Un nuevo sistema de marca de agua para bases de datos numéricas

Agustí Solanas, Josep Domingo-Ferrer, Francesc Sebé , Susana Bujalance

Departament d'Enginyeria Informàtica i Matemàtiques.
Universitat Rovira i Virgili
Av. Països Catalans, 26. 43007 Tarragona

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

Resumen Este artículo presenta un nuevo sistema de inserción de marcas de agua cuyo objetivo es prevenir la piratería de bases de datos numéricas. Este método es más robusto que los propuestos hasta el momento y permite que los datos marcados conserven la media y la desviación típica de los datos originales. El nuevo sistema combina el método propuesto por Agrawal-Haas-Kiernan con el sistema de marcaje llamado Expansión de Espectro, consiguiendo así un incremento en su resistencia al ruido. Se presentan también resultados experimentales comparando su robustez frente a la del sistema de Agrawal-Haas-Kiernan.

1 Introducción

Actualmente, la necesidad de evitar la copia ilegal de contenido digital se ha incrementado considerablemente. Se dispone de herramientas para elaborar copias perfectas sin ninguna degradación que indique que el documento (e.g. video, audio, imágenes, software, bases de datos) no es original; además dichas copias se pueden difundir fácilmente gracias a las nuevas tecnologías. Los distintos métodos de protección contra la copia de datos se agrupan en dos conjuntos:

1. **Protección del hardware/software**. Estos sistemas centran su atención en evitar que se efectúe la copia ilegal. Esto se consigue mediante restricciones introducidas en los datos y en los equipos de copia. Estos métodos se basan en la confianza de que nadie podrá manipular el sistema de protección. Esta suposición ha demostrado no ser realista.
2. **Inserción de marcas de agua**. Estos sistemas no se basan en evitar que se realice la copia. Su misión es, en caso de encontrarse una copia ilegal, poder demostrar los derechos de autor de los datos y que la copia es ilegal. Esto se consigue insertando mensajes ocultos entre los datos. Dichos mensajes pueden ser distintivos propios del producto, de su creador, del comprador, etc. Los objetivos más importantes que se plantean estos sistemas son: (1) evitar que los mensajes introducidos en los datos se puedan eliminar y (2) conseguir que los nuevos datos sigan siendo útiles a pesar de la distorsión causada durante la inserción del mensaje.

En este artículo se presenta un nuevo sistema de inserción de marcas de agua para la protección de bases de datos numéricas.

El artículo se estructura de la siguiente forma: En la sección 2 se presentan los trabajos previos más relevantes en el campo de las marcas de agua aplicadas a la protección de bases de datos. En la sección 3 se presenta nuestra nueva propuesta y los resultados experimentales se muestran en la sección 4. Finalmente, se presentan las conclusiones en la sección 5.

2 Trabajo previo

Durante años se ha estudiado a fondo el uso de marcas de agua aplicadas a la protección de datos multimedia. Se han diseñado sistemas para su uso en imágenes [PK95][RDB96][DVM97], sonido [Mor95], video [JV97][SST97], software [CT02], etc. Otras aplicaciones de las marcas de agua pueden encontrarse en [KP00]. Sin embargo, al considerar el caso particular de bases de datos, debemos tener en cuenta lo siguiente:

- Los datos pueden ser alfanuméricos y en consecuencia un pequeño cambio en su representación binaria puede afectar significativamente a su valor.
- La información que contiene una base de datos varía con frecuencia.
- Los datos no guardan una relación espacio-temporal y presentan una menor redundancia que los datos multimedia

2.1 Inserción de marcas de agua en bases de datos

El estudio del marcaje de bases de datos se inicia con el artículo de Agrawal-Haas-Kiernan [AHK03]. El artículo presenta un método que consiste en introducir una marca de agua en los bits de menor peso de un atributo de cada tupla seleccionada a partir de la clave primaria y una clave secreta.

El método presenta algunos puntos débiles:

- Si el atacante conoce el número de bits de menor peso donde se introduce la marca, éste podrá destruirla con facilidad.
- El sistema es poco resistente al ruido.

2.2 Sistemas de Expansión de Espectro

Este sistema fue diseñado para su utilización en la transmisión de información. Es la base de la técnica de acceso múltiple por división de código (CDMA) [Ver98]. El sistema presenta propiedades que lo hacen muy útil como método de inserción de marcas de agua (e.g. resistencia contra interferencias).

