Noise-Robust Watermarking for Numerical Datasets

Francesc Sebé, Josep Domingo-Ferrer, and Agustí Solanas

Rovira i Virgili University of Tarragona
Dept. of Computer Engineering and Maths
Av. Països Catalans 26, E-43007 Tarragona, Catalonia
{francesc.sebe,josep.domingo,agusti.solanas}@urv.net

Abstract. Watermarking has been used on multimedia for a variety of applications, among which intellectual property protection stands out. However, and in spite of a growing need, very few attempts have been made at using watermarking to protect the intellectual property of alphanumerical databases. We present in this paper a watermarking system for numerical databases, which is the first one of its kind to preserve means and variances of attributes in the database. Given a watermarked dataset and a candidate watermark, the recovery algorithm makes a decision whether the candidate watermark is embedded or not in the dataset. The probabilities of false positive and false negative can be made arbitrarily small by proper choice of a security parameter. We give an analytical expression for the information or data utility loss caused by watermark embedding. The proposed system can be made arbitrarily robust against noise addition. We also give empirical results showing a noise addition attack trying to remove the watermark must cause a distortion (and thus a loss of data utility) much larger than the watermark itself.

Keywords: Database watermarking, Watermark recovery decision, Intellectual property protection.

1 Introduction

Watermarking techniques have been widely used on multimedia data in order to transparently embed messages by including hidden information (*e.g.* the watermark) into the cover data (*e.g.* still images, sound streams or video streams)[KP00] Multimedia watermarking has a variety of uses [CMB00], the most typical being intellectual property protection: to prove ownership or to help tracing illegal copies of the cover data. Such is the case when the watermark contains a digital signature hidden by using a secret key[PK95][LVL97].

In contrast to its wide use over multimedia data, watermarking has found little application for protecting the copyright of numerical data or any kind of alphanumerical data, for that matter. The reason is not lack of need, for there are virtually no alternative approaches; rather, it might be that there are more constraints to care about when dealing with (numerical or alphanumerical) data than when dealing with multimedia. Specifically for numerical data, there are two

V. Torra et al. (Eds.): MDAI 2005, LNAI 3558, pp. 134–143, 2005.

main differences with multimedia that prevent straightforward use of multimedia watermarking techniques:

1. A number of statistics must be preserved for the watermarked data to stay analytically useful; at the very least, averages and variances of attributes should be preserved.
2. Numerical data are position-independent, that is, there is no clear connection between a numerical datum and its position in the dataset (unlike for an image, where a pixel value bears some relationship of similarity to the values of the surrounding pixels); in fact, a numerical dataset can be re-ordered and still be the same dataset.

Having pointed out the above differences, 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 attacks, bit flipping attacks, rounding attacks, subsampling attacks an so on[VPP+01].

The literature on data watermarking is very scarce: the main contribution is [AJK03] by Agrawal *et al.*, who proposed a watermarking system over databases which was able to resist a great number of attacks but did not preserve the average nor the variance values of the original data.

1.1 Contribution and Plan of This Paper

In this paper we tackle the problem of inserting a watermark into a numerical dataset while preserving the average and the variance of the original data. At the same time, we prove that the proposed watermarking system is robust against additive noise attacks. To do that, we present a method for *deciding* whether a watermark is embedded or not into a data vector. The probabilities of false positive and false negative decision can be made arbitrarily small. Also, the system can be made arbitrarily robust against noise addition.

The paper is organized as follows. Section 2 gives an overview of our watermarking system. Section 3 presents the proposed algorithms while defining the way in which their parameters must be computed to meet the aforementioned constraints. In Section 4 the robustness of the proposed algorithms against additive noise attacks is proven. Empirical results are given in Section 5. Finally, conclusions and future work are summarized in Section 6.

2 Watermarking System Overview

We assume that all attributes to be watermarked in the dataset are numerical and continuous.

2.1 Algorithms

In our proposal, a watermarking system is composed of two algorithms:

– *Mark embedding:* This is an algorithm taking as input the original unmarked dataset X, the watermark K and a positive number M which is a security

parameter. Its output is a dataset X' embedding K that is slightly different from X.

$$Embed(X, K, M) \rightarrow X'$$

– *Mark recovery:* This is an algorithm taking as input a dataset $\hat{X}$ and a watermark K. Its output is a value $\hat{M}$ whose interpretation is: if $\hat{M} > \frac{M}{2}$ we decide that K is embedded into $\hat{X}$; otherwise we decide that it is not.

$$Recover(\hat{X}, K) \rightarrow \hat{M}$$

$$\begin{cases} If\ \hat{M} > \frac{M}{2} \rightarrow K\ is\ in\ \hat{X} \\ If\ \hat{M} \leq \frac{M}{2} \rightarrow K\ is\ not\ in\ \hat{X} \end{cases}$$

2.2 Properties

We require our watermarking system to meet the properties below.

