analysis is performed in groups, compared with traditional relational databases, time series databases (TSDBs) have obvious time-grouping characteristics and are more valuable for analysis. Therefore, this paper proposes a robust watermarking algorithm for time series databases, effectively ensuring the consistency of statistical characteristics. Based on the time-group characteristics of TSDBs, we propose a three-step watermarking method, which is based on linear regression, error compensation, and watermark verification, named RCV. According to the properties of the linear regression model and error compensation, the proposed watermark method generates a series of data that have the same statistical characteristics. Then, the verification mechanism is performed to validate the generated data until it conveys the target watermark message. Compared with the existing methods, our method achieves superior robustness and preserves constant statistical properties better.r.

Keywords: database watermarking; time series database; statistical characteristics; time-group characteristics

CLC number: TP309.2 **Document code:** A

1 Introduction

In the era of the digital economy, all walks of life are generating massive amounts of data every day. Databases, such as time-series databases which consist of thousands of pieces of data, have a high commercial value because the analysis of such data can effectively help the development of the industry.

However, precisely because of their high commercial value, databases are also exposed to the risk of security issues, such as data breaches, unauthorized copying, and copyright violations. Such risks also exist in the field of multimedia, in which one common solution is the digital watermarking technique. By inserting different watermark signals into different multimedia data, such as images[1,2], video[3,4], 3D mesh[5–7], the copyright of which can be effectively protected. Therefore, to remedy such risk in databases, there has been some research on digital watermarking techniques for time series data[8-10]. Duy et al.[8] proposed a watermarking scheme that embeds watermark information based on modifying the mean modulation relationship of approximation coefficients in the wavelet domain. This scheme treats time series data as one-dimensional signals and obtains good robustness to signal processing noise. However, it cannot guarantee robustness against common attacks on databases.

Therefore, database watermarking was introduced by Agrawal and Kiernan[11], who provided a new technical solution for database security technologies. Database watermarking technology by embedding unique watermark information in the database to prove the copyright of the database and prevent malicious piracy or unauthorized use. In addition, database watermarking technology can also be used to track and prevent tampering. When pirated databases are freely distributed, identity information can be extracted by specific means as reliable digital evidence. To deal with the increasingly severe database security problems, the development of database watermarking has become a concerning and vital research topic.

Since then, several robust database watermarking schemes[12–17] have been proposed for copyright protection and traceability. Guo et al.[12] proposed a robust watermarking algorithm for relational databases based on fingerprint recognition. The algorithm embedded fingerprints to identify legitimate recipients of relational data and provided a digital confidence level to identify owners and illegal distributors. Guo et al.[13] proposed an improved LSB algorithm for watermarking digital attributes in relational databases to protect copyright. Franco-Contreras et al.[14] proposed a robust database watermarking scheme that can achieve semantic control of data dis-

DOI: 10.52396/JUSTC-2023-0091
JUSTC, **2024**, 54(4): 0401

tortion and extend quantization index modulation (QIM) to circular histograms of numerical attributes. However, traditional robust database watermarking schemes have a potential drawback, i.e., the embedding of the watermark will change the statistical characteristics of databases, which may influence the analysis of the whole database.

One solution for maintaining the consistency of the database is reversible watermarking[18–29], by designing a reversible manner to embed the watermark. The watermark as well as the original database can be recovered from the watermarked database. The first reversible watermarking scheme for databases was proposed in 2006[18], in which histogram expansion was used for reversible database watermarking, but the anti-attack performance of this scheme was poor. In 2009, the technique called difference expansion-based watermarking (DEW)[19] was utilized to watermark a database in a reversible way, but since the watermark is embedded in the integer part, the data distortion caused by it is very large. In Ref. [20], Jawad and Khan combined the DEW scheme with the GA to enhance the robustness of DEW. Imamoglu et al.[22] improved DEW with the firefly algorithm to reduce data distortion. Hu et al.[21] designed a robust reversible database watermarking based on distortion control, which uses the genetic algorithm to optimize histograms for watermark embedding. This method ensures the data distortion of a single attribute value within a certain range. The authors in Refs. [25,26] proposed a robust and reversible watermarking algorithm based on continuous columns in histograms. In 2020, Ge et al.[28] proposed a novel, robust, and reversible database watermarking technique, named histogram shifting watermarking based on random forest and genetic algorithm (RF-GAHCSW). However, the reversible process requires people with key permissions to restructure the non-destructive database for analysis, which is unsatisfactory in real life. First, the reverse operation is computationally complex. Second, the common case is that the watermarked databases are expected to be analyzed by people without such permissions.

The analysis of the database is often conducted on the statistical characteristics of the data, e.g. mean and variance. Therefore, the key point to ensure the analyzability of the database is keeping the statistical characteristics unchanged before and after watermark embedding. However, none of the existing methods can satisfy this goal. Therefore, designing a database watermarking method that maintains the statistical property invariance is currently an urgent demand.

To this end, this paper proposes a robust watermarking scheme for time series databases (TSDBs) that can effectively embed the watermark while maintaining the statistical characteristics at the same time. Statistical characteristics analysis often needs to be performed with groups of data. Compared with traditional relational databases which are clustered with single rows and single columns, TSDBs have obvious time-grouping characteristics where the information contained in the same group effectively reflects the characteristics of a certain time period. Based on the time-group characteristics of the time series database, we propose performing watermark embedding on a group basis rather than on a point basis. Specifically, we propose a three-step watermarking method, which is based on linear regression, error compensation, and watermark verification, named RCV. First, based on the linear regression model and error compensation, the proposed watermark method could generate a series of data that have the same statistical characteristics as the original database. Then, the verification mechanism is performed to validate the generated data until it conveys the target watermark message. The specific process of the RCV method is described in section 2.2.3. In watermark extraction, we will determine the final extracted watermark bits by the majority voting principle.

The main contributions of this article are summarized as follows:

(Ⅰ) We propose a robust watermarking scheme that can effectively keep the statistical characteristics unchanged. Based on the time-group characteristics of time series databases, we propose a mechanism named RCV to embed the watermark into data groups. We verify the validity of the proposed RCV scheme in statistical characteristics preservation both theoretically and practically.

(Ⅱ) Extensive experimental results indicate that our method has strong robustness and can resist common database attacks. For alteration attacks, extracting watermarks maintains a correct rate of 0.84 even with altered groups up to 90%. Our method maintains a high watermark extraction accuracy for deletion and insertion attacks.

2 The proposed method

2.1 Motivation

Since time series databases have obvious time-group characteristics, data editing operations on time series databases are often performed in groups of data (e.g. the data in a certain time period). Therefore, in this paper, we proposed embedding the watermark in groups rather than in individual data points. The group-based operation has two advantages: (ⅰ) Embedding the same watermark bit in the group of data is similar to spread spectrum watermarking, which can effectively improve the redundancy and enhance the robustness of watermarking; (ⅱ) The operation within the group is more conducive to the maintenance of statistical properties. Compared with the embedding in individual data points, the modifications between different data in the same group can compensate for each other and thus better satisfy the statistical characteristics preserving properties.

