International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems

Watermarking Numerical Data in the Presence of Noise

Francesc Sebé, Josep Domingo-Ferrer, Jordi Castellà-Roca
Rovira i Virgili University of Tarragona, Department of Computer Engineering and Mathematics, Av. Països Catalans 26, 43007 Tarragona, Catalonia, Spain.
{*francesc.sebe,josep.domingo,jordi.castella*}*@urv.cat*

Received (received date)
Revised (revised date)

Data mining aims at extracting knowledge from data. Sometimes this requires the data owner to make data available to the data analyst. Unless the analyst is trusted by the data owner, the latter may wish to have his intellectual property rights protected. Watermarking is a way to provide some protection, not only for multimedia data, but also for numerical data. It consists of imperceptibly embedding a secret mark into the data to be protected. The mark can later be used to resolve any subsequent disputes on data ownership. This paper presents a watermarking method that embeds a watermark into each attribute of a multivariate continuous numerical dataset. This watermarking method is shown to be robust against random noise addition attacks. Data quality is assured to the extent that the watermarked data nearly preserve the attribute means and the covariance matrix from the original dataset. Our proposal is the first known watermarking system for multivariate numerical datasets with such robustness and quality properties.

Keywords: Digital watermarking, Intellectual property protection, Numerical data, Noise addition attacks

1. Introduction

Watermarking [1] has been traditionally applied to multimedia content (*e.g.* still images, sound streams or video streams) in order to hide a message (*e.g.* the watermark) into the content.

Watermarking has a broad range of applications [2]. Intellectual property protection (IPR) is among the best-known ones: the watermark is used to prove ownership or to help tracing illegal copies of the protected data. In this application, the watermark contains owner-chosen data hidden by using a secret key [3,4].

Even though IPR is conceivable for any kind of content, watermarking has been seldom applied to data other than multimedia (*e.g.* numerical or alphanumerical data). The reason is not lack of need, for there are virtually no alternative IPR approaches for those data; rather, the hindrance seems to be that there are more constraints to care about when dealing with numerical or alphanumerical data than when dealing with multimedia. For example, a number of statistics must be preserved for watermarked numerical data to stay analytically useful; typical quality

requirements for numerical data are preservation of first and second-order moments for all attributes.

However, there are also similarities between numerical and multimedia watermarking: the watermark should be able to resist different kind of spiteful attacks such as additive noise, bit flipping, rounding, subsampling and so on [5].

The most important properties that a watermarking system must meet are:

(1) **Imperceptibility:** The alterations caused to the content by mark embedding must be imperceptible. For numerical or alphanumerical data, *quality* is a more appropriate term than imperceptibility: we look at preserving analytical validity of data rather than their perceptual properties. Data quality should not be reduced after watermarking, that is, statistical analyses relevant to the data users should not be substantially altered by the watermarking process.
(2) **Robustness:** This property measures how resistant the mark is against attacks aiming at removing it or making it unrecoverable. Ideal robustness is such that alterations necessary to remove the watermark by an attacker are so large that they cause the altered data to lose their quality, *i.e.* their analytical validity. We prove that our system is robust against random noise addition attacks.

Contrary to what happens for multimedia watermarking, the literature on data watermarking is very scarce. The seminal contribution by Agrawal *et al.* [6] proposed a watermarking system for databases which is able to resist a great number of attacks but does not preserve the average nor the variance values of the original data. Recently, in [7] a watermarking system for univariate continuous data was presented; in that system, a watermark was embedded into a univariate dataset (data vector) while preserving its first and second-order moments (mean and variance). The proposal was proven to be robust against noise addition attacks.

1.1. *Contribution and plan of this paper*

This paper extends the proposal [7] to multivariate datasets. What we deliver here is a watermarking method such that the first and second-order moments are nearly preserved *for all attributes*. In a multivariate context, this means preserving not only means and variances, but also the entire covariance matrix expressing the relationships between the various attributes. Covariance preservation implies preservation of Pearson's correlations as well.

Section 2 gives an overview of the watermarking system. Mathematical background is given in Section 3. The mark embedding procedure is detailed in Section 4. Robustness against noise addition is shown in Section 5. Section 6 contains experimental results. Section 7 is a conclusion.

2. Watermarking System Overview

We assume that all attributes to be watermarked in the dataset are numerical and continuous. We will denote by X an $n \times d$ matrix representing an original dataset

consisting of n records and d attributes. The objective of our watermarking system is to modify X into X' so that each of the d attributes of X' embeds a watermark K.