– *Imperceptibility:* The watermarked dataset X' should be similar to the original X in order to stay useful. This means that its statistical properties should be preserved as much as possible. In our system we ensure that the mean and variance of all attributes are preserved, *i.e.* $\overline{X'_i} = \overline{X_i}$ and $S^2_{X'_i} = S^2_{X_i}$, for all pairs of corresponding attributes (X_i, X'_i), where X_i is in X and X'_i in X'.
– *Low false positive probability:* The probability of recovering a watermark K from a dataset X which does not embed K should be low. In our system, this means that

$$P\left[Recover(X, K) > \frac{M}{2}\right] < \epsilon$$

where ϵ can be made arbitrarily small by choosing an appropiate parameter M.
– *Correctness:* Given a watermarked dataset X', $Recover(X', K)$ should always return a value greater than $\frac{M}{2}$. In our system $Recover(X', K)$ always returns M.
– *Robustness:* Given a watermarked dataset X', obtaining an attacked dataset X'' so that $Recover(X'', K) < \frac{M}{2}$ without knowledge of K implies a high degradation on the quality of X''. In this paper, we focus on noise addition attacks and we prove that, for a noise addition attack to have a non-negligible probability of success, the mean square error between X' and X'' is much higher than the one between X and X'.

3 Our Watermarking System

For the sake of clarity, we will describe our system in reverse order. First, the mark recovery algorithm will be specified. Understanding how the mark is recovered gives insight on how the mark should be embedded, which makes the subsequent mark embedding algorithm easier to understand.

3.1 Mark Recovery

As previously mentioned, recovery takes two parameters as input: a dataset $\hat{X}$ and the watermark K. Without loss of generality, we assume the dataset to consist of a single attribute, that is, $\hat{X} = \{\hat{x}_1, \ldots, \hat{x}_n\}$, where $\hat{x}_i$ is a scalar. If the dataset has several attributes, mark recovery (and embedding) is independently performed for each attribute.

The algorithm is next detailed:

Algorithm 1 (Recover($\hat{X}, K$))

1. *Generate a binary sequence* $S = \{s_1, \ldots, s_n\}$, $s_i = \{-1, 1\}$, $p(s_i = 1) = p(s_i = -1) = 1/2$ *using a pseudo-random generator* G *seeded by* K.
2. *Compute* $\hat{M}$ *as*
$$\hat{M} = \frac{1}{n} \sum_{i=1}^{n} s_i \hat{x}_i$$
3. *Return* $\hat{M}$.

As previously said, we decide that K is embedded into $\hat{X}$ if $\hat{M} > \frac{M}{2}$. We wish our system to have a low false positive probability. The following lemma and corollary address this subject.

Lemma 1. *Given a random dataset* $X = \{x_i\}$ *and a pseudo-random binary sequence* $S = \{s_i\}$, *with* $s_i = \{-1, 1\}$ *and* $p(s_i = 1) = p(s_i = -1) = 1/2$, *it holds that*
$$\frac{1}{n} \sum_{i=1}^{n} s_i x_i \sim N\left(0, \frac{E[X^2]}{n}\right)$$
where $N(\mu, \sigma^2)$ *denotes a Gaussian distribution with* μ *mean and* σ^2 *variance.*

Proof: By the independence of X and S, we have $E[SX] = E[S]E[X]$. Now, since $E[S] = 0$, it holds that $E[SX] = 0$

Next, $Var[SX] = E[(SX)^2] - (E[SX])^2 = E[S^2X^2] - (E[S]E[X])^2$, since $E[S] = 0$ and $s_i^2 = 1$, we conclude that $Var[SX] = E[X^2]$.

