Distortion Optimization of Model-Based Secure Embedding Schemes for Data-Hiding

Benjamin Mathon, Patrick Bas, François Cayre, Fernando Pérez Gonzalez

Distortion Optimization of Model-Based Secure Embedding Schemes for Data-Hiding

Benjamin Mathon[1], Patrick Bas[1], François Cayre[1], and Fernando Pérez-González[2]

[1] Gipsa-Lab
Département Images et Signal
961 rue de la Houille Blanche
Domaine universitaire - BP 46
38402 Saint Martin d'Hères cedex, France
{benjamin.mathon,patrick.bas,francois.cayre}@gipsa-lab.inpg.fr
[2] Universidad de Vigo
Departamento de Teoría de la Señal y las Comunicaciones
Signal Processing in Communications Group
36200 Vigo, Spain
fperez@gts.tsc.uvigo.es

Abstract. This paper is the continuation of works about analysis of secure watermarking schemes in the case of WOA (Watermarked Only Attack) framework. In previous works, two new BPSK spread-spectrum watermarking modulations, Natural Watermarking (NW) and Circular Watermarking (CW), have been proposed and have been shown to be more secure than classical modulations. Because security is guaranted using specific distributions of watermarked contents, we propose to use optimal *model-based embedding* to ensure security while minimizing the overall distortion. Additionally, we propose a new secure watermarking scheme based on distribution of vector norms in the Gaussian case. We illustrate model-based embedding performance in the case of Gaussian signals and show that this approach not only allows to achieve excellent level of security in the WOA framework, but also allows to minimize distortion. Finally, a comparison of the robustness of the proposed embedding schemes is performed.

Key words: Watermarking, Security, Hungarian method, Distortion Optimization.

1 Introduction

Watermarking is a mean to hide information into digital contents (images, sounds, videos). This hidden information can be used for copyright or copy protection applications, integrity checking, or fingerprinting in order to control each copy of a numerical document. Our works focus on copyright and copy protection. The embedding of the message must meet many constraints:

- *imperceptibility*: the hidden information must not impair the original content for regular use,
- *capacity*: in multi-bit watermarking, one must ensure a sufficient number of bits can be reliably hidden into the host content,
- *robustness*: the hidden message should still be readable after common media processing,
- *security*: this last constraint can be viewed as *"the inability by unauthorized users to access, remove, read or write the hidden message"* [1]. Security in general is based on Kerckhoffs' principle [2]: the key is the only unknown parameter for the adversary.

This secret key is used to embed and detect (zero-bit watermarking) or decode (multi-bit watermarking) the watermark. Security is different from robustness. Robustness attacks relate to watermark survivability under common processing (in the case of still images, one may want to resist geometrical deformations, compression or noise addition). These attacks are generally not intentional.

On the contrary, security attacks are intentional and relate to the estimation of a part or all the secret key [3]. In multi-bit watermarking, the key is defined by the location of a set of codewords in a subspace. To embed a message in a host content, it must be placed in the decoding region of the right codeword. If an adversary learns the secret key, he can alter the message while minimizing the attack distortion with a 100% probability: his attack becomes deterministic. For example, he can design an attack in order to "push" the watermarked content into the nearest (wrong) decoding region, he can read the hidden message and copy the watermark in another content (copy attack [4]).

This paper deals with the case of WOA (Watermarked Only Attack) [5], the adversary has only access to several marked contents and to the source code of the watermarking algorithm (Kerckhoffs' principle). One should notice that the adversary can model the distribution of the host contents (Gaussian mixture or generalized Gaussian distribution for DCT or wavelet coefficients for example) since he knows the watermarking space. The adversary's goal is twofold. On one hand, he wants to model the conditional distribution of marked contents given the secret key, and on the other hand, he wants to estimate each codeword location.

Based on these ideas, we can find [6] the definition of four security classes to rank watermarking schemes security in the WOA framework (see Fig. 1 for relationships among them).

The first class is *insecurity*. In this class, the conditional distribution (given the key) of marked contents is not the same for all keys. By exhaustive search (or a more involved technique), he can estimate both the private subspace and the codewords.

A watermarking scheme belongs to the second class, *key-security*, if, for a subset of keys, the conditional distribution of the marked contents given each key is the same. The adversary can find this subset of keys and he can find the secret subspace but he cannot gain more information about the codewords.