We will consider that the j-th attribute X'_j of X' embeds the watermark K if it satisfies

$$\frac{1}{n}\sum_{i=1}^{n} s_{ij}x'_{ij} > \frac{M_j}{2} \tag{1}$$

where $s_{ij} \in \{-1, 1\}$ denotes the i-th element of a pseudo-random sequence S_j. Sequence S_j is generated using a pseudo-random number generator G_j seeded with the watermark K. As to M_j, it is a positive real number that acts as a security parameter. We will discuss its selection later on.

The watermarking procedure should preserve the quality of data. As said above, we require the means and the covariance matrix to be nearly preserved.

In summary, we require the watermarked dataset X' constructed by our method to satisfy the following constraints:

$$\frac{1}{n}\sum_{i=1}^{n} s_{ij}x'_{ij} = M_j,\ \forall j \in (1 \dots d) \tag{2}$$

$$\bar{X}'_j \approx \bar{X}_j,\ \forall j \in (1 \dots d) \tag{3}$$

$$CovMatrix(X') \approx CovMatrix(X) \tag{4}$$

Note that Constraint (2) ensures that the watermark will be detected as described in Expression (1).

2.1. *Selecting M_j*

A watermarking system can fail in two ways. The first one is when the watermark recovery algorithm finds a mark in a non-marked dataset, *i.e. false positive*. The other possibility is not to find a watermark in a marked dataset, *i.e. false negative*.

Since our marking system is required to meet Constraint (2), the probability of a false negative is 0 provided that X' has not been tampered with.

This is not the case for the false positive case. It is very important to keep the false positive probability as low as possible. Otherwise, finding a watermark in a data set would not guarantee its ownership. The following lemma and corollary show how to choose M_j so that the probability of false positive is less than a pre-defined ϵ.

Lemma 1. *Given the j-th attribute X_j from a random dataset $X = \{x_{ij}\}$ and a pseudo-random binary sequence $S_j = \{s_{ij}\}$, with $s_{ij} = \{-1, 1\}$ and $p(s_{ij} = 1) = p(s_{ij} = -1) = 1/2$, it holds that*

$$\frac{1}{n}\sum_{i=1}^{n} s_{ij}x_{ij} \sim N\left(0, \frac{E[X_j^2]}{n}\right)$$

where $N(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2.

Proof. By the independence of X_j and S_j, we have $E[S_j X_j] = E[S_j]E[X_j]$. Now, since $E[S_j] = 0$, it holds that $E[S_j X_j] = 0$. Next,

$$Var[S_j X_j] = E[(S_j X_j)^2] - (E[S_j X_j])^2 = E[S_j^2 X_j^2] - (E[S_j]E[X_j])^2$$

Since $E[S_j] = 0$ and $s_{ij}^2 = 1$, we conclude that $Var[S_j X_j] = E[X_j^2]$.

Now, from the Central Limit Theorem, we have that

$$\frac{\frac{\sum_{i=1}^{n}(s_{ij}x_{ij})}{n} - E[S_j X_j]}{\sqrt{\frac{Var[S_j X_j]}{n}}} = \frac{\frac{\sum_{i=1}^{n}(s_{ij}x_{ij})}{n}}{\sqrt{\frac{E[X_j^2]}{n}}}$$

follows a $N(0,1)$ distribution. Thus,

$$\frac{1}{n}\sum_{i=1}^{n} s_{ij}x_{ij} \sim N\left(0, \frac{E[X_j^2]}{n}\right)$$

□

Corollary 1. *Given an attribute X_j from an original, unmarked random dataset X and a watermark K, taking*

$$M_j \geq \sqrt{\frac{2E[X_j^2]}{n\epsilon}}$$

ensures that the probability of false positive is at most ϵ, that is, $P\left[\frac{1}{n}\sum_{i=1}^{n} s_{ij}x_{ij} > \frac{M_j}{2}\right] \leq \epsilon$.

Proof. We have that $\frac{1}{n}\sum_{i=1}^{n} s_{ij}x_{ij} \sim N\left(0, \frac{E[X_j^2]}{n}\right)$. Since the normal distribution is symmetric and $M_j > 0$, we can use the Chebyshev inequality for symmetric distributions to bound the right tail of the above normal as

$$P\left[\frac{1}{n}\sum_{i=1}^{n} s_{ij}x_{ij} > \frac{M_j}{2}\right] \leq \frac{2E[X_j^2]}{M_j^2 n} \tag{5}$$

We now wish to upper-bound the right-hand side of Inequality (5) by ϵ, that is,

$$\frac{2E[X_j^2]}{M_j^2 n} \leq \epsilon$$

Solving for M_j yields the expression in the corollary. □

3. Mathematical Background

Next, we explain the mathematical background on which our watermarking system is based. Specifically, we discuss preservation of means, variances and covariances.