Now, from the Central Limit Theorem, we have that
$$\frac{\sum_{i=1}^{n}(s_i x_i) - nE(SX)}{\sqrt{Var[SX]n}} = \frac{\frac{\sum_{i=1}^{n}(s_i x_i)}{n} - E(SX)}{\sqrt{\frac{Var[SX]}{n}}} = \frac{\frac{\sum_{i=1}^{n}(s_i x_i)}{n}}{\sqrt{\frac{E[X^2]}{n}}}$$
follows a $N(0, 1)$ distribution. Thus,
$$\frac{1}{n} \sum_{i=1}^{n} s_i x_i \sim N\left(0, \frac{E[X^2]}{n}\right)$$
□

Corollary 1. *Given a random dataset X and a watermark K, the probability $P\left[Recover(X,K) > \frac{M}{2}\right]$ can be made arbitrarily small by increasing M.*

Proof. Since $\frac{1}{n}\sum_{i=1}^{n} s_i x_i \sim N\left(0, \frac{E[X^2]}{n}\right)$ and $M > 0$, from the Chebyshev bound we have that

$$P\left[\frac{1}{n}\sum_{i=1}^{n} s_i x_i > \frac{M}{2}\right] \leq \frac{2E[X^2]}{M^2 n}$$

It can be seen that this upper bound can be made arbitrarily small by increasing M. □

3.2 Mark Embedding

Next we describe the mark embedding algorithm. It takes as input a dataset $X = \{x_1, \ldots, x_n\}$, a secret key K and a security parameter M. It generates as output a dataset X' meeting $Recover(X', K) = M$

Algorithm 2 (Embed(X, K, M))

1. *Generate a binary sequence $S = \{s_1, \ldots, s_n\}$, $s_i = \{-1, 1\}$, $p(s_i = 1) = p(s_i = -1) = 1/2$ using a pseudo-random generator G seeded by K.*
2. *Using a random seed, generate a pseudo-random sequence $T = \{t_1, \ldots, t_n\}$ whose elements t_i follow a Gaussian $N(0,1)$ distribution*
3. *Let us denote by $X' = \{x'_1, \ldots, x'_n\}$ the resulting watermarked dataset. Its elements are computed as*

$$x'_i = ax_i + b + s_i|t_i|\lambda$$

where $|\cdot|$ denotes the absolute value operator.

Next, we describe how to choose parameters a, b and λ.

Parameter Choice. Our objective is to embed K into X while meeting the imperceptibility and correctness properties. This is achieved by obtaining a watermarked dataset X' that is slightly different from X. To do that, we wish the following:

1. The mean of X should be preserved by X', that is,

$$\overline{X'} = \overline{X} \tag{1}$$

2. The variance of X should be preserved by X', that is,

$$S^2_{X'} = S^2_X \tag{2}$$

3. Correctness is met in the above sense; to ensure that, we force that $Recover(X', K) = M$. By construction of the recovery algorithm, this is equivalent to

$$\frac{1}{n}\sum_{i=1}^{n} s_i x'_i = M \tag{3}$$

In this way, we will choose parameters a, b, λ so that the aforementioned constraints are met.

According to Constraint (1) we must equal $\overline{X'}$ and $\overline{X}$. Now,

$$\overline{X'} = \frac{1}{n}\sum_{i=1}^{n} x'_i = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b + s_i|t_i|\lambda)$$

$$= \frac{a}{n}\sum_{i=1}^{n} x_i + b + \frac{\lambda}{n}\sum_{i=1}^{n} s_i|t_i| = a\overline{X} + b + \lambda\overline{S|T|}$$

Now equalling with $\overline{X}$, we obtain a first equation:

$$\overline{X} = a\overline{X} + b + \lambda\overline{S|T|} \tag{4}$$

According to Constraint 2 we must preserve the variance of X into X'. Now,

$$S^2_{X'} = S^2_{ax_i+b+s_i|t_i|\lambda} = a^2 S^2_X + \lambda^2 S^2_{S|T|}$$

By equalling with S^2_X, we obtain a second equation:

$$S^2_X = a^2 S^2_X + \lambda^2 S^2_{S|T|} \tag{5}$$

Finally, according to Constraint 3 we must force $Recover(X', K) = M$.

$$Recover(X', K) = \frac{1}{n}\sum_{i=1}^{n} s_i x'_i = \frac{1}{n}\sum_{i=1}^{n} s_i(ax_i + b + s_i|t_i|\lambda)$$

$$= \frac{a}{n}\sum_{i=1}^{n} s_i x_i + \frac{b}{n}\sum_{i=1}^{n} s_i + \frac{\lambda}{n}\sum_{i=1}^{n}|t_i| = a\overline{XS} + b\overline{S} + \lambda\overline{T}$$

By equalling the above expression to M, we obtain a third equation

$$M = a\overline{XS} + b\overline{S} + \lambda\overline{T} \tag{6}$$

We can now solve the system formed by Equations (4), (5) and (6) for a, b and λ.