Based on this idea, we propose a three-step watermarking scheme named RCV which is realized by a “regression, compensation, and verification” operation. Based on the properties of the linear regression model, watermarking can be effectively achieved with statistical characteristics preserving.

2.2 Framework

In this section, we mainly introduce the proposed watermarking framework. The whole framework can be divided into three phases: preprocessing, watermark embedding, and watermark extraction. Before introducing each phase, we first describe the common components of the time series database.

DOI: 10.52396/JUSTC-2023-0091
JUSTC, **2024**, 54(4): 0401

2.2.1 Composition of time series database

For better illustration, we give an example of a time series database[30]①, as shown in Table 1. The time series database mainly contains four parts: point, timestamp, tag, and field. The definition of each of them can be expressed as:

• Point: the piece of data in the database, for example, "67.20" in the second row of the "Price" column.

• Timestamp: a column of points that must exist in a time series database, which indicates the time point when the data were collected.

• Tag: a column of points representing the attribute of the collected data, which generally does not change with time, such as the "Information" column in Table 1.

• Field: a column of points representing the measured value of the data, which fluctuates smoothly over time, such as "Price" and "Demand" in Table 1.

In addition, for a quick reference, we list the notations used in this paper, as shown in Table 2.

2.2.2 Preprocessing

In the preprocessing phase, any form of watermark information (such as pictures, text, sounds, etc.) will be converted into binary bit sequences W with length l. $W \in \{0,1\}^l$ will be the watermark to be embedded. Then we cluster the original database D into m groups according to its timestamp and embed a 1-bit watermark messages in each group. It should be noted that timestamp points are usually not allowed to be modified, so we group them according to timestamp points. The function used for grouping is denoted as $\mathcal{G}$, which is used in both the embedding side and extraction side. After grouping, we could obtain m groups with the different timestamps, denoted as T^i, $i \in [1,m]$. The grouping function $\mathcal{G}$ is, specifically defined as $T^i = \mathcal{G}(D,m) = \{D_{ij} | N*(i-1)/m < j \le N*i/m,\ i \in [1,m]\}$. For each group, there are n timestamp points, denoted as t_j^i, $j \in [1,n]$. Then we have to determine which bit will be embedded in each group. This operation is realized by a mapping function $\mathcal{F}_t$. For the ith group, $\mathcal{F}_t$ receives the timestamp points T^i, and a secret key K_s②, and watermark length l as inputs, and outputs the index of the watermark to be embedded in this group denoted as k^i:

$$k^i = \mathcal{F}_t(T^i, K_s, l), \tag{1}$$

Table 1. The Circuit Load Dataset.

Timestamp	Information	Price	Demand
1041350400	Uniform Singapore Energy Price	67.28	3268.933
1041352200	Uniform Singapore Energy Price	67.20	3208.756
1041354000	Uniform Singapore Energy Price	61.32	3143.460
1041355800	Uniform Singapore Energy Price	56.23	3087.796
1041357600	Uniform Singapore Energy Price	56.04	3042.699

① The "Circuit Load Dataset" records the electricity price and load data of a certain device over a period of time, and an observation value is recorded every half an hour.

② K_s should be selected from a large key space so that it is computationally infeasible for an attacker to guess the key.

Table 2. Notations used in the paper.

Symble	Description
W	Original watermark
W_e	Extracted watermark
K_s	The secret key
l	The length of the watermark
N	Number of points in the database
n	Number of points in the group
m	Total number of groups
$\mathbb{T}$	The set of timestamp points
T^i	The ith group of timestamp points
t_i^j	Timestamp point of the jth data in group i
$\mathcal{G}$	Timestamp grouping function
$\mathcal{F}_t$	Function to map the watermark index to timestamp groups
$\mathcal{F}_v$	Verification function
$\mathbb{X}$	Confidential points (data to be watermarked)
X^i	The ith group of $\mathbb{X}$
x_j^i	Confidential point of the jth data in X^i
$\mathbb{Y}$	Watermarked confidential points
Y^i	The ith group of $\mathbb{Y}$
y_j^i	Confidential point of the jth data in Y^i
$\mathbb{S}$	Nonconfidential points (reference data)
S^i	The ith group of $\mathbb{S}$
s_j^i	Nonconfidential point of the jth data in S^i
A^i	Noise of standard normal distribution for group i
B^i	Regression residuals of group i
C^i	Compensation parameter of group i
$\sigma_{B^i}^2$	Variance of B^i
$\sigma_{X^i}^2$	Variance of X^i
$\sigma_{S^i}^2$	Variance of S^i
$\sigma_{Y^i}^2$	Variance of Y^i
$\sigma_{X^iS^i}$	Covariance between X^i and S^i
$\sigma_{Y^iS^i}$	Covariance between Y^i and S^i

where $k^i \in [1,l]$. In this paper, $\mathcal{F}_t$ is achieved by

$$\mathcal{F}_t(T^i, K_s, l) = \mathrm{mod}\left(\mathcal{H}\left(K_s, \sum_{j=1}^{n} t_j^i\right), l\right), \tag{2}$$

where $\mathrm{mod}(\cdot)$ indicates the modulo operation, and $\mathcal{H}$ indicates the Hashing operation. Based on Eq. (2), we could determine the index of the watermark bit be embedded in the group i. It should be noted that m should be larger than l for full watermark embedding. In addition, according to the definition of $\mathcal{F}_t$, the watermark bit with the same index might be embedded more than once.

2.2.3 Watermark embedding

For a typical time series database, the points of the "timestamp" and "tag" columns often have less information than the "field" column. Therefore, it would be better to em-

DOI: 10.52396/JUSTC-2023-0091
JUSTC, 2024, 54(4): 0401

bed the watermark into "field" points. Assume that in "field" points, there are several column points with a confidential attribute such as salary information, and several column points with a nonconfidential attribute such as behavior information. The data we need to maintain statistical characteristics rely more on points with confidential attributes. Therefore, in this paper, we propose embedding the watermark into the confidential points while leaving the nonconfidential points unmodified. The nonconfidential points could effectively serve as a reference to maintain the statistical characteristics of the confidential points. In addition, each point in the "field" column corresponds to a "timestamp" feature, so the group information of "timestamp" can be directly applied to the "field" points.

Denote the column of points we want to embed as $\mathbb{X}$, and the referenced column of points as $\mathbb{S}$. Since we have clustered the database with m groups, the goal is to embed each watermark bit into each group of $\mathbb{X}$ to generate the watermarked points $\mathbb{Y}$.