3.1. *Preserving means*

Let X and X' be two datasets with d common attributes and n records. Let x_{ij}, x'_{ij} be the values taken by the j-th attribute for the i-th record of X and X', respectively. Mean preservation for the j-th attribute can be expressed as:

$$\bar{X}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n} = \frac{\sum_{i=1}^{n} x'_{ij}}{n} = \bar{X}'_j$$

If this equation holds $\forall j \in (1 \dots d)$, then the means of X' match those of X.

3.2. *Preserving variances*

Let X and X' be two datasets with d common attributes and n records. Let x_{ij}, x'_{ij} be the values taken by the j-th attribute for the i-th record and $\bar{X}_j$, $\bar{X}'_j$ be the averages of the j-th attributes in X and X' respectively. Preserving attribute variances can be written as

$$\frac{\sum_{i=1}^{n}(x_{ij} - \bar{X}_j)^2}{n} = \frac{\sum_{i=1}^{n}(x'_{ij} - \bar{X}'_j)^2}{n}, \quad \forall j \in (1 \cdots d)$$

For the j-th attribute, the above is equivalent to

$$\frac{\sum_{i=1}^{n} {x_{ij}}^2}{n} - \bar{X}_j^2 = \frac{\sum_{i=1}^{n} {x'}_{ij}^{\,2}}{n} - \bar{X}_j'^2$$

From the previous expression, we can see that, if $\bar{X}_j = \bar{X}'_j$ (*i.e.* if first-order moments are preserved), and $\frac{\sum_{i=1}^{n} {x_{ij}}^2}{n} = \frac{\sum_{i=1}^{n} {x'_{ij}}^2}{n}$, $\forall j \in (1 \dots d)$, then the variance of corresponding j-th attributes in X and X' will be the same.

3.3. *Preserving covariances*

Considering the same datasets X and X' above, the preservation of the covariance between any pair $1 \leq j < k \leq d$ of attributes can be written as:

$$\frac{\sum_{i=1}^{n}(x_{ij} - \bar{X}_j)(x_{ik} - \bar{X}_k)}{n} = \frac{\sum_{i=1}^{n}(x'_{ij} - \bar{X}'_j)(x'_{ik} - \bar{X}'_k)}{n}$$

For any j, k, the above can be rewritten as

$$\frac{\sum_{i=1}^{n} x_{ij}x_{ik}}{n} - \bar{X}_j\bar{X}_k = \frac{\sum_{i=1}^{n} x'_{ij}x'_{ik}}{n} - \bar{X}'_j\bar{X}'_k$$

From the previous expression, we can see that if $\bar{X}_j = \bar{X}'_j$, $\forall j \in (1 \dots d)$ (first-order moments are preserved) and $\frac{\sum_{i=1}^{n} x_{ij}x_{ik}}{n} = \frac{\sum_{i=1}^{n} x'_{ij}x'_{ik}}{n}$, $\forall j, k$, $1 \leq j < k \leq d$, then the covariance between the j-th and k-th attributes in X will match the corresponding one in X'.

4. Mark Embedding Procedure

4.1. *Objective*

The objective of our mark embedding procedure is to construct a marked dataset X' from X so that its attributes embed the watermark K. As explained in Section 2, this can be expressed as

$$\frac{1}{n}\sum_{i=1}^{n} s_{ij}x'_{ij} = M_j,\ \forall j \in (1\dots d)$$

We require means to be preserved. As justified in Section 3.1, this is ensured if

$$\frac{\sum_{i=1}^{n} x_{ij}}{n} = \frac{\sum_{i=1}^{n} x'_{ij}}{n},\ \forall j \in (1\dots d)$$

Once means are preserved, as justified in Section 3.2 preserving variances is ensured if

$$\frac{\sum_{i=1}^{n} {x_{ij}}^2}{n} = \frac{\sum_{i=1}^{n} {x'_{ij}}^2}{n}\ \forall j \in (1\dots d)$$

Finally, Section 3.3 shows that once means are preserved, covariances are preserved if

$$\frac{\sum_{i=1}^{n} x_{ij}x_{ik}}{n} = \frac{\sum_{i=1}^{n} x'_{ij}x'_{ik}}{n},\ \ \forall i,j,\ \ 1 \le j < k \le d$$