La primera aplicación de este sistema en el campo de la inserción de marcas de agua fue en [HG98], donde se propone un algoritmo para su aplicación al marcaje de video.

3 Nuestro sistema de marca de agua

Nuestro método está formado por dos algoritmos: el de inserción de la marca y el de recuperación de la misma.

A continuación se describe con más detalle cada uno de ellos. La notación utilizada de ahora en adelante se resume en la Tabla 1.

Tabla 1: Notación

υ	Conjunto de atributos que serán marcados
$1/\gamma_t$	Fracción de tuplas marcadas
$1/\gamma_a$	Fracción de atributos marcados de cada tupla
M	Matriz de inserción
G	Generador de valores pseudoaleatorios
s	Secuencia pseudoaleatoria de {-1,1}
K	Clave secreta
R	Una relación de la base de datos
r	Una tupla de R
W	La marca de agua

3.1 Inserción de la marca

El objetivo del algoritmo de inserción es el de introducir una marca en los datos de forma que ésta sea resistente a ataques que intenten eliminarla, distorsionando lo menos posible el valor de los datos además de permitir que estos conserven la media y la desviación típica de los datos originales. El algoritmo consta de tres etapas:

1. Seleccionar aquellas posiciones de la base de datos donde introducir la marca
2. Generar una secuencia pseudoaleatoria
3. Marcar los atributos seleccionados conservando la media y la desviación típica de los datos originales

3.1.1 Selección de las posiciones de inserción

Para llevar a cabo la selección, se ha diseñado una modificación del algoritmo propuesto en [AHK03] que permite obtener una matriz de inserción. El nuevo algoritmo inicializa un generador pseudoaleatorio mediante la clave primaria $r.P$ de cada tupla de una relación R y la clave secreta K que sólo conoce el propietario de la base de datos (línea 3 del Algoritmo 1). Una vez inicializado el generador, la función Siguiente(G) retorna un número pseudoaleatorio que depende del valor con que se haya inicializado el generador (líneas 4 y 6). Del total de tuplas η, se marca el porcentaje

correspondiente al valor indicado por γ_t (línea 4) y de cada tupla candidata se marca el porcentaje de atributos correspondiente a γ_a (línea 6). La ejecución del algoritmo da como resultado una matriz *M* de ceros y unos donde cada elemento de la matriz representa un elemento de la base de datos. Los unos corresponden a las posiciones de los valores que se marcarán. Los ceros indican aquellas posiciones que permanecerán inalteradas.

Algoritmo 1 Obtención de la matriz de inserción

```
1)  Función ObtenerMatrizInserción(K,R) retorna R
2)      para tupla r ∈ R hacer
3)          Inicializar G con r.P concatenado con K
4)          si (Siguiente(G) mod γ_t =0) entonces
5)              para cada atributo a∈υ hacer
6)                  si (Siguiente(G) mod γ_a =0) entonces
7)                  M_{r,a} = 1
8)                  sino
9)                  M_{r,a} = 0
10)                 fin_si
11)             fin_para
12)         fin_si
13)     fin_para
14)     retorna M
15) fin función ObtenerMatrizInserción
```

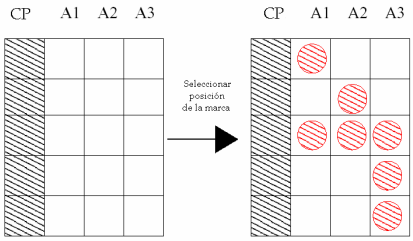

Figura 1: Esquema de selección de las posiciones de inserción de la marca de agua mediante el Algoritmo 1 y creación de la matriz de inserción. Los círculos de la derecha muestran los atributos donde se insertará la marca, su posición se corresponde con las posiciones de la matriz donde hay un 1.

3.1.2 Generación de una secuencia pseudoaleatoria

En esta etapa se genera una secuencia pseudoaleatoria *s* (ver Figura 2). Dicha secuencia se usa para insertar la marca de agua en la siguiente etapa (ver Figura 4). Se

pretende que la marca de agua sea resistente a la transposición de filas, por lo tanto la secuencia pseudoaleatoria *s* debe depender de un valor que permanezca invariante a la transposición. En este caso se utiliza la clave primaria *r.P* de cada tupla de la relación *R* concatenada con la clave secreta del usuario *K* para inicializar el generador *G* de la secuencia. La secuencia esta formada únicamente por una sucesión de {–1, 1} y contiene tantos términos como tuplas η tenga la relación.