Besides, the statistical characteristics of each group of $\mathbb{X}$ (denoted as $X^i \in \mathbb{X}, i \in [1,m]$), *i.e.*, the mean of X^i, the variance of X^i should be the same as that of $Y^i \in \mathbb{Y}, i \in [1,m]$, and the covariance between X^i and S^i is equal to the covariance between Y^i and S^i.

To achieve this goal, we propose RCV, a regression-compensation-verification-based method for watermarking. Based on X^i and S^i, and the standard normal distribution data points A^i, we first use linear regression models and conduct a error compensation mechanism to generate a set of data points Y^i, guaranteeing that the statistical characteristics of X^i and Y^i are consistent. Then we perform a watermark verification mechanism $\mathcal{F}_v$ to validate whether the watermarked Y^i could convey the watermarked message. If Y^i passes the verification, that is, Y^i has embedded watermark information, then proceed to the next group of data. Otherwise, we re-execute the regression-compensation process to generate Y^i and repeat until Y^i can pass the verification mechanism. To more clearly describe the RCV scheme, our specific examples will be given in Section 3.1.1.

Specifically, for the ith group, we use a linear regression model with parameter $\bar{\alpha}_0^i$ and $\bar{\alpha}_1^i$ to predict the value of $x_j^i \in X^i, j \in [1,n]$ with $s_j^i \in S^i, j \in [1,n]$,

$$\hat{x}_j^i = \bar{\alpha}_0^i + \bar{\alpha}_1^i \times s_j^i, \tag{3}$$

where $\hat{x}_j^i$ is the predicted point. Then we sample a set of noise A^i with the same size of X^i from the standard normal distribution. A^i is further represented by another linear regression model with parameter $\bar{\beta}_0$, $\bar{\beta}_1$ and $\bar{\beta}_2$, *i.e.*,

$$\hat{a}_j^i = \bar{\beta}_0^i + \bar{\beta}_1^i \times s_j^i + \bar{\beta}_2^i \times x_j^i, \tag{4}$$

where $\hat{a}_j^i$ indicates the predicted value of $a_j^i \in A^i$. After obtaining $\hat{A}^i$, we calculate the differences between $\hat{A}^i$ and A^i, denoted as B^i, which can be formulated as:

$$B^i = A^i - \hat{A}^i. \tag{5}$$

Then we calculated the compensation parameter C^i according to B^i with Eq. (6):

$$c_j^i = \frac{b_j^i}{\sigma_{B^i}} \xi, \tag{6}$$

where $c_j^i \in C^i$, $b_j^i \in B^i$, and ξ can be calculated as

$$\xi^2 = \sigma_{X^i}^2 - \frac{\sigma_{X^i S^i}^2}{\sigma_{S^i}^2}, \tag{7}$$

where $\sigma_{X^i}^2$, $\sigma_{S^i}^2$, and $\sigma_{X^i S^i}$ are the variance of X^i, S^i, and the covariance between X^i and S^i, respectively. After determining the compensation parameter C^i, the final watermarked data Y^i of the group i can be calculated as:

$$y_j^i = \hat{x}_j^i + c_j^i. \tag{8}$$

In this manner, the generated Y^i can maintain the same mean, the same variance, and the same covariance (with S^i) as that of X^i. The relevant proof can be found in Section 2.2.5.

Then, we will conduct a verification mechanism $\mathcal{F}_v$ to check whether the generated Y^i can convey the watermark bit, *i.e.* whether $\mathcal{F}_{\sqsubseteq}(\mathcal{Y}^i, \mathcal{K}_s) = \mathcal{W}(\|^i)$. The verification mechanism in this paper is shown as Eq. (9):

$$\mathcal{F}_v(Y^i, K_s) = \mod\left(\mathcal{H}\left(K_s, \sum_{j=1}^{n} \lfloor y_j^i \rfloor\right), 2\right), \tag{9}$$

where $\lfloor \cdot \rfloor$ indicates the round down function. The whole embedding process will be conducted until the generated Y^i has passed the verification. Then we replace all the $\mathbb{X}$ with $\mathbb{Y}$ to generate the final watermarked database D_w.

The embedding algorithm is illustrated as Algorithm 1.

Algorithm 1: Watermark embedding algorithm

Input: Secret key K_s, original database D, watermark W, group number m

Output: Watermarked database D_w

$n, T^i = \mathcal{G}(\mathbb{T} \in D, m), i \in [1,m]$
for $i = 1 \to m$ **do**
  $k^i = \mathcal{F}_t(T^i, K_s, l)$
  do
    Generate A^i;
    for $j = 1 \to n$ **do**
      $\hat{x}_j^i = \bar{\alpha}_0^i + \bar{\alpha}_1^i \times s_j^i$;
      $\hat{a}_j^i = \bar{\beta}_0^i + \bar{\beta}_1^i \times s_j^i + \bar{\beta}_2^i \times x_j^i$;
    end
    $B^i = A^i - \hat{A}^i$;
    for $j = 1 \to n$ **do**
      $c_j^i = \frac{b_j^i}{\sigma_{B^i}} \xi$;
      $y_j^i = \hat{x}_j^i + c_j^i$;
    end
    $v^i = \mathcal{F}_v(Y^i, K_s)$
  while $v^i != W(k^i)$;
end
$D_w = D(\mathbb{X} \Rightarrow \mathbb{Y})$

DOI: 10.52396/JUSTC-2023-0091
JUSTC, **2024**, 54(4): 0401

2.2.4 Watermark extraction

In this section, we will introduce the mechanism to extract the watermark from the watermarked database D_{w}. Specifically, we first use the grouping algorithm $\mathcal{G}$, which is the same as the embedding phase, to cluster the database with the secret key K_{s} and the grouping numbers m. For each group i, we utilize the mapping function $\mathcal{F}_l$ to determine the index denoted as k^i of the extracted watermark according to each T^i, as shown in Eq. (1) and Eq. (2). Then we take the embedded column Y_i and perform the verification mechanism $\mathcal{F}_v$ on it to extract the embedded watermark bit W_{e}. In embedding, the watermark bit with the same index might be embedded into more than one group, in extraction, for different groups of watermarks corresponding to the same index, we will determine the final extracted watermark bits by majority voting principle. Specifically, for each index of watermark $k \in [1,l]$, we record the number of bits "0" (denoted as Num_0^k) and the number of bits "1" (denoted as Num_1^k) that are extracted from all groups corresponding to k. If Num_0^k is larger than Num_1^k, we regard the watermark bit of index k as "0", otherwise, we regard the bit as "1".

For example, suppose there are 3 groups all corresponding to the 7th bit, and the extracted watermark of these 3 groups is $\{1,1,0\}$, so the 7th bit watermark will be determined as 1 according to the majority voting principle.

