# Marcaje robusto de datos numéricos

Francesc Sebé, Josep Domingo-Ferrer, Agustí Solanas, Susana Bujalance
Universitat Rovira i Virgili.
Dept. Enginyeria Informàtica i Matemàtiques.
Av.Països Catalans, 26. E-43007. Tarragona

{francesc.sebe, josep.domingo}@urv.net
{agusti.solanas, susana.bujalance}@urv.net

## Resumen

La inserción de marcas de agua se ha aplicado sobre datos multimedia para múltiples fines, entre ellos la protección de la propiedad intelectual. Sin embargo, y a pesar de la creciente necesidad, apenas se ha estudiado el marcaje de conjuntos de datos numéricos. En este artículo presentamos un sistema de inserción de marcas de agua sobre conjuntos de datos numéricos, siendo éste el primer método que preserva la media y la variancia de los datos del conjunto original.

## 1. Introducción

Las técnicas de inserción de marcas de agua han tenido una extensa aplicación sobre datos multimedia con el objetivo de insertar, de forma transparente, mensajes que contienen información oculta (e.i. la marca de agua) en los datos (e.i. imágenes, sonidos o vídeos)[2]. Contrariamente a su extendido uso sobre datos multimedia, la inserción de marcas de agua apenas se ha estudiado para la protección de datos numéricos o alfanuméricos. Este hecho no se debe a la poca necesidad de protección en este campo, puesto que no contamos con métodos alternativos; sin embargo, la razón podría residir en la existencia de más restricciones de las que ocuparse cuando se trabaja con datos numéricos o alfanuméricos que cuando se aplica a datos multimedia. Concretamente los datos numéricos presentan dos diferencias básicas con los datos multimedia que no permiten usar los métodos de inserción de marcas de agua ya conocidos para multimedia para el marcaje de datos numéricos:

1. Una serie de propiedades estadísticas deben permanecer en los datos para que estos sean de utilidad: por lo menos las medias y variancias de los atributos deben conservarse.
2. Los datos numéricos están situados de forma independiente, es decir, no guardan relación con los datos que los rodean (contrariamente a las imágenes, donde los *píxeles* están estrechamente relacionados con los *píxeles* de su alrededor), en resumen, un conjunto de datos numéricos puede ser reordenado y aún así ser el mismo conjunto.

También existen ciertas similitudes entre los datos numéricos y multimedia: una marca de agua debe resistir diferentes tipos de ataques, como por ejemplo la adición de ruido, inversión de bits, ataques de redondeo, ataques de subconjunto y muchos más [3].

La literatura sobre la inserción de marcas de agua en conjuntos de datos numéricos es muy escasa: la contribución principal es [1] de Agrawal-Haas-Kiernan, donde se propone un sistema de marcaje para bases de datos que resiste un buen número de ataques pero no preserva los valores de la media y la variancia de los datos originales al tiempo que su resistencia al ruido es baja.

En este artículo tratamos el problema de la inserción de marcas de agua en datos numéricos conservando la media y la variancia de los datos originales. El método presentado es robusto a los ataques de inserción de ruido. El artículo se distribuye como sigue. La sección 2 da una descripción de nuestro sistema de marcaje. La

sección 3 presenta el sistema que proponemos a la vez que define la manera de calcular los parámetros para cumplir con las restricciones mencionadas. En la sección 4 se demuestra la robustez de los algoritmos propuestos ante ataques de inserción de ruido. Los resultados experimentales se muestran en la sección 5. Finalmente, las conclusiones y el trabajo futuro se resumen en la sección 6.

## 2. Descripción del sistema de marcaje

En lo que sigue, suponemos que los datos que serán marcados son numéricos y continuos.

### 2.1 Algoritmos

En nuestra propuesta, el sistema de marcaje mediante marcas de agua se compone de dos algoritmos:

- Inserción de la marca: Este algoritmo tiene como entradas el conjunto de datos sin marcar X, la marca de agua K y un número positivo M que es el parámetro de seguridad. Su salida es el conjunto de datos marcado $X'$ con K que es ligeramente diferente de X.