4.2. *Optimization problem*

By combining the above preservation objectives with a relaxed version of Constraint (2), mark embedding can be regarded as a constrained optimization problem:

$$\min_{x'_{ij}} \sum_{j=1}^{d} \frac{1}{\sigma^2_{x_j}} \left(\frac{\sum_{i=1}^{n} x_{ij}}{n} - \frac{\sum_{i=1}^{n} x'_{ij}}{n} \right)^2 + \sum_{j=1}^{d} \frac{1}{(\sigma^2_{x_j})^2} \left(\frac{\sum_{i=1}^{n} {x_{ij}}^2}{n} - \frac{\sum_{i=1}^{n} {x'_{ij}}^2}{n} \right)^2 +$$

$$+ \sum_{1\le j<k\le d} \frac{1}{\sigma^2_{x_j}\sigma^2_{x_k}} \left(\frac{\sum_{i=1}^{n} x_{ij}x_{ik}}{n} - \frac{\sum_{i=1}^{n} x'_{ij}x'_{ik}}{n} \right)^2 \qquad (6)$$

subject to

$$M_j \le \frac{1}{n}\sum_{i=1}^{n} s_{ij}x'_{ij} \le (1+\xi)M_j,\ \forall j \in (1\dots d)$$

Parameter ξ has the role of relaxing the equality constraint (2), which is necessary for the above optimization problem to be feasible. Note that relaxation is such that $\frac{1}{n}\sum_{i=1}^{n} s_{ij}x'_{ij} = M'_j$ for $M'_j \ge M_j$. As will be discussed later, this does not decrease security.

We will denote by $\Delta(X, X')$ the result of evaluating the objective function (6) between datasets X and X'.

4.3. *Solving the optimization problem*

We start with the original dataset X and values $M_j, j \in (1 \dots d)$. We compute the initial X' as

$$x'_{ij} = x_{ij} + s_{ij}\left(M_j - \frac{\sum_{i'=1}^{n} s_{i'j}x_{i'j}}{n}\right)$$

Note that all the attributes X_j of X' contain the watermark. This is so because,

$$\frac{\sum_{i=1}^{n} s_{ij}x'_{ij}}{n} = \frac{\sum_{i=1}^{n} s_{ij}x_{ij}}{n} + M_j - \frac{\sum_{i'=1}^{n} s_{i'j}x_{i'j}}{n} = M_j,\ \forall j \in (1 \dots d)$$

However, there is no guarantee that the means and the covariance matrix of X' resemble those of X, let alone match them. To that end, we use the following hill-climbing heuristic:

(1) Compute $e_1 := \Delta(X, X')$ (Δ denotes the objective function).
(2) While **not end** loop
   (a) Randomly select two indexes i, j and randomly perturb value x'_{ij}.
   (b) If $M_j \leq \frac{1}{n}\sum_{i'=1}^{n} s_{i'j}x'_{i'j} \leq (1+\xi)M_j$ then
      i. Let $e_2 := e_1$.
      ii. Recompute $e_1 := \Delta(X, X')$ (X' has been modified).
      iii. If $e_1 \geq e_2$ then restore the previous value for x'_{ij} and set $e_1 := e_2$.
   (c) Else restore the previous value for x'_{ij}.
(3) Return the dataset X'.

The end condition for the above algorithms depends on a trade-off between data quality and computation time. From our experiments, we have found that the objective value Δ decreases asymptotically with time. In other words, Δ decreases quickly at the beginning, but, as time goes on, it decreases more and more slowly. Thus, even if a longer execution time yields a lower Δ, at some point it does not pay to spend any more computing time. Note that reaching $\Delta = 0$ would ensure a perfect preservation of means and the covariance matrix.

5. Robustness Against Noise Addition

Let us now study the effect of noise addition on watermark recovery.

5.1. *Probability of watermark removal*

Lemma 2. *Given a watermarked attribute X'_j and an altered version obtained through random noise addition $X''_j = X'_j + D$, where D is the noise, it holds that*

$$\frac{1}{n}\sum_{i=1}^{n} s_{ij}x''_{ij} \sim N\left(M'_j, \frac{E[D^2]}{n}\right)$$

where $M'_j = E[S_j X'_j] \geq M_j$.

Proof. First, $E[S_jX''_j] = E[S_j(X'_j + D)] = E[S_jX'_j + SD] = E[S_jX'_j] + E[S_jD] = M'_j + E[S_j]E[D]$. Now, since $E[S_j] = 0$ and the optimization procedure imposes $M'_j \geq M_j$, we conclude that $E[S_jX''_j] = M'_j \geq M_j$.