Consideramos que la clave primaria de cada tupla permanece intacta tras cualquier ataque que sufran los datos, ya que su modificación inutilizaría los datos. En el caso que la base de datos no posea ninguna clave la manera de proceder sería la siguiente (Figura 3):

1. Utilizar el valor de los bits de mayor peso de cada elemento como clave primaria y concatenar este número con la clave secreta para obtener la secuencia pseudoaleatoria.
2. Insertar la marca únicamente en los bits de menor peso

Se considera que la alteración de los bits de mayor peso provoca una alteración inaceptable de los datos.

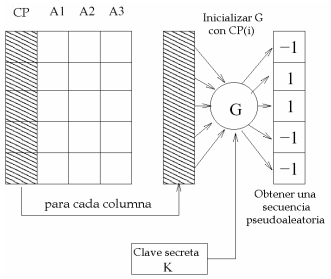

Figura 2: Generación de una secuencia pseudoaleatoria basada en los resultados de un generador de números pseudoaleatorios inicializado mediante la clave primaria de la tupla concatenada con la clave secreta del propietario de la base de datos

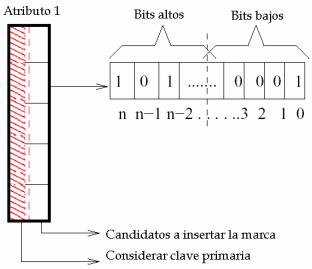

Figura 3: En el caso de una base de datos sin clave primaria, los bits de mayor peso de un atributo serían usados como tal.

3.1.3 Marcaje de los datos

El algoritmo de marcaje se usa de forma independiente para cada uno de los atributos de la relación. Una vez obtenida la matriz de posiciones *M* y generada la secuencia pseudoaleatoria *s* de {-1,1}, se inserta una marca *W* en cada uno de los elementos correspondientes $x_{r,a}$ de la relación *R* tales que $M_{r,a}=1$.

Se puede expresar como:

Marcar(R, s, K, M, W) R'

El marcaje que utilizaremos consiste en:

1. Para cada atributo a generar una secuencia pseudoaleatoria $Z_a=\{z_{1,a},...,z_{n_a,a}\}$ cuyos elementos sigan una distribución Gausiana N(0,1). Donde n_a es el número de elementos del atributo a tales que $M_{r,a}=1,\forall r\in R$
2. Siendo $Y_a=\{y_{1,a},...,y_{n_a,a}\}$ el conjunto de los valores de X_a tales que $M_{r,a}=1,\forall r\in R$, debemos obtener los elementos marcados $\{y'_{1,a},...,y'_{na,a}\}$ a partir de la siguiente fórmula:

$$y'_{i,a}=Ay_{i,a}+B+s_{i,a}\,|\,z_{i,a}\,|\,\lambda$$

donde |·| es el operador de valor absoluto, $s_{i,a}$ es la proyección de los elementos de *s* tal que $M_{r,a}=1,\forall r\in R$ y A, B, λ son coeficientes a determinar.

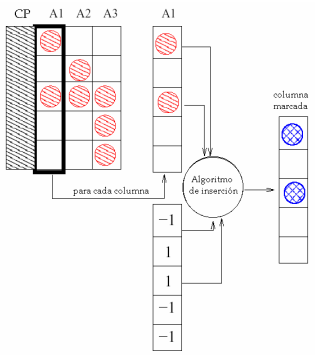

Figura 4: Esquema del algoritmo de marcaje. Para cada tupla, dadas las posiciones donde se insertará (matriz M) y la secuencia pseudoaleatoria de {-1,1}, el algoritmo introduce una marca en los correspondientes atributos.