3.3 Information Loss

By construction, the previous choice of a, b and λ guarantees that the mean and variance of X are preserved into X'. The information loss can be measured as the mean square error between elements of X and X', *i.e.*

$$E[(X - X')^2] = \frac{1}{n}\sum_{i=1}^{n}(x_i - x'_i)^2$$

We can rewrite $E[(X - X')^2]$ as

$$E[(X - aX - b - S|T|\lambda)^2] = E[((1-a)X - b - S|T|\lambda)^2]$$

$$= E[\big((1-a)^2X^2 + 2(a-1)(b + S|T|\lambda)X + (b + S|T|\lambda)^2\big)]$$

$$= E[\big((1-a)^2X^2 + 2(a-1)bX + 2(a-1)\lambda S|T|X + b^2 + 2b\lambda S|T| + S^2|T|^2\lambda^2\big)]$$

$$= (1-a)^2E[X^2] + 2(a-1)bE[X] + 2(a-1)\lambda E[T]E[X] + b^2 + 2b\lambda E[T] + \lambda^2 E[T^2]$$

Since $T \sim N(0,1)$, we have that $E(T) = 0$ and $E(T^2) = 1$. Thus, the information or data utility loss can be written as

$$E[(X - X')^2] = (1-a)^2E[X^2] + 2(a-1)bE[X] + b^2 + \lambda^2 \tag{7}$$

So far, the approach followed has been to compute a, b, λ from the system of equality constraints (4), (5) and (6) and then compute the information loss using Expression (7). If the information loss obtained is too high or reducing information loss has a very high priority, then one might sacrifice to some extent mean and variance preservation and also reduce robustness by allowing $Recover(X', K)$ to output a number less than M. With this alternative approach, one would compute a, b, λ minimizing Expression (7) subject to constraints for mean-variance preservation and robustness relaxed as follows:

- Change Equation (4) by just requiring that $\overline{X'} \in \{\overline{X} \pm \varepsilon_{\overline{X}}\}$ for some $\varepsilon_{\overline{X}} > 0$;
- Change Equation (5) by just requiring that $S^2_{X'} \in \{S^2_X \pm \varepsilon_{S^2_X}\}$ for some $\varepsilon_{S^2_X} > 0$;
- Change Equation (6) by just requiring that $\frac{1}{n}\sum_{i=1}^{n} s_i x'_i > M/2$

4 Robustness Against Noise Addition

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

4.1 Probability of Watermark Removal

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

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

Proof: First, $E[SX''] = E[S(X' + D)] = E[SX' + SD] = E[SX'] + E[SD] = M + E[S]E[D]$, since $E[S] = 0$, we conclude that $E[SX''] = M$.

Second, $Var[SX''] = Var[S(X' + D)] = Var[SX'] + Var[SD]$. Since $Var[SX'] = 0$ we have $Var[SX''] = Var[SD]$.

Now, $Var[SD] = E[S^2D^2] - E[SD]^2 = E[D^2] - (E[S]E[D])^2$. Since $E[S] = 0$, we obtain $Var[SD] = E[D^2]$.

From the Central Limit Theorem, we have that

$$\frac{\sum_{i=1}^{n}(s_i x_i'') - nE(SX'')}{\sqrt{Var[SX'']n}} = \frac{\frac{\sum_{i=1}^{n}(s_i x_i'')}{n} - E(SX'')}{\sqrt{\frac{Var[SX'']}{n}}} = \frac{\frac{\sum_{i=1}^{n}(s_i x_i'')}{n} - M}{\sqrt{\frac{E[D^2]}{n}}}$$

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

$$\frac{1}{n}\sum_{i=1}^{n} s_i x_i'' \sim N\left(M, \frac{E[D^2]}{n}\right)$$

□

Corollary 2. *Given a watermarked dataset X' and an altered version obtained through noise addition $X'' = X' + D$, where D is the noise, the probability $P\left[Recover(X'', K) < \frac{M}{2}\right]$ can be made arbitrarily small by increasing M.*

Proof. Since $\frac{1}{n}\sum_{i=1}^{n} s_i x_i'' \sim N\left(M, \frac{E[D^2]}{n}\right)$ and $M > 0$, from the Chebyshev bound we have that

$$P\left[\frac{1}{n}\sum_{i=1}^{n} s_i x_i'' < \frac{M}{2}\right] \leq \frac{2E[D^2]}{M^2 n}$$

It can be seen that this upper bound can be made arbitrarily small by increasing M. □

4.2 Information Loss

A noise addition attack introduces a distortion on the attacked dataset.