Second, $Var[S_jX''_j] = Var[S(X'_j + D)] = Var[S_jX'_j] + Var[S_jD]$. Since $Var[S_jX'_j] = 0$ (the watermarked version is fixed) we have $Var[S_jX''_j] = Var[S_jD]$.

Now,

$$Var[S_jD] = E[S_j^2D^2] - E[S_jD]^2 = E[D^2] - (E[S_j]E[D])^2 = E[D^2]$$

because $s_{ij}^2 = 1$ and $E[S_j] = 0$.

From the Central Limit Theorem, we have that

$$\frac{\frac{\sum_{i=1}^{n}(s_{ij}x''_{ij})}{n} - E[S_jX''_j]}{\sqrt{\frac{Var[S_jX''_j]}{n}}} = \frac{\frac{\sum_{i=1}^{n}(s_{ij}x''_{ij})}{n} - M'_j}{\sqrt{\frac{E[D^2]}{n}}}$$

follows a $N(0,1)$ distribution. Thus,

$$\frac{1}{n}\sum_{i=1}^{n} s_{ij}x''_{ij} \sim N\left(M'_j, \frac{E[D^2]}{n}\right)$$

where $M'_j \geq M_j$ □

Corollary 2. *Given a watermarked attribute X'_j and an altered version obtained through random noise addition $X''_j = X'_j + D$, where D is the noise, the probability of the attack succeeding in removing the watermark, $P\left[\frac{1}{n}\sum_{i=1}^{n} s_{ij}x''_{ij} < \frac{M_j}{2}\right]$, is at most the probability that a random variable following a $N(0,1)$ distribution takes a value greater than $\frac{M_j}{2}\sqrt{\frac{n}{E[D^2]}}$*

Proof. We have that $\frac{1}{n}\sum_{i=1}^{n} s_{ij}x''_{ij} \sim N\left(M'_j, \frac{E[D^2]}{n}\right)$, where $M'_j \geq M_j$. The worst case (weakest mark) is $M'_j = M_j$. In this case, we have

$$P\left[N\left(M_j, \frac{E[D^2]}{n}\right) < \frac{M_j}{2}\right] = P\left[N\left(0, \frac{E[D^2]}{n}\right) > \frac{M_j}{2}\right] =$$

$$P\left[N(0,1) > \frac{M_j}{2}\sqrt{\frac{n}{E[D^2]}}\right]$$ □

Note that large values for M_j reduce the probability that a random noise addition attack succeeds in removing the watermark K from data.

6. Experimental Results

6.1. *Test data set*

We used two of the reference test microdata sets from the European CASC project [8]:

- The "Census" dataset contains 1080 records with 13 numerical attributes. It was used in CASC and in [9,10,11,12,13].
- The "Tarragona" dataset contains 834 records with 13 numerical attributes. It was used in CASC and in [13,14].

6.2. *Information loss measures*

Data quality after watermarking was studied using the information loss measures proposed in [15]. More exactly, the mean variation was used to mesure information loss for means ($IL(\bar{X}, \bar{X}')$), variances ($IL(S, S')$) and covariances ($IL(V, V')$). The measure for covariances only considers nondiagonal components in the covariance matrix. Although not explicitly addressed by our marking system, data similarity between raw records $IL(X, X')$ was also measured. In this case, we use the measure proposed in [11].

Let X and X' be the original and the masked data sets. Let V and V' be the covariance matrices of X and X'. Let S_j denote the standard deviation of the j-th attribute of X.

Information loss for raw data is measured as

$$IL(X, X') = \frac{\sum_{j=1}^{d} \sum_{i=1}^{n} \frac{|x_{ij} - x'_{ij}|}{\sqrt{2} S_j}}{nd}$$

Information loss for means is measured as

$$IL(\bar{X}, \bar{X}') = \frac{\sum_{j=1}^{d} \frac{|\bar{X}_j - \bar{X}'_j|}{|\bar{X}_j|}}{d}$$

Information loss for variances is measured as

$$IL(S, S') = \frac{\sum_{j=1}^{d} \frac{|v_{jj} - v'_{jj}|}{|v_{jj}|}}{d}$$

and information loss for covariances is measured as

$$IL(V, V') = \frac{\sum_{j=1}^{d} \sum_{1 \le i < j} \frac{|v_{ij} - v'_{ij}|}{|v_{ij}|}}{\frac{d(d-1)}{2}}$$

6.3. *Numerical results on watermarked data quality*

We watermarked each of the 13 attributes of both test datasets with the following parameters. We considered several probabilities of false positive $\epsilon = 0.1$, $\epsilon = 0.05$ and $\epsilon = 0.025$. In all cases, a relaxation parameter $\xi = 0.025$ was taken. Table 1 shows the results.