Determinación de los parámetros

El objetivo es insertar la marca de agua W dentro de la base de datos R, consiguiendo que dicha marca sea imperceptible y que preserve ciertas propiedades. Se desea que:

1. La media de cada uno de los atributos X_a que se encuentran en R se conserve en R'. Esto se consigue exigiendo:

$$\overline{Y_a} = \overline{Y'_a} \quad (\mathbf{1})$$

2. La variancia de cada atributo *a* de R se conserve en R'. Esto se consigue exigiendo:

$$S^2_{Y_a} = S^2_{Y'_a} \quad (\mathbf{2})$$

3. Garantizar la correcta recuperación de la marca. Según el algoritmo Recuperar(·) descrito en la sección 3.2, se tiene que *Recuperar(* X_a *,K)=W*. Por la construcción del algoritmo de recuperación, el requisito de corrección se puede expresar mediante la siguiente restricción para cada tributo *a*

$$\frac{1}{n_a}\sum_{i=1}^{n_a} s_{i,a}\, y'_{i,a} = W \quad (\mathbf{3})$$

A continuación se halla A, B, λ de forma que se verifiquen las restricciones expuestas anteriormente. La media $\overline{Y'_a}$ del atributo *a* de la base de datos marcada R' se puede expresar como:

$$\overline{Y'_a} = \frac{1}{n_a}\sum_{i=1}^{n_a} y'_{i,a} = \frac{1}{n_a}\sum_{i=1}^{n_a} (A y_{i,a} + B + \mathrm{s_{i,a}} \mid \mathrm{z_{i,a}} \mid \lambda)$$

$$= \frac{A}{n_a}\sum_{i=1}^{n_a} y_{i,a} + B + \frac{\lambda}{n_a}\sum_{i=1}^{n_a} \mathrm{s_{i,a}} \mid \mathrm{z_{i,a}} \mid = A\overline{Y_a} + B + \lambda\overline{S_a|Z_a|}$$

Igualando lo obtenido con $\overline{Y_a}$, siguiendo la restricción (1), obtenemos que:

$$\overline{Y_a} = A\overline{Y_a} + B + \lambda\overline{S_a|Z_a|} \quad (\mathbf{4})$$

La variancia $S^2_{Y'_a}$ del atributo *a* pude ser expresada como:

$$\mathrm{S^2_{Y'_a}} = \mathrm{A^2 S^2_{Y_a}} + \lambda^2 \mathrm{S^2_{S_a|Z_a|}}$$

Igualando con $S^2_{Y_a}$, siguiendo la restricción (2), se da que:

$$S^2_{Y_a} = A^2 S^2_{Y_a} + \lambda^2 S^2_{S_a|Z_a|} \quad (5)$$

Finalmente, teniendo en cuenta la restricción (3) podemos forzar que *Recuperar(a,K)=W* para cada tributo *a*. El lado izquierdo de la restricción 3 puede expresarse como:

$$\frac{1}{n_a}\sum_{i=1}^{n_a} s_{i,a}\, y'_{i,a} = \frac{1}{n_a}\sum_{i=1}^{n_a} s_{i,a}\,(A y_{i,a} + B + s_{i,a}|z_{i,a}|\lambda) =$$

$$\frac{1}{n_a}\sum_{i=1}^{n_a} s_{i,a}\, y_{i,a} + \frac{B}{n_a}\sum_{i=1}^{n_a} s_{i,a} + \frac{\lambda}{n_a}\sum_{i=1}^{n_a}|z_{i,a}| = A\overline{Y_a S_a} + B\overline{S_a} + \lambda\overline{|Z_a|}$$

Igualando esta expresión a W, obtenemos:

$$W = A\overline{Y_a S_a} + B\overline{S_a} + \lambda\overline{|Z_a|} \quad (6)$$

Resolviendo el sistema que forman la ecuación (4), (5) y (6) podremos hallar el valor de A, B y λ de forma que los datos marcados verifiquen las propiedades exigidas.

3.2 Recuperación de la marca

El proceso de recuperación de la marca consiste en:

1. Obtener la matriz de inserción
2. Generar la secuencia pseudoaleatoria
3. Recuperar la marca de forma independiente para cada atributo de la relación

Únicamente puede aplicar este algoritmo de recuperación el propietario de la base de datos, ya que sólo él conoce la clave secreta. Los dos primeros pasos son idénticos a sus equivalentes en la inserción. En cambio, la etapa número tres se define como *Recuperar(a,K)* y se desarrolla como sigue:

1. Computar

$$\hat{W}_a = \frac{1}{n_a}\sum_{i=1}^{n_a} s_{i,a}\,\hat{y}_{i,a}$$

donde $\hat{y}_{i,a}$ es el elemento *i-ésimo* del atributo *a* de $\hat{R}$ con $M_{r,a} = 1$. Es necesario destacar que $\hat{R}$ no es necesariamente R' ya que la base de datos marcada puede haber sufrido algún ataque y por lo tanto se habrá modificado.