The third class is called *subspace-security*. In this case, the conditional distribution of marked contents given the key will be the same for all keys. Therefore the pirate cannot gain any information about the key (he has not access to the secret subspace).

The last class *stego-security* relates to steganography: the distribution of marked contents is the same than that of the host contents. The adversary cannot decide whether the contents are marked or not.

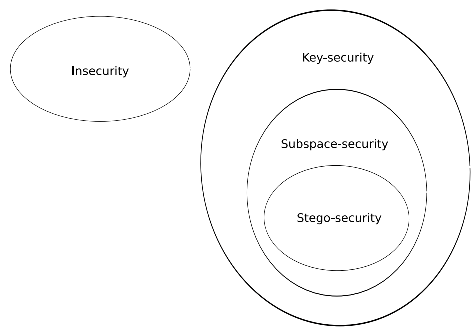


Fig. 1. Security classes in WOA framework.

Obviously, security in the WOA framework is strongly linked with the distribution of the contents after watermark embedding. In [7], authors have applied two secure modulations (namely Circular and Natural Watermarking, resp. CW and NW in the sequel) to still images. These modulations are used to modify the distribution of marked contents in order to be stego-secure or subspace-secure for NW and key-secure for CW. Note that such an approach is similar to the one proposed by Sallee in steganography [8], where perfect secrecy [9] is guaranteed by constraining the distributions of stego contents to be identical with the distributions of cover contents.

In this paper, we propose a new method to watermark signals in order to fit a chosen *target distribution* in an optimal way. This method uses the Hungarian algorithm, which minimizes distortion on average (in the sense of Euclidean distance) between points of host distribution and points of the target distribution. Section 2 recalls basics on BPSK-based SS watermarking schemes with an emphasis on unsecure and secure modulations and presents a watermarking scheme based on distribution of norm of signals. Section 3 presents our model-based embedding scheme by using Hungarian method. Section 4 compares our implementations of classical versus model-based embedding from security, distortion and robustness point of view over 2000 Gaussian signals.

2 Secure watermarking schemes

This section recalls definitions of unsecure and secure modulations, moreover we propose also a new secure embedding scheme based on the modification of the norm of the host vector.

2.1 Notations and definitions

We first list the conventions used in this paper. Data are written in small letters. Vectors and matrices are set in bold fonts. Vectors are written in small letters and matrices in capital ones. $\mathbf{x}(i)$ is the i-th component of a vector $\mathbf{x}$ and $\mathbf{x}_j$ is the vector $\mathbf{x}$ associated to a j-th observation. We write $(\mathbf{x}(0), \mathbf{x}(1), \mathbf{x}(2), ...)$ the content of a vector $\mathbf{x}$. We note [a;b],]a;b[, [a;b[and]a;b] real-intervals. Sets are noted in capital letters and $vect\,(A)$ represents the vector space generated by A. $p\,(\mathbf{x}_j)$ denotes the distribution of vectors $\mathbf{x}_j$. $\sigma^2_{\mathbf{x}}$ denotes the variance of a signal $\mathbf{x}$ and $\langle .|. \rangle$ denotes the usual scalar product.

We want to hide a message $\mathbf{m}$ of N_c bits in a host vector $\mathbf{x} \in \mathbb{R}^{N_v}$. So we create a watermark signal $\mathbf{w} \in \mathbb{R}^{N_v}$ in order to obtain $\mathbf{y} = \mathbf{x}+\mathbf{w}$ the watermarked signal. The secret key K used to embed and decode the message $\mathbf{m}$ is the seed of a PRNG. With K, we generate N_c Gaussian carriers $\mathbf{u}_i \in \mathbb{R}^{N_v}$. Each carrier is able to hide one bit. In order to have a null ISI (Inter Symbol Interference), carriers must be orthogonal. Thanks to $s : \{0, 1\} \to \mathbb{R}$, a modulation, we can create $\mathbf{w}$ by:

$$\mathbf{w} = \sum_{i=0}^{N_c-1} \mathbf{u}_i s(\mathbf{m}(i)). \tag{1}$$

Distortion is assessed by means of the WCR (Watermark-to-Content Ratio):

$$WCR_{[dB]} = 10 \log_{10} \left(\frac{\sigma^2_{\mathbf{w}}}{\sigma^2_{\mathbf{x}}} \right). \tag{2}$$