From Table 1 we can observe that, in general, means, variances and covariances are very well preserved since their information loss measures are close to 0. A measure 0 would indicate perfect preservation. The actual measures are nonzero because minimization of the objective function in Expression (6) is heuristic. Thus,

Table 1. Information loss measures between the original data set X and its watermarked version X' for $\epsilon = 0.1$, $\epsilon = 0.05$ and $\epsilon = 0.025$.

Dataset	ϵ	$IL(X, X')$	$IL(\bar{X}, \bar{X}')$	$IL(S, S')$	$IL(V, V')$
"Census"	0.1	0.18140	0.00021	0.00469	0.01846
	0.05	0.24579	0.00244	0.04449	0.19066
	0.025	0.34971	0.00583	0.15890	0.32288
"Tarragona"	0.1	0.11285	0.01381	0.02608	0.01251
	0.05	0.16176	0.01981	0.05351	0.01807
	0.025	0.23094	0.02831	0.10862	0.02655

in general, we do not reach final zero values for the objective function: for instance, for the "Census" data set for $\epsilon = 0.1$ we got $\Delta(X, X') = 0.000658$, for $\epsilon = 0.05$ we got $\Delta(X, X') = 0.01335$, and for $\epsilon = 0.025$ we got $\Delta(X, X') = 0.088585$. Note that a stronger watermark (smaller ϵ) results in a watermarked data set of lower quality. This shows the tradeoff between robustness and imperceptibility that underlies watermarking.

6.4. *Numerical results on attacked data quality*

In this section we study the robustness of our watermarking system against Gaussian random noise addition attacks. The experiments were made on the watermarked datasets X' obtained in the previous section. For $j = 1$ to 13, we took each attribute j of the 13 watermarked attributes in X' and perturbed it by adding Gaussian noise $D_j \sim N(0, \sigma_j^2)$.

The variance σ_j^2 of the noise and the probability of watermark removal are related by Corollary 2 (note that, from the distribution of D_j, it holds that $E[D_j^2] = \sigma_j^2$). Thus, the experimental results below refer to the probability of mark removal rather than to the noise variance σ_j^2.

Table 2 shows information loss measures between the watermarked set X' (for $\epsilon = 0.1$) and its attacked version X'' through noise addition. Tables 3 and 4 show the same measures between the watermarked datasets generated for $\epsilon = 0.05$ and $\epsilon = 0.025$, respectively, and their corresponding attacked versions.

From Tables 2, 3 and 4 we can see that the distortion caused by noise addition is much larger than the distortion caused by mark embedding. For instance, for the "Census" data set, for X' with $\epsilon = 0.1$, the noise that must be added to get a probability of mark removal close to 0.1 causes $IL(X', X'') = 1.96077$. This value is 10.8 times $IL(X, X')$ (distortion caused during mark embedding). This relation stays similar for the watermarked version with $\epsilon = 0.05$ and $\epsilon = 0.025$. The larger the probability of watermark removal, the larger the distortion that must be inflicted to data. Choosing a watermark strength that reduces the quality of X' to the minimum acceptable analytical validity will ensure that any attacked version is useless.

Table 2. Information loss measures between the watermarked data set X' ($\epsilon = 0.1$) and its attacked version X''.

Dataset	$P[removal]$	$IL(X', X'')$	$IL(\bar{X}', \bar{X}'')$	$IL(S', S'')$	$IL(V', V'')$
Census	0.1	1.96077	0.05411	12.5939	8.02629
	0.2	2.98571	0.08239	29.2721	20.447
	0.3	4.79245	0.13226	75.5496	56.1438
	0.4	9.92052	0.27378	324.214	253.394
Tarragona	0.1	1.05051	0.12631	3.51227	0.23868
	0.2	1.60666	0.19318	8.19904	0.44565
	0.3	2.48301	0.298554	19.5565	0.90768
	0.4	5.46263	0.656819	94.5265	4.8314

Table 3. Information loss measures between the watermarked data set X' ($\epsilon = 0.05$) and its attacked version X''.

Dataset	$P[removal]$	$IL(X', X'')$	$IL(\bar{X}', \bar{X}'')$	$IL(S', S'')$	$IL(V', V'')$
Census	0.1	2.71835	0.07664	24.1751	12.6308
	0.2	4.13928	0.1167	56.1571	30.8348
	0.3	6.64317	0.18729	144.837	82.3354
	0.4	13.7509	0.38769	621.264	363.588
Tarragona	0.1	1.46615	0.1785	6.8297	0.4033
	0.2	2.24235	0.2730	15.9524	0.7864
	0.3	3.46545	0.4219	38.0648	1.9111
	0.4	7.62399	0.9283	184.057	10.1273

Noise addition perturbs variances and covariances as well as means (which were nearly preserved during mark embedding). However, means suffer less alteration than second-order statistics, because the added noise has zero mean.