Given a watermarked dataset X' and its attacked version $X'' = X' + D$, we will measure its distortion as the mean square error between X' and X'',

$$\frac{1}{n}\sum_{i=1}^{n}(x_i' - x_i'')^2$$

Its average corresponds to $E[(X' - X'')^2]$. Thus, it is

$$E[(X' - X' - D)^2] = E[D^2]$$

5 Numerical Example

We have used a dataset extracted from the 1995 U.S. Current Population Survey extracted using the U. S. Bureau of the Census Data Extraction System. This dataset contains 1080 records and 13 numerical attributes. Among these attributes, we have used PTOTVAL (total personal income). The 1080 values that PTOTVAL takes over the 1080 records are our data vector X. Table 1 shows the performance of our system against a Gaussian noise addition attack $N(0, \sigma_D^2)$.

Table 1. Performance figures

M	$IL(X, X')$	$100{*}P[false\ positive]$	σ_D^2	$IL(X', X'')$	$100 * P[mark\ removal]$	$\frac{IL(X',X'')}{IL(X,X')}$
100	13693.117	30.85%	10^4	10022.434	0.00%	0.732
100	13693.117	30.85%	10^6	1002243.426	5.05%	73.193
100	13693.117	30.85%	10^8	100224342.622	43.64	7319.323
200	59125.869	16.11%	10^4	10022.434	0.00%	0.170
200	59125.869	16.11%	10^6	1002243.426	0.05%	16.951
200	59125.869	16.11%	10^8	100224342.622	37.45%	1695.101
300	137444.551	6.81%	10^4	10022.434	0.00%	0.073
300	137444.551	6.81%	10^6	1002243.426	0.00%	7.292
300	137444.551	6.81%	10^8	100224342.622	31.21%	729.198

We can clearly observe that the distortion produced by the noise necessary to remove the watermark is much greater than the one produced by the insertion of the watermark on the original data. As an extreme case, when the noise added to remove the watermark is 7319.323 times larger than the noise caused by watermark embedding, the attacker has only probability 0.4364 of success in removing the watermark.

6 Conclusion and Future Work

The literature on watermarking alphanumerical data is quite scarce. This is surprising, because there is an increasing demand for copyright protection of databases. We have presented a spread-spectrum watermarking technique for numerical data which has the interesting feature of preserving means and variances of attributes; to the best of our knowledge, ours is the first data watermarking algorithm to do so.

Based on the whole watermarked dataset, the recovery algorithm makes a decision whether a particular watermark K is embedded in the data. The probabilities of false positive and false negative can be made arbitrarily small by proper choice of the security parameter M, even after a noise addition attack.

An analytical expression of the information loss (distortion) caused by watermark embedding on the original data has been given. As mentioned above, robustness against noise addition attacks can be arbitrarily increased by tuning M. Empirical results have been presented showing that removing a watermark

via noise addition is only possible at the cost of completely damaging the utility of the attacked dataset.

The proposed system has been described for univariate datasets. In case of multivariate datasets, it can be independently applied to the various attributes.

Future research will involve:

- Refining the proposed system to withstand other attacks beyond noise addition and bit-flipping (which is a special case of noise addition). In particular, we aim at making correct watermark recovery decisions in presence of the attacks mentioned in [AJK03]: subsetting, mix-and-match, etc.
- 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".

References

[AJK03] R. Agrawal, P. J. Haas, and J. Kiernan. Watermarking relational data: Framework, algorithms and analysis. *VLDB journal*, vol. 12, no. 2, pp. 157-169, 2003.

[CMB00] I. J. Cox, M. L. Miller, and J. A. Bloom. Watermarking applications and their properties. In *Proceedings of ITCC'2000*, pp. 6–10. IEEE Computer Society, 2000.

[KP00] S. Katzenbeisser and F. A. P. Petitcolas. *Information Hiding: techniques for steganography and digital watermarking*. Artech House, 2000.

[LVL97] G. C. Langelaar, J. C. A. VanderLubbe, and R. L. Lagendijk. Robust labeling methods for copy protection of data. In *Proceedings of SPIE 3022, Storage and Retrieval for Image and Video Databases V*, pp. 298–309, 1997.

[PK95] I. Pitas and T. H. Kaskalis. Applying signatures on digital images. In *IEEE Workshop on Nonlinear Signal and Image Processing*, Thessaloniki, Greece, pp. 460–463, 1995.

[VPP+01] 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*, vol. 30, no. 8, pp. 118–127, 2001.