We model robustness attacks by adding Gaussian noise $\mathbf{n}$. So we consider the attacked vector $\mathbf{r} = \mathbf{y} + \mathbf{n}$. Attack strength is assessed by means of the $WCNR$ (Watermarked Content-to-Noise Ratio):

$$WCNR_{[dB]} = 10 \log_{10} \left(\frac{\sigma^2_{\mathbf{y}}}{\sigma^2_{\mathbf{n}}} \right). \tag{3}$$

Decoding is classically obtained by correlations z:

$$z_{\mathbf{r},\mathbf{u}_i} = \sum_{j=0}^{N_v-1} \mathbf{r}(j)\mathbf{u}_i(j). \tag{4}$$

We consider $\widehat{\mathbf{m}}$ the estimated message, so we have for each bit:

$$\hat{\mathbf{m}}(i) = \begin{cases} 0 \text{ if } z_{\mathbf{r},\mathbf{u}_i} > 0, \\ 1 \text{ if } z_{\mathbf{r},\mathbf{u}_i} < 0. \end{cases} \tag{5}$$

We measure robustness of the watermarking scheme by BER (Bit Error Rate) between the estimated and the original message:

$$BER(\mathbf{m}, \hat{\mathbf{m}}) = \frac{1}{N_c} \sum_{i=0}^{N_c-1} \mathbf{m}(i) \oplus \hat{\mathbf{m}}(i). \tag{6}$$

For a pirate, there is no difference between estimating the carriers or getting K in the WOA context. According to [6], the security of a watermarking scheme relies on the properties of the conditional distribution $p(\mathbf{y}_j|\mathrm{K})$. In the case of spead-spectrum techniques, points of conditional distribution given the carriers are the N_c-tuples $(z_{\mathbf{y},\mathbf{u}_0}, ..., z_{\mathbf{y},\mathbf{u}_{N_c-1}})$ in the subspace spanned by the carriers.

2.2 Unsecure SS modulations:

Classical modulation SS (Spread Spectrum) is given by:

$$s_{SS}(\mathbf{m}(i)) = \gamma(-1)^{\mathbf{m}(i)}. \tag{7}$$

This modulation is analog to BPSK modulation. Parameter γ is used to set the power of watermark. It is a function of WCR. ISS (Improved Spread Spectrum) [10] uses side-information to improve robustness:

$$s_{ISS}(\mathbf{m}(i)) = \alpha(-1)^{\mathbf{m}(i)} - \lambda\frac{\langle \mathbf{x}|\mathbf{u}_i\rangle}{\|\mathbf{u}_i\|^2}, \tag{8}$$

where $\langle .|.\rangle$ denotes the usual scalar product, α and λ are computed to achieve host-interference rejection and error probability minimisation given Noise-to-Content power Ratio:

$$NCR_{[dB]} = 10\log_{10}\left(\frac{\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{x}}^2}\right), \tag{9}$$

where $\mathbf{n}$ denotes Gaussian noise. Previous works [5] have shown that SS and ISS are unsecure, and that carriers estimation is possible (see tests on [7]).

2.3 Secure modulations:

NW (Natural Watermarking) [6] modulation is defined by:

$$s_{NW}(\mathbf{m}(i)) = \left((-1)^{\mathbf{m}(i)}\frac{\langle \mathbf{x}|\mathbf{u}_i\rangle}{|\langle \mathbf{x}|\mathbf{u}_i\rangle|} - 1\right)\frac{\langle \mathbf{x}|\mathbf{u}_i\rangle}{\|\mathbf{u}_i\|^2}. \tag{10}$$

NW belongs to the so-called stego-secure class and is suitable for steganography applications.
CW (Circular Watermarking) [6] modulation is defined by:

$$s_{CW}(\mathbf{m}(i)) = \alpha(-1)^{\mathbf{m}(i)}\mathbf{d}(i) - \lambda\frac{\langle \mathbf{x}|\mathbf{u}_i\rangle}{\|\mathbf{u}_i\|^2}, \tag{11}$$

where α and λ are computed the same way than with ISS and $\mathbf{d}$ is generated at each embedding as follows from $\mathbf{g} \sim \mathcal{N}(0,1)$, this parameter is used to randomly spread the correlations of the mixed signals on the whole decoding regions:

$$\mathbf{d}(i) = \frac{|\mathbf{g}(i)|}{\|\mathbf{g}\|}. \tag{12}$$

This parameter $\mathbf{d}$ enables the following property of circularity :

$$p\left(z_{\mathbf{y},\mathbf{u}_0}, ..., z_{\mathbf{y},\mathbf{u}_{N_c-1}}\right) = p\left(\sqrt{\sum_{i=0}^{N_c-1} z^2_{\mathbf{y},\mathbf{u}_i}}\right). \tag{13}$$