2.2.5 Statistical characteristics analysis

In this section, we will analyze the statistical characteristics of the database before and after embedding.

Since in this paper, we use the linear regression model to predict confidential points $\mathbb{X}$ from nonconfidential points $\mathbb{S}$, two properties of the linear regression model are first introduced. Take the simple linear regression model with parameter p, q, and n points as an example, as shown in Eq. (10):

$$y_i = px_i + q, i \in [1,n]. \tag{10}$$

Algorithm 2: Watermark extraction algorithm

Input: Secret Key K_{s}, watermarked database D_{w}, group number m, watermark length l

Output: Watermark W_{e}

$E \in \mathbb{R}^{l\times 2} = 0$;

$n, T^i = \mathcal{G}(\mathbb{T} \in D_w, m), i \in [1,m]$

for $i = 1 \to m$ **do**

  $k^i = \mathcal{F}_l(T^i, K_{\mathrm{s}}, l)$;

  $v^i = \mathcal{F}_v(Y^i, K_{\mathrm{s}})$;

  $E(k^i, v^i) = E(k^i, v^i) + 1$;

end

for $k = 1 \to l$ **do**

  if $E(k,0) \geq E(k,1)$ **then**

    $W_{\mathrm{e}}(k) = 0$;

  else

    $W_e(k) = 1$;

  end

end

Denoting the predicted value as $\hat{y}_i = px_i + q$, and the predicted residual as $e_i = px_i + q - y_i = \hat{y}_i - y_i$. The two properties are as follows: (ⅰ) The mean value of e_i is 0. (ⅱ) e_i is orthogonal to x_i.

Here we give the proof. The purpose of model training is to solve the optimal parameters p and q, which makes the regression result as close as possible to the true value. Without loss of generality, we use the least squares method to solve the optimization problem. Denoting the input data as x_i, $i = 1,2,\ldots,n$, we can obtain a set of predicted values $f(x_i)$ according to Eq. (10) and calculate the squared loss between $f(x_i)$ and the existing real value y_i based on the squared loss function as follows:

$$\mathcal{L}(p,q) = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - y_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(px_i + q - y_i)^2, \tag{11}$$

where $\mathcal{L}(p,q)$ is the squared loss with parameters p and q. For the ideal case, the partial derivative of $\mathcal{L}$ on p and q should be 0, i.e.:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial p} &= \frac{1}{n}\sum_{i=1}^{n} 2(px_i + q - y_i)\cdot x_i = \\ &\frac{2}{n}\left(p\sum_{i=1}^{n}x_i^2 - \sum_{i=1}^{n}(y_i - q)x_i\right) = 0,\end{aligned} \tag{12}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial q} &= \frac{1}{n}\sum_{i=1}^{n} 2(px_i + q - y_i) = \\ &\frac{2}{n}\left(\sum_{i=1}^{n}(px_i - y_i)\right) + 2q = 0.\end{aligned} \tag{13}$$

According to Eq. (13), we can obtain:

$$\frac{1}{n}\sum_{i=1}^{n}(px_i + q - y_i) = \frac{1}{n}\sum_{i=1}^{n}(e_i) = \mathcal{E}(e_i) = 0, \tag{14}$$

where $\mathcal{E}(e_i)$ indicates the mean value of e_i. Based on Eq. (14), we could see property (ⅰ) is satisfied. For property (ⅱ), it could be proved with Eq. (12). According to Eq. (12), we can rewritten the equation as:

$$\sum_{i=1}^{n}(px_i + q - y_i)\cdot x_i = \sum_{i=1}^{n} e_i \cdot x_i = 0. \tag{15}$$

Therefore, e_i is orthogonal to x_i.

Besides, according to Eq. (12) and Eq. (13), we could determine the value of p as:

$$\begin{aligned}p &= \frac{\sum_{i=1}^{n}(x_i y_i - \mathcal{E}(Y)x_i)}{\sum_{i=1}^{n}(x_i^2 - \mathcal{E}(X)x_i)} = \\ &\frac{\sum_{i=1}^{n}(x_i - \mathcal{E}(X))(y_i - \mathcal{E}(Y))}{\sum_{i=1}^{n}(x_i - \mathcal{E}(X))^2} = \frac{\sigma_{XY}}{\sigma_X^2},\end{aligned} \tag{16}$$

DOI: 10.52396/JUSTC-2023-0091
JUSTC, **2024**, 54(4): 0401

where $\mathcal{E}(X)$ indicates the mean value of $x_i \in X$, $\mathcal{E}(Y)$ indicates the mean value of $y_i \in Y$, σ_{XY} is the covariance of $x_i \in X$ and $y_i \in Y$ and σ_X^2 is the variance of X. The value of q can be determined as:

$$q = \mathcal{E}(Y) - p\mathcal{E}(X). \tag{17}$$

Now we analyze of the statistical characteristics of Y^i and X^i for specific group i. Based on the proposed embedding mechanism, the value $y_j^i \in Y^i$ could be generated as:

$$y_j^i = \hat{x}_j^i + c_j^i = \bar{\alpha}_0 + \bar{\alpha}_1 \times s_j^i + c_j^i. \tag{18}$$

Mean: For the mean of Y^i, we can get:

$$\mathcal{E}(Y^i) = \mathcal{E}\left(\hat{X}^i\right) + \mathcal{E}(C^i) = \mathcal{E}\left(\hat{X}^i\right) + \mathcal{E}\left(\frac{B^i}{\sigma_{B^i}}\xi\right), \tag{19}$$

where B^i is the predicted residual of A^i, as shown in Eq. (5). According to property (i), $\mathcal{E}(B^i)$ is 0, so Eq. (19) could be rewritten as:

$$\mathcal{E}(Y^i) = \mathcal{E}\left(\hat{X}^i\right) + \mathcal{E}(B^i) = \mathcal{E}\left(\hat{X}^i\right) = \mathcal{E}(X^i). \tag{20}$$

Therefore, the mean value of Y^i is equal to the mean value of X^i.