Note. The reason that $IL(S, S')$ and $IL(V, V')$ are larger for "Census" than for "Tarragona" is that the variances and covariances in the former dataset are smaller (both measures are computed as mean variations rather than as mean absolute errors).

6.5. *On the influence of computation time*

The optimization procedure described in Section 4.2 is a hill-climbing heuristic. This means that it works iteratively by decreasing the objective function $\Delta(X, X')$ at each iteration. Time is of great importance, since the longer the procedure iterates, the lower values for the objective function are obtained leading to better data quality

Table 4. Information loss measures between the watermarked data set X' ($\epsilon = 0.025$) and its attacked version X''.

Dataset	$P[removal]$	$IL(X', X'')$	$IL(\bar{X}', \bar{X}'')$	$IL(S', S'')$	$IL(V', V'')$
Census	0.1	3.63468	0.108757	42.9366	10.5526
	0.2	5.5346	0.165607	99.7015	24.3429
	0.3	8.88253	0.265785	257.075	62.5518
	0.4	18.3862	0.550153	1102.45	267.562
Tarragona	0.1	2.0212	0.2523	12.9632	0.7057
	0.2	3.0912	0.3858	30.2912	1.6039
	0.3	4.7774	0.5963	72.299	4.1362
	0.4	10.5103	1.3120	349.688	21.698

preservation. In this section we show the influence of time on data quality.

Table 5 shows the evolution of the different data quality measures and the objective function during the execution of the hill-climbing procedure for the "Census" data set (parameter $\epsilon = 0.05$). Figure 1 shows this evolution graphically. It can be seen that the objective function $\Delta(X, X')$ decreases asymptotically (no significant improvement is obtained after some time, so computation can be stopped at that point).

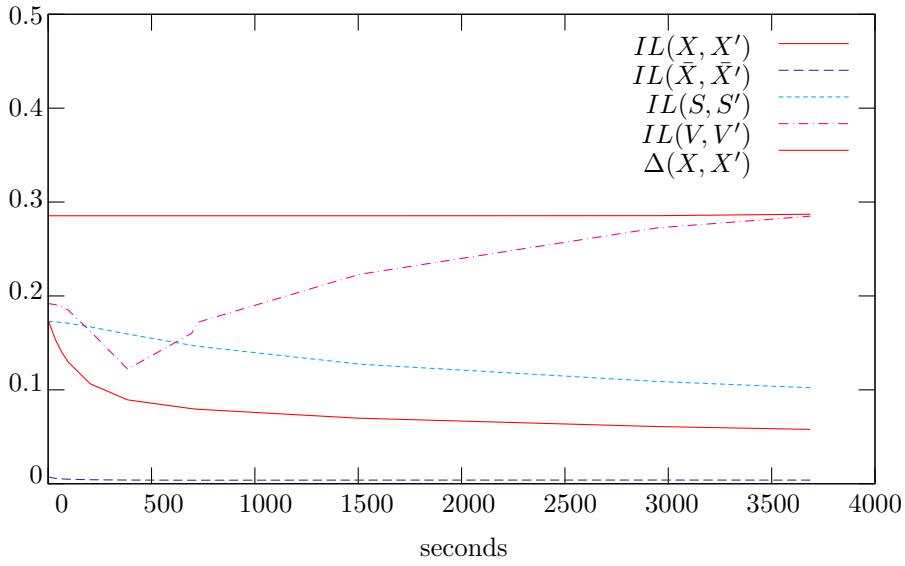
Table 5. Influence of time on data quality.