The circularity of the distribution allows us to say that for a subset of several keys (all bases of $vect(\{\mathbf{u}_i\})$), the distribution of marked signals will be the same. CW belongs to the so-called key-secure security class.

2.4 A new secure embedding based on the χ^2 distribution

Based on previous security assessment, we are able to propose a stego-secure watermarking scheme χ^2W (CHI2 Watermarking) by modifying norm of Gaussian host signals after embedding, while keeping the same distribution between original and watermarked contents. Distribution of these norms can be modeled by a χ^2 law. So we define codewords location by real-intervals in the set of norms of Gaussian vectors. Finally, in order to embed a secret message, we chose randomly a norm in the corresponding interval and we multiply the host vector in order to have the desired norm. Contrary to BPSK modulations, the watermarking subspace (space of norms) is not private, secret relies only on the partition of the real-positive axis representing the norms. There is no security flaw because we works on the WOA framework. Adversaries do not know the embedded message. However, this embedding is easy to implement, it enables to achieve stego-secure embedding since the distributions of original and watermarked contents are the same, and brings another scheme to compare with in term of robustness.

We use previous notations of Part. 2.1. We want to create a watermarked signal $\mathbf{y} = \alpha\mathbf{x}$, $\alpha \in \mathbb{R}^+$. This method is based on the distribution of the norms of host Gaussian vectors. In fact, if $\mathbf{x} \in \mathbb{R}^{N_v}$ with $\mathbf{x} \sim \mathcal{N}(0,1)$, $\|\mathbf{x}\|^2 \sim \chi^2(N_v)$ (Chi-2 law of degrees N_v), $\|.\|$ representing the euclidean norm. Codewords are sets in a partition of $[0, +\infty[$. To embed a secret message $\mathbf{m}$ in a host vector $\mathbf{x}$, we randomly choose a norm $\|\mathbf{y}\|^2$ in the corresponding real-interval and we compute:

$$\mathbf{y} = \sqrt{\frac{\|\mathbf{y}\|^2}{\|\mathbf{x}\|^2}}\mathbf{x}. \tag{14}$$

We obtain the watermark signal $\mathbf{w} = \mathbf{y} - \mathbf{x}$. This process can be considered as a variant of Moulin and Briassouli stochastic embedding [11] who work on different host distributions. Different consequences arise from this new embedding scheme:

- The means of choosing a norm in the right codeword is not optimal (from the distortion point of view): we generate real numbers until we have one of them in the desired interval.
- To define the secret partition on the real-positive axis, we use an estimator of the quantile function of χ^2 distribution.
- With the condition of message equiprobability, we must have more than one codeword for each message. Without this condition, a pirate can find the secret partition by using a quantile function (he separates the p.m.f. into 2^{N_c} parts of probability $\frac{1}{2^{N_c}}$). We denote N_w the number of codewords used.

Fig. 2 shows a secure partition of real positive axis and the associated distribution with parameters $N_c = 2, N_w = 8, N_v = 55$.

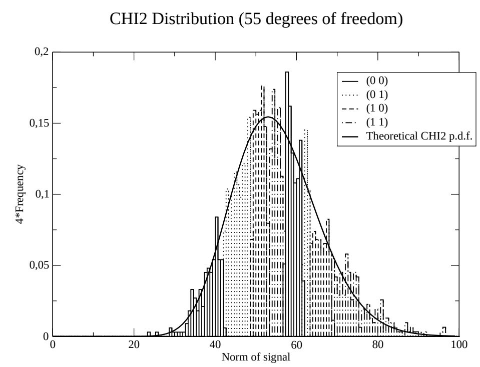


Fig. 2. Secret partition of real-positive axis and associated distribution with parameters $N_c = 2, N_w = 8, N_v = 55$, each message is coded into 2 codewords. We have constructed the bins by generating 2000 Gaussian vectors for each message and calculating their respectives norms.