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

- Recuperación de la marca: Este algoritmo tiene como entradas un conjunto de datos $\hat{X}$ y la marca de agua K. Su salida es el valor $\hat{M}$ cuya interpretación es: si $\hat{M} > \frac{M}{2}$ decidimos que la marca K se encuentra en los datos $\hat{X}$ ; si no se verifica decidiremos que no se encuentra.

$$\text{Re}\,cuperar(\hat{X}, K) \rightarrow \hat{M}$$

### 2.2 Propiedades

Se pretende que nuestro sistema cumpla las siguientes propiedades.

- Imperceptibilidad: El conjunto marcado $X'$ debería ser similar al original X para que fuera útil. Sus propiedades estadísticas deberían preservarse tanto como fuera posible. En nuestro sistema nos aseguramos que la media y la variancia se preserve, i.e. $\overline{X'} = \overline{X}$ y $S^2_{X'} = S^2_X$.
- Baja probabilidad de falsos positivos: La probabilidad de recuperar una marca de agua K de un conjunto de datos X donde no ha sido insertada K debería ser baja. En nuestro sistema, esto significa que

$$P\left[\text{Re}\,cuperar(X, K) > \frac{M}{2}\right] < \varepsilon$$

donde $\varepsilon$ puede ser arbitrariamente pequeño eligiendo un valor apropiado del parámetro M.

- Corrección: Dado un conjunto de datos marcados $X'$, Recuperar($X'$,K) debería siempre retornar un valor mayor que $\frac{M}{2}$. En nuestro sistema Recuperar($X'$,K) siempre retorna M.
- Robustez: Dado un conjunto de datos marcado $X'$, obtener un conjunto de datos atacado $X''$ tal que Recuperar($X''$,K)< $\frac{M}{2}$ sin conocimiento de K implica una alta degradación de la calidad de $X''$. En este artículo, nos centramos en los ataques de adición de ruido y demostramos que, para que una adición de ruido tenga un impacto significativo, el error cuadrático medio entre $X'$ y $X''$ es mayor que el que hay entre X y $X'$.

## 3. Nuestro sistema de inserción de marcas de agua

Para una mayor claridad, describiremos nuestro método en orden inverso. Primero, se especificará el algoritmo de recuperación de la marca. Entender cómo se recupera la marca nos ayudará a entender cómo debería insertarse.

### 3.1 Recuperación de la marca

Tal como se ha mencionado, la recuperación toma dos parámetros como entrada: un conjunto de datos $\hat{X}$ y la marca de agua K. Sin pérdida de generalidad, podemos asumir que el conjunto de datos consiste en un solo atributo, que es $\hat{X}=\{\hat{x}_1,...,\hat{x}_n\}$, donde $\hat{x}_i$ es un escalar. Si el conjunto de datos tiene varios atributos, la recuperación de la marca (y la inserción) se lleva a cabo de la misma forma sobre cada uno de los atributos.

El algoritmo se explica detalladamente a continuación:

**Algoritmo 1** (*Recuperación* ( $\hat{X}$ , K))

1. Generar una secuencia binaria $S=\{s_1,...,s_n\}$, $s_i=\{-1,1\}$, $p(s_i=1)=p(s_i=-1)=1/2$ usando un generador pseudoaleatorio G inicializado con K.
2. Calcular $\hat{M}$ como $\hat{M}=\frac{1}{n}\sum_{i=1}^{n}s_i\hat{x}_i$
3. Retornar $\hat{M}$

Como hemos dicho anteriormente, decidimos que K está insertada en $\hat{X}$ si $\hat{M}>\frac{M}{2}$. Deseamos que nuestro sistema tenga una baja probabilidad de falsos positivos. El siguiente lema y corolario tratan esta cuestión.