Variance: The variance of Y^i, we can be calculated as:

$$\begin{aligned}
&\sigma_{Y^i}^2 = \mathcal{E}\left((Y^i)^2\right) - \mathcal{E}^2(Y^i) = \\
&\mathcal{E}\left(\left(\hat{X}^i + C^i\right)^2\right) - \mathcal{E}^2(X^i) = \\
&\mathcal{E}\left(\hat{X}^{i\,2}\right) + \mathcal{E}\left((C^i)^2\right) + 2\mathcal{E}\left(\hat{X}^i(C^i)\right) - \mathcal{E}^2(X^i) = \\
&\mathcal{E}\left(\hat{X}^{i\,2}\right) + \mathcal{E}\left(\left(\frac{B^i}{\sigma_{B^i}}\xi\right)^2\right) + 2\mathcal{E}\left(\hat{X}^i\left(\frac{B^i}{\sigma_{B^i}}\xi\right)\right) - \mathcal{E}^2(X^i) = \\
&\mathcal{E}\left((\bar{\alpha}_0 + \bar{\alpha}_1 \cdot S^i)^2\right) + \xi^2 - \mathcal{E}^2(X^i).
\end{aligned} \tag{21}$$

According to Eq. (16) and Eq. (17),

$$\begin{aligned}
\bar{\alpha}_1 &= \frac{\sigma_{X^iS^i}}{\sigma_{S^i}^2}, \\
\bar{\alpha}_0 &= \mathcal{E}(X^i) - \bar{\alpha}_1\mathcal{E}(S^i).
\end{aligned} \tag{22}$$

Based on Eq. (7), Eq. (21) could be written as:

$$\begin{aligned}
&\sigma_{Y^i}^2 = \mathcal{E}\left((\bar{\alpha}_0 + \bar{\alpha}_1 \cdot S^i)^2\right) + \xi^2 - \mathcal{E}^2(X^i) = \\
&\mathcal{E}^2(X^i) + \frac{\sigma_{X^iS^i}^2}{\sigma_{S^i}^2} + \xi^2 - \mathcal{E}^2(X^i) = \sigma_{X^i}^2.
\end{aligned} \tag{23}$$

Therefore, $\sigma_{Y^i}^2$ is equal to $\sigma_{X^i}^2$.

Covariance: The covariance between Y^i and S^i can be written as:

$$\begin{aligned}
&\sigma_{Y^iS^i} = \mathcal{E}(Y^iS^i) - \mathcal{E}(Y^i)\mathcal{E}(S^i) = \\
&\mathcal{E}((\bar{\alpha}_0 + \bar{\alpha}_1S^i + C^i)S^i) - \mathcal{E}(\bar{\alpha}_0 + \bar{\alpha}_1S^i + C^i)\mathcal{E}(S^i) = \\
&\bar{\alpha}_1\left(\mathcal{E}\left((S^i)^2\right) - \mathcal{E}^2(S^i)\right) = \bar{\alpha}_1\sigma_{S^i}^2 = \sigma_{X^iS^i}.
\end{aligned} \tag{24}$$

Therefore, when conducting an ideal linear regression and satisfying Eq. (22), the mean value, variance value, and covariance value with reference column S^i of the embedded column Y^i is equal to that of the original column X^i.

3 Experimental results and analysis

Experiments are implemented on a common PC with an Intel Core i5 CPU and RAM of 16 GB. It should be noted that the proposed RCV scheme could be applied to all time series databases rather than specific databases with special characteristics. Without loss of generality, we take one available public time series database, the Circuit Load Data of Singapore[30], to evaluate the performance of our scheme, in which we focus on three attributes: timestamp, electricity load, and real-time electricity price. The equipment collects power load and electricity price information every half an hour, i.e., 48 data points per day, and a total of 340080 data points are collected from January 1, 2003, to May 22, 2022. In our experiment, the length of the embedded watermark is set as 70 bits with a group size of 48 data points, the same size as one day. Note that the length and grouping size of the watermark information can be selected reasonably by the owner according to the specific situation of the database. For the time series database, the statistical analysis for a certain time period is more important, so we pay more attention to how the statistical characteristics of each group remain unchanged. To better analyze the information contained in the data, we recommend that the group size be a day, a week, or a meaningful period of data points.

The following experiments are illustrated in two aspects. The first part analyzes the statistical characteristics and compares the statistical preserving properties of the proposed RCV scheme with other watermarking schemes. In the second part, the robustness of our scheme is compared with some watermarking schemes.

3.1 Statistical characteristics analysis

This subsection is mainly divided into two parts. The first part proves the effectiveness of the proposed scheme in preserving statistical characteristics. Then, the statistical characteristics of RCV are compared and analyzed with the existing database watermarking schemes, including GAHSW[21], DEW [19], and time series data watermarking scheme Signal[8].

3.1.1 Local statistical characteristics preservation

The proposed RCV does not degrade data availability after embedding watermarks while keeping the local statistical properties of the data the same. The watermarked data Y^i and X^i obtained by each group have exactly the same mean and variance as the original data. Additionally, the covariance between Y^i and S^i is exactly the same as the covariance between X^i and S^i. To better illustrate the statistical characteristics preservation of the proposed RCV scheme, we randomly select a timestamp points group T^i as an example for analysis. In this paper, we grouped the time series database by days, and each group has 48 data points, including the values of the confidential data $\mathbb{X}$ (electricity price) and the non confidential data $\mathbb{S}$ (electricity load). As shown in Table 3, the mean and the variance of X^i are 3304.7456 and 52423.3529, respectively. The correlation between X^i and S^i is 3808.8394. Regressing X on S, we obtain $\bar{\alpha}_1 = 9.1830$ and $\bar{\alpha}_0 = 2650.2113$ based on Eq. (3) and calculate the predicted values for each observation. Then, the next column A^i repres-

DOI: 10.52396/JUSTC-2023-0091
JUSTC, 2024, 54(4): 0401

Table 3. An illustration of local statistical characteristics preservation.