2. Retornar la marca $\hat{W}_a$ recuperada del atributo *a* de la base de datos $\hat{R}$.

Consideraremos que la marca W se ha recuperad del atributo *a* si la marca extraída $\hat{W}_a$ verifica : $\hat{W}_a > \frac{W}{2}$

En el caso que la base de datos $\widehat{R}$ fuese igual a R', es decir, no hubiera sufrido ningún ataque, entonces la marca se recuperaría en su totalidad, esto es, se encontraría la marca en todos y cada uno de los atributos. Sin embargo, para bases de datos donde se ha aplicado un ataque, es necesario un criterio que permita decidir si la marca W se encuentra en $\widehat{R}$, ya que se habrá obtenido una serie de atributos donde se considerará encontrada la marca y otro grupo de atributos donde la marca no se reconoce. Para tomar esta decisión debemos fijar un umbral. Este umbral se representa mediante un porcentaje mínimo de atributos donde debe encontrarse la marca (e.g. 75% de recuperación exitosa); si se recupera un menor porcentaje de marca se considera que ésta no está contenida en los datos.

4 Resultados experimentales

Se ha elaborado un análisis de la respuesta de nuestro sistema a los ataques de adición de ruido gausiano y transposición de tupas, y se ha comparado con la respuesta ofrecida por el método de Agrawal-Haas-Kiernan.

4.1 Adición de ruido gausiano

Este ataque consiste en añadir ruido gausiano al valor de los elementos a de la base de datos ya marcada R' con la finalidad de destruir la marca, es decir, hacerla indetectable por el sistema de recuperación. Si llamamos $R'(a)$ al valor del elemento a de la relación R', entonces el ataque se puede definir como:

$$R''(a) = R'(a) + N$$

Donde N sigue una distribución gausiana $N(\mu, \sigma^2)$.

Se considerará que el ataque ha tenido éxito cuando éste haya destruido la marca, es decir, ésta no se pueda recuperar de $R''(a)$. En nuestro caso, la marca no se recupera si:

$$\widehat{W}_a = \frac{1}{n_a} \sum_{i=1}^{n_a} s_{i,a}\, y''_{i,a} < \frac{W}{2}$$

Todos los resultados que se han llevado a cabo han sido probados sobre una relación de 1080 tuplas. Dicha relación contiene una clave primaria y 12 atributos. La Figura 5 muestra el porcentaje de atributos donde perdura la marca después de haber añadido ruido gausiano. Podemos comprobar que cuanto mayor es W más difícil es destruir la marca.

Como ya hemos comentado, el método propuesto por Agrawal-Haas-Kiernan no es muy resistente al ruido gausiano. En la Figura 6, se compara nuestro algoritmo con el método de Agrawal-Haas-Kiernan y se observa la mejora que aporta nuestro sistema. La figura muestra el porcentaje de éxito de un ataque, es decir, las posibilidades que

tiene un atacante de destruir la marca. Se ha usado una marca W=10 y W=25 frente a las tres posibilidades del sistema de Agrawal-Haas-Kiernan: marcar los 4, 6 u 8 bits menos significativos. Como se puede ver en la imagen, nuestra marca resiste el ruido gausiano notablemente en comparación con el método de Agrawal-Haas-Kiernan, ya que los ataques tienen escasas posibilidades de éxito.

4.2 Ataques de transposición

Este tipo de ataques consisten en intercambiar el orden de las tuplas. Nuestro sistema resiste este ataque, igual que el método de Agrawal-Haas-Kiernan, puesto que tanto el algoritmo de inserción como el de recuperación no dependen del orden de los datos; sinó que dependen de la clave primaria de los mismos, junto a la clave secreta del propietario de la base de datos, las cuales permanecen inalteradas tras un ataque de transposición.