**Lema 1.** Dado un conjunto aleatorio $X=\{x_i\}$ y una secuencia pseudoaleatoria $S=\{s_i\}$, con $s_i=\{-1,1\}$ y $p(s_i=1)=p(s_i=-1)=1/2$, se tiene que

$$\frac{1}{n}\sum_{i=1}^{n}s_ix_i \sim N\left(0,\frac{E[X^2]}{n}\right)$$

donde $N(\mu,\sigma^2)$ denota una distribución gausiana con media $\mu$ y variancia $\sigma^2$.

**Corolario 1.** *Dado un conjunto de datos aleatorio X y una marca de agua K, la probabilidad* $P\left[\text{Re}\,cuperar(X,K)>\frac{M}{2}\right]$ *se puede hacer arbitrariamente pequeña incrementando M.*

### 3.2 Inserción de la marca

A continuación describimos el algoritmo de inserción de la marca de agua. Este toma como entradas al conjunto de datos $X=\{x_1,...,x_n\}$, una clave secreta K y un parámetro de seguridad M. El algoritmo genera un conjunto de datos $X'$ que cumple que $\text{Re}\,cuperar(X',K)=M$

**Algoritmo 2** ($Insertar(X,K,M)$)

1. Generar una secuencia binaria $S=\{s_1,...,s_n\}$, $s_i=\{-1,1\}$, $p(s_i=1)=p(s_i=-1)=1/2$ usando un generador pseudoaleatorio G inicializado con K.
2. Usando el generador de números pseudoaleatorios ya inicializado, generamos una secuencia $T=\{t_1,...,t_n\}$ cuyos elementos siguen una distribución gausiana N(0,1)
3. Denotamos como $X'=\{x'_i,...,x'_n\}$ al conjunto de datos marcados. Sus elementos se calculan como

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

donde $|\cdot|$ es el operador de valor absoluto.

A continuación, describimos cómo elegir los parámetros $a$, $b$ y $\lambda$.

### 3.3 Elección de los parámetros

Nuestro objetivo es insertar K dentro de X teniendo en cuenta las propiedades de imperceptibilidad y corrección. Esto es, obtener un conjunto de datos marcado $X'$ que sea muy parecido a X. Para conseguir esto, debe cumplirse lo siguiente:

1. La media de X debe conservarse en $X'$, esto es,

$$\overline{X'} = \overline{X} \quad (1)$$

2. La variancia de X debe preservarse en $X'$, esto es,

$$S^2_{X'} = S^2_X \quad (2)$$

3. La corrección debe garantizarse en el sentido descrito previamente; para asegurarnos, forzaremos que $\text{Re}\,cuperar(X', K) = M$. Según la construcción del algoritmo de recuperación, esto es

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

De esta forma, elegimos los parámetros $a$, b, $\lambda$ forzándolos a mantener las restricciones anteriores.

De acuerdo a la restricción (1) podemos igualar $\overline{X'}$ y $\overline{X}$. Es decir,

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

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

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

Igualando con $\overline{X}$, obtenemos la primera ecuación:

$$\overline{X} = a\overline{X} + b + \lambda\overline{S|T|} \quad (4)$$

De acuerdo con la restricción (2) debemos preservar la variancia de X en $X'$. Por tanto,

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

Igualando a $S^2_X$ obtenemos la segunda ecuación:

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

Finalmente, de acuerdo con la restricción (3) debemos forzar $\text{Re}\,cuperar(X', K) = M$.

$$\text{Re}\,cuperar(X', K) = \frac{1}{n}\sum_{i=1}^{n} s_i x'_i =$$

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

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

$$a\overline{XS} + b\overline{S} + \lambda\overline{|T|}$$

Igualando la expresión anterior a M, obtenemos la tercera ecuación:

$$M = a\overline{XS} + b\overline{S} + \lambda\overline{|T|} \quad (6)$$

Podemos resolver el sistema de ecuaciones formado por (4), (5) y (6) para hallar $a$, $b$, $\lambda$.

### 3.4 Pérdida de información

La elección anterior de los parámetros $a$, $b$ y $\lambda$ garantiza que la media y la variancia de X se conservan en $X'$. La pérdida de información puede medirse como el error cuadrático medio entre los elementos de X y de $X'$, i.e.