j	S^i	X^i	A^i	B^i	C^i	Y^i
1	67.28	3268.9330	1.0196	1.1248	149.2936	3417.3360
2	67.20	3208.7560	0.4891	0.4946	65.6496	3332.9573
3	61.32	3143.4600	0.0881	0.0551	7.3148	3220.6265
4	56.23	3087.7960	1.1048	1.0400	138.0377	3304.6081
5	56.04	3042.6990	0.1171	−0.0209	−2.7782	3162.0474
6	56.04	3005.6520	−0.0014	−0.2014	−26.7276	3138.0980
7	15.50	2972.3830	−1.2546	−1.0225	−135.7172	2656.8303
8	0.45	2943.7910	−0.4615	−0.0961	−12.7568	2641.5869
9	0.36	2925.4070	0.3230	0.6588	87.4400	2740.9572
10	56.26	2920.2680	0.8871	0.5417	71.8971	3238.7430
11	56.26	2926.7580	−0.2640	−0.5986	−79.4506	3087.3953
12	61.16	2950.6090	1.0661	0.7125	94.5663	3306.4088
13	61.16	2973.8190	−1.7847	−2.0995	−278.6736	2933.1689
14	61.16	2957.3940	−0.1732	−0.5155	−68.4239	3143.4186
15	61.15	2963.5860	−0.3141	−0.6459	−85.7377	3126.0129
16	63.64	3038.5090	1.6665	1.4300	189.8157	3424.4320
17	67.53	3128.4260	−0.0151	−0.1480	−19.6449	3250.6932
18	68.95	3219.8390	−0.1103	−0.1074	−14.2533	3269.1246
19	70.82	3303.0970	−0.3868	−0.2672	−35.4602	3265.0899
20	76.45	3369.2720	−1.0612	−0.8986	−119.2726	3232.9777
21	81.48	3425.1100	0.8361	1.0316	136.9284	3535.3690
22	81.49	3465.0740	0.2986	0.5609	74.4454	3472.9779
23	84.29	3476.7590	0.0792	0.3273	43.4482	3467.6930
24	81.49	3465.4220	−1.3827	−1.1199	−148.6502	3249.8823
25	81.49	3452.4650	1.2789	1.5201	201.7647	3600.2972
26	81.49	3452.7290	−1.5737	−1.3321	−176.8113	3221.7212
27	81.49	3455.1140	1.1530	1.3986	185.6428	3584.1753
28	81.49	3452.6240	1.1965	1.4380	190.8654	3589.3979
29	81.49	3447.3760	−0.3799	−0.1473	−19.5474	3378.9851
30	81.48	3434.1760	−0.1369	0.0738	9.7966	3408.2373
31	81.48	3427.7340	−0.2859	−0.0860	−11.4202	3387.0205
32	81.49	3427.4710	0.0870	0.2863	38.0002	3436.5327
33	81.48	3420.4680	0.2455	0.4332	57.5070	3455.9477
34	81.41	3410.7270	1.6185	1.7908	237.7016	3635.4995
35	80.56	3399.8730	0.1039	0.2683	35.6126	3425.6050
36	80.56	3389.1320	−0.2204	−0.0740	−9.8275	3380.1649
37	81.41	3411.2840	−0.2401	−0.0669	−8.8806	3388.9173
38	86.27	3509.3370	−0.9111	−0.6323	−83.9281	3358.4990
39	89.92	3595.1740	−1.3489	−0.9705	−128.8231	3347.1220
40	89.92	3610.6350	−2.0792	−1.6749	−222.3222	3253.6228
41	89.92	3608.7420	0.5950	0.9961	132.2110	3608.1560
42	89.86	3607.5690	−0.4955	−0.0956	−12.6874	3462.7066
43	88.80	3601.6250	−1.1069	−0.7042	−93.4716	3372.1885
44	88.82	3571.8300	0.5348	0.8874	117.7875	3583.6313
45	86.90	3530.8770	−0.7458	−0.4386	−58.2231	3389.9893
46	84.17	3479.0890	−3.1900	−2.9366	−389.7892	3033.3537
47	81.56	3417.4420	1.0579	1.2396	164.5407	3563.7160
48	74.12	3331.4750	−1.5362	−1.4087	−186.9884	3143.8656
		3304.7456		Mean		3304.7456
		52423.3529		Variance		52423.3529
		3808.8394		Correlation with S^i		3808.8394

ents the set of random variables generated from a univariate normal distribution with mean 0 and variance 1. Next, based on Eq. (4), A^i is regressed on both X^i and S^i and we obtain $\bar{\beta}_0 = 4.5530$, $\bar{\beta}_1 = 0.0120$, and $\bar{\beta}_2 = -0.0017$. The prediction residuals calculated by Eq. (5) are shown in the next column as B^i. Based on $\sigma^2_{X^i}$, $\sigma^2_{S^i}$, and $\sigma_{X^iS^i}$ and using Eq. (7), we can calculate that $\xi = 132.0865$. Finally, we can calculate C^i and Y^i for each observation according to Eq. (6) and Eq. (8).

It is easily verified that Y^i has the same mean and variance as that of X^i. Besides, the correlation between Y^i and S^i is

DOI: 10.52396/JUSTC-2023-0091
JUSTC, **2024**, 54(4): 0401

3808.8394, which is exactly the same as that between X^i and S^i. Thus, the results of the analysis for which the mean and covariance are sufficient statistics, such as regression analysis, will be exactly the same when using Y^i in place of X^i.

In the verification mechanism, we need to verify whether the watermark information in the obtained group Y^i is the same as the watermark information to be embedded. Using Eq. (2), we calculate that the 7th-bit watermark "1" is to be embedded in the group. Then, Eq. (9) is used to calculate whether the watermark information carried in this group Y^i is "1". If right, then Y^i passes the verification; Otherwise, regenerate A^i and repeat the above steps until Y^i passes the verification.

3.1.2 Comparative experiment of statistical characteristics preservation

In this subsection, we compare the statistical characteristics preservation performance of the proposed watermarking scheme and the comparative ones, including Signal[8], GAHSW[21], and DEW[19]. As shown in Table 4, the mean, variance, and covariance between the attribute $\mathbb{X}$ and attribute $\mathbb{S}$ of the watermarked database with the proposed RCV are the same as the corresponding statistical characteristics of the original database. Specifically, our scheme can not only keep the local statistical characteristics unchanged but also keep the overall statistical characteristics unchanged. For the comparative schemes, the statistical characteristics of the database watermarked by GAHSW change slightly, yet the variance and covariance values of DEW and Signal have relatively large variations. The evaluation results verify the effectiveness of our proposed scheme in keeping the statistical characteristics unchanged.

3.2 Robustness analysis

Referring to the experimental setting of existing database watermark schemes, we consider three common attacks, i.e., insertion, deletion, and alteration, to evaluate the robustness of the proposed RCV scheme. The watermark extraction accuracy (*Acc*), i.e., the ratio of the correctly extracted bits in the extracted watermark bits, is used as the robustness metric, which can be calculated as follows:

$$Acc = \frac{\sum_{i=1}^{l} w_i \oplus w_i^{\det}}{l}, \tag{25}$$

where w_i is an embedded watermark bit and $w_i^{\det}$ is the extracted watermark bit. It should be noted that a higher *Acc* means higher robustness of watermarking schemes.

3.2.1 Comparative experiment of robustness

Considering that the high-value data of time series databases are often concentrated, the attacker tends to modify important parts of the continuous data. Therefore, we evaluate the *Acc* of the watermarked database under insertion, deletion, or alteration with percentages from 10% to 90% in steps of 10% of the data groups. The experimental results are illustrated in the following charts, in which the vertical axis represents the rate of successful watermark detection, and the horizontal axis represents the change of groups in attack percentage according to the size of the database.