5 Conclusiones y mejoras futuras

Este artículo propone un nuevo sistema de inserción de marcas de agua destinado a la protección de bases de datos numéricas. El método es más robusto que los métodos anteriores ya que presenta resistencia al el ruido gausiano y a las transposiciones a la vez que conserva la media y la desviación típica de los datos. Este sistema se puede considerar una síntesis entre el método de Agrawal-Haas-Kiernan y el marcaje basado en Expansión de Espectro que obtiene como resultado una mejor tolerancia al ruido.

Como propuestas para futuros trabajos consideramos:

- Mejorar la robustez del sistema haciéndolo resistente a más ataques como por ejemplo: ataques de subconjunto, inversión de bits, etc.
- Estudiar como pueden favorecer al sistema otros dominios de inserción (e.g. domino de Fourier).
- Obtener un procedimiento automático para determinar el umbral de detección de la marca.

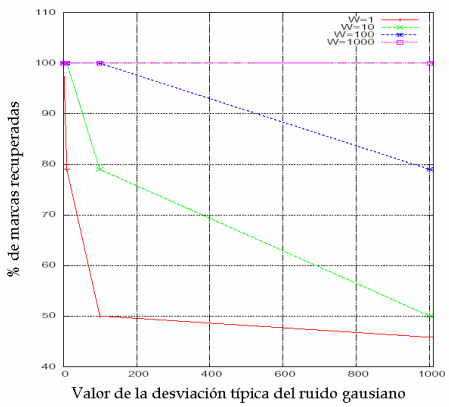

Figura 5: Porcentaje de atributos para los cuales se recupera exitosamente la marca después de añadir ruido gausiano. Los resultados estan en función de la desviación típica del ruido, para varios valores de W

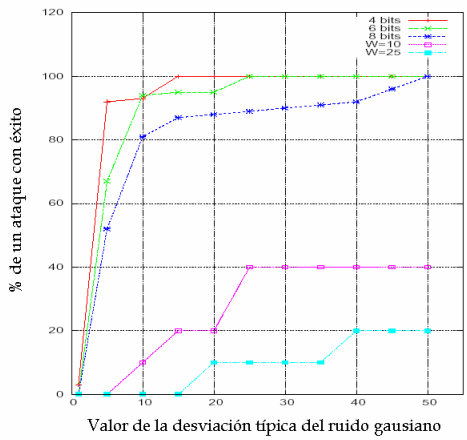

Figura 6: Comparación entre la respuesta a la adición de ruido gausiano de nuestro método y el propuesto por Agrawal-Haas-Kiernan-Haas-Kiernan. Los resultados están expresados en porcentaje de éxito de un ataque en función de la desviación típica del ruido.

Referencias

[AHK03] R. Agrawal-Haas-Kiernan, P. J. Haas, and J. Kiernan. Watermarking relational data: Framework, algorithms and analysis. *VLDB journal*, 2003.

[CT02] C. S. Collberg and C. Thomborson. Watermarking, tamper-proofing and obfuscation: tools for software protection.
IEEE Transactions on Software Engineering, 28(8):735–746, 2002.
http://dx.doi.org/10.1109/TSE.2002.1027797.

[DVM97] J. F. Delaigle, C. De Vleeschuwer, and B. Macq. *Watermarking Using a Matching Model Based on the Human Visual System*. Marly le Roi, 1997.

[HG98] F. Hartung and B. Girod. Watermarking of uncompressed and compressed video. *Signal Processing*, 66(3):283–301, Mayo 1998.

[JV97] F. Jordan and T. Vynne. Motion vector watermarking. Patent, 1997. Laboratoire de Traitement des Signaux ´Ecole Polytechnique Fédérale de Lausanne.

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

[Mor95] N. Moreau. *Techniques de compression des signaux*. Collection technique et scientifique des télécomunications. Masson, 1995.

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

[RDB96] J.J.K. ´O Ruanaidth, W.J. Downling, and F.M. Boland. Phase watermarking of digital images. In *Proceedings of the IEEE international Conference on Image Processing*, volumen 3, páginas 239–242, Septiembre 1996.

[SST97] M. D. Swanson, B. Shu, and A. H. Tew-fik. Data hiding for video-in-video. In *Proceedings of the International Conference on Image Processing*, volumen 1, páginas 676–679, Santa Barbara CA, 1997.

[Ver98] S. Verdú. *Multiuser Detection*. Cambridge University Press, 1998.