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

Podemos escribir $E\left[(X - X')^2\right]$ como

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

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

$$(1-a)^2 E\left[X^2\right] +$$

$$2(a-1)bE\left[X\right] +$$

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

$$2b\lambda E\left[T\right] + \lambda^2 E\left[T^2\right]$$

Puesto que T ~ N(0,1), tenemos que $E(T)=0$ y $E(T^2)=1$. Entonces, la pérdida de información de los datos se podrá escribir como

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

## 4. Robustez contra la adición de ruido

A continuación estudiaremos el efecto que causa la adición de ruido en la recuperación de la marca.

**Lema 2.** Dado un conjunto de datos marcado $X'$ y una versión alterada mediante la adición de ruido $X''=X'+D$, donde D es el ruido, se tiene que

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

**Corolario 2.** *Dado un conjunto de datos marcado $X'$ y una versión alterada mediante la adición de ruido $X''=X'+D$, donde D es el ruido, la probabilidad $P\left[\text{Recuperar}(X'',K)<\frac{M}{2}\right]$ se vuelve arbitrariamente pequeña a medida que se incrementa M.*

## 5. Ejemplos numéricos

Hemos usado un conjunto de datos extraído de una encuesta realizada en 1995 sobre la población de EE.UU. La tabla 1 muestra la pérdida de información y la probabilidad de falso positivo para diferentes valores de M. La tabla 2 muestra la distorsión causada por un ataque de adición de ruido gausiano $N(0,\sigma_D^2)$ junto a la probabilidad de eliminar la marca y la relación entre la distorsión producida durante el ataque y aquella producida durante la inserción.

| M | $IL(X,X')$ | 100*P(Falso Positivo) |
|---|---|---|
| 100 | 13693.117 | 30.85% |
| 200 | 59125.869 | 16.11% |
| 300 | 137444.551 | 6.81% |

**Tabla 1.** Pérdida de información y probabilidad de falso positivo para distintos valores de M.

| M | $\sigma_D^2$ | $IL(X',X'')$ | 100* P(borrar) | $\frac{IL(X',X'')}{IL(X,X')}$ |
|---|---|---|---|---|
| 100 | $10^4$ | 10022 | 0.00% | 0.73 |
| 100 | $10^6$ | $10^6$ | 5.05% | 73.19 |
| 100 | $10^8$ | $10^8$ | 43.64 % | 7319.32 |
| 200 | $10^4$ | 10022 | 0.00% | 0.17 |
| 200 | $10^6$ | $10^6$ | 0.05% | 16.95 |
| 200 | $10^8$ | $10^8$ | 37.45% | 1695.10 |
| 300 | $10^4$ | 10022 | 0.00% | 0.07 |
| 300 | $10^6$ | $10^6$ | 0.00% | 7.29 |
| 300 | $10^8$ | $10^8$ | 31.21% | 729.19 |

**Tabla 2.** Pérdida de información y probabilidad de eliminar la marca para distintos valores de M y de $\sigma_D^2$.

Se puede observar claramente que la distorsión producida por el ruido necesario para destruir la marca es mucho mayor que la producida al insertar la marca en los datos originales.

## 6. Conclusiones y trabajo futuro

Se ha presentado un sistema robusto de marca de agua para datos numéricos que mantiene la media y la variancia de los datos originales. El método presentado es robusto al ruido. Temas abiertos para futuras investigaciones son:

- Mejorar la robustez del sistema haciéndolo resistente a más ataques como por ejemplo: ataques de subconjunto, reordenaciones, etc.
- Extender el sistema a datos no numéricos.

## Agradecimientos

Los autores están parcialmente subvencionados por el Ministerio de Educación y Ciencia a través del proyecto SEG2004-04352-C04-01 "PROPRIETAS".

## Referencias

[1] 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.

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

[3] S. Voloshynovskiy, S. Pereira, T. Pun, J.J. Eggers, and J.K. Su. Attacks on digital watermarks: Classification, estimation-based attacks and bench marks. IEEE Communications Magazine, vol. 30, no. 8, pp. 118–127, 2001.