To simulate the alteration attack, we alter the watermarked database with different ratios of data groups, mainly altering the measured values of the database. Fig. 1 shows *Acc* of the extracted watermarks of RCV, Signal, DEW, and GAHSW under alteration attack. We found that, for all schemes, the *Acc* of the extracted watermark decreases as the number of altered groups increases. Nevertheless, RCV is superior among other methods under alteration attack. This result is because our scheme has a higher embedding rate of watermark information, which makes the same amount of modification can be embedded with more duplicate information and is more resistant to an alteration attack. The GAHSW scheme embeds watermark information into the database through HSW (histogram shifting of prediction error expansion watermarking), which is mainly embedded by moving the left and right sides of the histogram. The embedding formula proposed in this scheme can only be embedded when the absolute value between the prediction error and the peak bin is equal, so the embedding rate of the watermark is not very high. The DEW scheme embeds watermark information through differential technology but also needs to meet certain conditions before embedding the watermark bit into the tuple. Compared with these schemes, our scheme embeds the watermark in each group and has a higher embedding rate. Therefore, even after an alteration of up to 90% of the data groups, the *Acc* of RCV is still higher than 80%.

In the deletion attack, we delete different ratios of the data groups from the watermarked database randomly. As shown in Fig. 2, for RCV, we can see that even only a small portion of the preserved database is dufficient for successful watermark extraction. Meanwhile, when the database suffers from a heavy deletion attack, e.g., 90% of the tuples of the database are deleted, Signal and DEW could only be extracted with *Acc* values of 49% and 10%, respectively, lacking the robustness to database deletion. For the Signal scheme, watermark information is embedded by modifying the average modulation relationship of the approximate coefficients in the wavelet domain. Although this way of treating time series as one-dimensional signals has good robustness against noise attacks, it will destroy the synchronization structure of the watermark signal and affect the robustness against database deletion attacks. Compared with GAHSW and DEW, RCV has higher data redundancy when embedding the watermark.

Table 4. Statistical characteristics of the database watermarked with different schemes.

	Original data	RCV	Signal	GAHSW	DEW
Mean	5137.4943	5137.4943	5134.7117	5136.9959	5145.5192
Variance	894129.9816	894129.9816	1245253.7502	894133.9886	918689.1496
Correlation with $\mathbb{S}$	19989.5356	19989.5356	19438.6446	19997.6797	–3243.3402

DOI: 10.52396/JUSTC-2023-0091
JUSTC, **2024**, 54(4): 0401

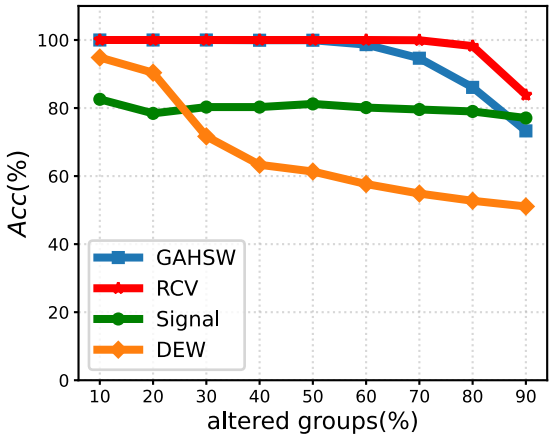


Fig. 1. A Comparison of watermark extraction *Acc* of RCV with Signal, DEW, and GAHSW after alteration attack.

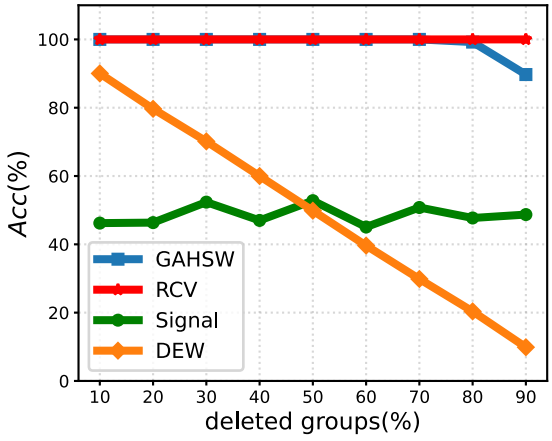


Fig. 2. A Comparison of watermark extraction *Acc* of RCV with Signal, DEW, and GAHSW after deletion attack.

Therefore, even though many groups are deleted, the watermark can be correctly extracted as long as the remaining groups contain the bits of the watermark.

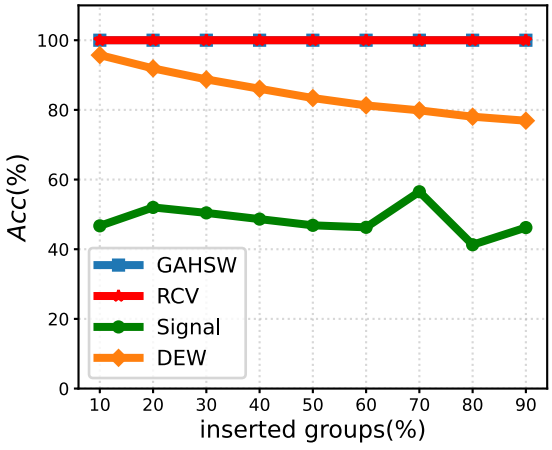


Fig. 3. A Comparison of watermark extraction *Acc* of RCV with Signal, DEW, and GAHSW after insertion attack.

For the insertion attack, data groups are randomly created and inserted between groups, aiming to weaken the embedded watermark. As shown in Fig. 3, the proposed RCV and GAHSW have excellent resilience to insertion attacks, while DEW and Signal are not very robust to it. Similarly, the insertion attack also destroys the synchronization structure of the watermark signal, resulting in the failure of the signal to extract the correct watermark.

4 Conclusions

Based on the inherent time-group characteristics of time series databases, this paper proposes a robust watermarking scheme for time series databases, which effectively ensures the consistency of statistical characteristics. Specifically, we devise a three-step scheme RCV, composed of regression, compensation, and verification operations. Based on the characteristics of the linear regression model, the statistical characteristics of the database can be kept constant. The experimental results show that the proposed method outperforms the existing methods in terms of effectiveness, robustness, and fidelity of statistical characteristics.

To maintain the availability of data, we design a database watermarking scheme with statistical feature preservation. However, in some scenarios where data accuracy is more needed, we have a higher goal for data availability. In future work, we will first further expand our RCV watermarking scheme. The current scheme keeps the single-row and single-column statistical characteristics unchanged and extends to the multirow and multicolumn statistical characteristics unchanged. We will further study the characteristics of the high-dimensional mean vector and covariance matrix in order to design a more versatile database watermarking scheme. Then, we consider the study of data lossless database watermarking technology, in order to meet the requirements of more stringent data accuracy scenarios.

Acknowledgements

This work was supported by the Natural Science Foundation of China (62072421, U2336206, 62102386, 62372423, and U20B2047), Fundamental Research Funds for the Central Universities (WK2100000041).

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Yelu Yu received her B.E. degree from the North China Electric Power University in 2021. She is currently a master's student at the University of Science and Technology of China. Her research interests include privacy and security in database.