$Time(s)$	$IL(X, X')$	$IL(\bar{X}, \bar{X}')$	$IL(S, S')$	$IL(V, V')$	$\Delta(X, X')$
0	0.285416	0.00716441	0.172802	0.191979	0.174101
36	0.285415	0.00576381	0.172319	0.190643	0.152986
65	0.285415	0.00516061	0.171767	0.188635	0.140434
97	0.285415	0.0048284	0.170964	0.184893	0.129682
204	0.285414	0.00426026	0.167073	0.162366	0.106372
384	0.285414	0.00391139	0.159455	0.122031	0.089291
694	0.285414	0.00376451	0.147388	0.160358	0.079921
724	0.285414	0.00377415	0.146365	0.17204	0.079336
1505	0.285415	0.00386262	0.127393	0.222994	0.069759
2947	0.285515	0.00394272	0.10894	0.272527	0.060893
3686	0.287033	0.00384417	0.102284	0.28497	0.057879

7. Conclusions and Future Work

We have presented a watermarking system for multivariate continuous numerical data. Our system produces a watermarked data set that nearly preserves the means

Fig. 1. Influence of time on data quality



and the covariance matrix of the original data. An upper bound on the probability of watermark removal by noise addition has been derived. Empirical results show that our system is robust against random noise addition attacks in the sense that the distortion necessary to remove a watermark from data is so large that the attacked data become useless.

Future research will include:

- Refining the proposed system to withstand attacks different from noise addition, such as data permutation.
- Extending the system to non-numerical data types, that is, categorical attributes.

Acknowledgments

The authors are partly supported by the Spanish Ministry of Science and Education through project SEG2004-04352-C04-01 "PROPRIETAS", and by the Catalan government under grants 2002 SGR 00170 and 2005 SGR 00446.

References

1. S. Katzenbeisser and F. A. P. Petitcolas, *Information Hiding: techniques for steganography and digital watermarking* (Artech House, Norwood, 2000).
2. I. J. Cox, M. L. Miller, and J. A. Bloom, "Watermarking applications and their properties", *Proceedings of ITCC'2000* (IEEE Computer Society, 2000) 6–10.
3. I. Pitas and T. H. Kaskalis, "Applying signatures on digital images", *IEEE Workshop on Nonlinear Signal and Image Processing*, Thessaloniki, Greece (1995) 460–463.

4. G. C. Langelaar, J. C. A. VanderLubbe, and R. L. Lagendijk, "Robust labeling methods for copy protection of data", *Proceedings of SPIE 3022, Storage and Retrieval for Image and Video Databases V* (1997) 298–309.
5. S. Voloshynovskiy, S. Pereira, T. Pun, J. J. Eggers, and J. K. Su, "Attacks on digital watermarks: Classification, estimation-based attacks and benchmarks" *IEEE Communications Magazine*, **30**(8) (2001) 118–127.
6. R. Agrawal, P. J. Haas, and J. Kiernan, "Watermarking relational data: Framework, algorithms and analysis", *VLDB journal*, **12**(2) (2003) 157–169.
7. F. Sebé, J. Domingo-Ferrer and A. Solanas, "Noise-robust watermarking for numerical datasets", *Lecture Notes in Computer Science*, **3558** (2005) 134–143.
8. R. Brand, J. Domingo-Ferrer and J. M. Mateo-Sanz, "Reference data sets to test and compare SDC methods for protection of numerical microdata", European Project IST-2000-25069 CASC, http://neon.vb.cbs.nl/casc (2002).
9. J. Domingo-Ferrer, J. M. Mateo-Sanz and V. Torra, "Comparing SDC methods for microdata on the basis of information loss and disclosure risk", in *Pre-proceedings of ETK-NTTS'2001 (vol. 2)*, Luxemburg: Eurostat (2001) 807–826.
10. R. Dandekar, J. Domingo-Ferrer and F. Sebé, "LHS-based hybrid microdata vs rank swapping and microaggregation for numeric microdata protection", in *Lecture Notes in Computer Science*, **2316** (2002) 153–162.
11. W. E. Yancey, W. E. Winkler and R. H. Creecy, "Disclosure Risk Assessment in Perturbative Microdata Protection", *Lecture Notes in Computer Science*, **2316** (2002) 135–152.
12. J. Domingo-Ferrer and V. Torra, "Ordinal, continuous and heterogenerous k-anonymity through microaggregation", *Data Mining and Knowledge Discovery* **11**(2) (2005) 195–212.
13. M. Laszlo and S. Mukherjee, "Minimum spanning tree partitioning algorithm for microaggregation", in *IEEE Transactions on Knowledge and Data Engineering* **17**(7) (2005) 902–911.
14. J. Domingo-Ferrer and J. M. Mateo-Sanz, "Practical data-oriented microaggregation for statistical disclosure control", in *IEEE Transactions on Knowledge and Data Engineering* **14**(1) (2002) 189–201.
15. J. Domingo-Ferrer and V. Torra, "A quantitative comparison of disclosure control methods for microdata", in *Confidentiality, Disclosure and Data Access*, eds. P. Doyle, J. Lane, J. Theeuwes and L. Zayatz. (Elsevier, Amsterdam, 2002) 113-135.