Han Fang received his B.S. degree in 2016 from Nanjing University of Aeronautics and Astronautics (NUAA) and the Ph.D degree in 2021 from University of Science and Technology of China (USTC). Currently, he is a research fellow at School of Computing, National University of Singapore. His research interests include image watermarking, informa-

DOI: 10.52396/JUSTC-2023-0091
JUSTC, **2024**, 54(4): 0401

tion hiding, and adversarial machine learning.

Weiming Zhang received his M.S. degree and Ph.D. degree in 2002 and 2005, respectively, from the Zhengzhou Information Science and Technology Institute. Currently, he is a Professor with the School of Cyber Science and Technology, University of Science and Technology of China. His research interests include information hiding and multimedia security.

References

[1] Xu J, Chen H, Yang X, et al. Verifiable image revision from chameleon hashes. *Cybersecurity*, **2021**, *4*: 34.

[2] Yuan G, Hao Q. Digital watermarking secure scheme for remote sensing image protection. *China Communications*, **2020**, *17*: 88–98.

[3] Sun J, Jiang X, Liu J, et al. An anti-recompression video watermarking algorithm in bitstream domain. *Tsinghua Science and Technology*, **2020**, *26*: 154–162.

[4] Munir R, Harlili. A secure fragile video watermarking algorithm for content authentication based on Arnold cat map. In: 2019 4th International Conference on Information Technology (InCIT). Bangkok, Thailand: IEEE, **2019**: 32–37.

[5] Wang F, Zhou H, Fang H, et al. Deep 3D mesh watermarking with self-adaptive robustness. *Cybersecurity*, **2022**, *5*: 24.

[6] Hamidi M, Haziti M E, Cherifi H, et al. A robust blind 3-D mesh watermarking based on wavelet transform for copyright protection. In: 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). Fez, Morocco: IEEE, **2017**: 1–6.

[7] Hou J U, Kim D G, Lee H K. Blind 3D mesh watermarking for 3D printed model by analyzing layering artifact. *IEEE Transactions on Information Forensics and Security*, **2017**, *12*: 2712–2725.

[8] Duy T P, Tran D, Ma W. An intelligent learning-based watermarking scheme for outsourced biomedical time series data. In: 2017 International Joint Conference on Neural Networks (IJCNN). Anchorage, AK, USA: IEEE, **2017**: 4408–4415.

[9] Kaur S, Singhal R, Farooq O, et al. Digital watermarking of ECG data for secure wireless commuication. In: 2010 International Conference on Recent Trends in Information, Telecommunication and Computing. Kerala, India: IEEE, **2010**: 140–144.

[10] Edward Jero S, Ramu P, Swaminathan R. Imperceptibility-Robustness tradeoff studies for ECG steganography using Continuous Ant Colony Optimization. *Expert Systems With Applications*, **2016**, *49*: 123–135.

[11] Agrawal R, Kiernan J. Watermarking relational databases. In: Proceedings of the 28th international conference on Very Large Data Bases. New York: ACM, **2002**: 155–166.

[12] Guo F, Wang J, Li D. Fingerprinting relational databases. In: Proceedings of the 2006 ACM symposium on Applied computing. New York: ACM, **2006**: 487–492.

[13] Guo F, Wang J, Zhang Z, et al. An improved algorithm to watermark numeric relational data. In: Proceedings of the 6th international conference on Information Security Applications. New York: ACM, **2005**: 138–149.

[14] Franco-Contreras J, Coatrieux G. Robust watermarking of relational databases with ontology-guided distortion control. *IEEE Transactions on Information Forensics and Security*, **2015**, *10*: 1939–1952.

[15] Sion R, Atallah M, Prabhakar S. Rights protection for relational data. In: Proceedings of the 2003 ACM SIGMOD international conference on Management of data, New York: ACM, **2003:** 98–109.

[16] Shehab M, Bertino E, Ghafoor A. Watermarking relational databases using optimization-based techniques. *IEEE Transactions on Knowledge and Data Engineering*, **2008**, *20*: 116–129.

[17] Gross-Amblard D. Query-preserving watermarking of relational databases and XML documents. In: Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. New York: ACM, **2003**: 191-201.

[18] Zhang Y, Yang B, Niu X. Reversible watermarking for relational database authentication. **2006**, *Journal of Computer, 17* (2): 59–65.

[19] Gupta G, Pieprzyk J. Reversible and blind database watermarking using difference expansion. *International Journal of Digital Crime and Forensics*, **2009**, *1*: 42–54.

[20] Jawad K, Khan A. Genetic algorithm and difference expansion based reversible watermarking for relational databases. *Journal of Systems and Software*, **2013**, *86*: 2742–2753.

[21] Hu D, Zhao D, Zheng S. A new robust approach for reversible database watermarking with distortion control. *IEEE Transactions on Knowledge and Data Engineering*, **2019**, *31*: 1024–1037.

[22] Imamoglu M B, Ulutas M, Ulutas G. A new reversible database watermarking approach with firefly optimization algorithm. *Mathematical Problems in Engineering*, **2017**, *2017*: 1387375.

[23] Farfoura M E, Horng S J, Wang X. A novel blind reversible method for watermarking relational databases. *Journal of the Chinese Institute of Engineers*, **2013**, *36*: 87–97.

[24] Iftikhar S, Kamran M, Anwar Z. RRW —a robust and reversible watermarking technique for relational data. *IEEE Transactions on Knowledge and Data Engineering*, **2015**, *27*: 1132–1145.

[25] Li Y, Wang J, Jia H. A robust and reversible watermarking algorithm for a relational database based on continuous columns in histogram. *Mathematics*, **2020**, *8*: 1994.

[26] Li Y, Wang J, Luo X. A reversible database watermarking method non-redundancy shifting-based histogram gaps. *International Journal of Distributed Sensor Networks*, **2020**, *16*: 1550147720921769.

[27] Tang X, Cao Z, Dong X, et al. PKMark: A robust zero-distortion blind reversible scheme for watermarking relational databases. In: 2021 IEEE 15th International Conference on Big Data Science and Engineering (BigDataSE). Shenyang, China: IEEE, **2021**: 72–79.

[28] Ge C, Sun J, Sun Y, et al. Reversible database watermarking based on random forest and genetic algorithm. In: 2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC). Chongqing, China: IEEE, **2020**: 239–247.

[29] Wang W, Liu C, Wang Z, et al. FBIPT: A new robust reversible database watermarking technique based on position tuples. In: 2022 4th International Conference on Data Intelligence and Security (ICDIS). Shenzhen, China: IEEE, **2022**: 67–74.

[30] Uniform Singapore Energy Price and Demand Forecast, **2022**. https://www.emcsg.com/MarketData/PriceInformation#priceDataView. Accessed March, 01, 2023

DOI: 10.52396/JUSTC-2023-0091
JUSTC, **2024**, 54(4): 0401