Note that the decoding regions are the same for NW and CW (they are delimited by hyperplanes) but different from the ones related to χ^2W (delimited by hyperspheres). Fig. 3 shows 2D representations of coding regions for spread-spctrum schemes and for χ^2W. We use $N_c = 2$.

3 Minimisation of the embedding distortion

We have seen that the level of security of the previously watermarking methods is given by the distribution of the signals after watermarking. Note that these distributions are defined by the distribution of correlations for BPSK modulations and the distribution of the norms for χ^2W. However, from the distortion point of view, these implementations are not optimal. We propose in the next section a new scheme which can ensure a given distribution of our marked signals

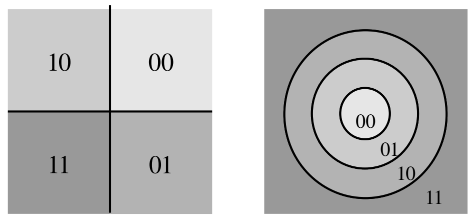


Fig. 3. 2D representations of coding regions for NW and CW (left part) and for χ^2W (right part) for 2 bits.

while minimizing the embedding distortion. We want to associate each point of host distribution with each point of a chosen distribution with a minimal average Euclidean distance between these two points (distortion is proportional to the distance). The Hungarian algorithm is a mean to solve this problem. To explain this algorithm, some reminders about graph theory are useful.

3.1 Minimal cost perfect matching in a bipartite graph

A bipartite graph is a graph $G = (V, E)$ with the following property: there exists a partition $V = A \sqcup B$, each edge of E is of the form $[a, b]$ with $a \in A$ and $b \in B$. Moreover G satisfy $|A| = |B| = N_m$. A weighted bipartite graph $G = (V, E, P)$ is a bipartite graph where each edge is weighted by a function $P : E \to \mathbb{R}$. A perfect matching M of G is defined as a subset of E with N_m elements where each vertex is incident with exactly one member of M. In this paper, we are interested in the Assignement Problem (AP), we search the minimal cost perfect matching M^*, i.e., a perfect matching whose the sum of weights of edges is minimal. More precisely we search:

$$M^* = \arg\min_M \sum_{t \in M} P(t). \tag{15}$$

Fig. 4 shows a weighted bipartite graph and its minimal cost perfect matching.

3.2 The Hungarian method for the AP

The Hungarian method [12] is an efficient algorithm to solve the AP in a weighted bipartite graph in polynomial time ($O({N_m}^3)$). We consider $G = (V, E, P)$ a weighted bipartite graph with:

- $V = A \sqcup B$,
- $A = \{a_0, ..., a_{N_m-1}\}$,
- $B = \{b_0, ..., b_{N_m-1}\}$.

We consider $\mathbf{D} \in \mathcal{M}_{N_m,N_m}(\mathbb{R})$, a matrix initialized with $\mathbf{D}(i, j) = P([a_i, b_j])$. The goal is to choose N_m elements of this matrix in order to have each row and each column containing one chosen element. In fact, the minimal cost perfect

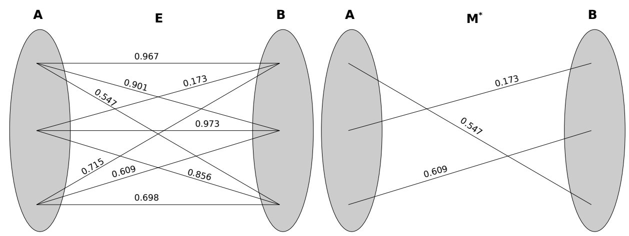


Fig. 4. Example of a weighted bipartite graph with two partitions of three vertices and minimal cost perfect matching found by Hungarian Method (weights between two vertices are noted on each corresponding edge).

matching is the set of edges corresponding to these choosen elements. The Hungarian algorithm does the following:

1. Subtract the entries of each row by the row minimum: each row has at least one zero, all entries are positive or zero.
2. Subtract the entries of each column by the column minimum: each row and each column has at least one zero.
3. Select rows and columns across which to draw lines, in such a way that all the zeros are covered and that no more lines have been drawn than necessary.
4. A test for optimality:
   – If the number of the lines is n, choose a combination from the modified cost matrix in such a way that the sum is zero.
   – If the number of the lines is < n, go to 5.
5. Find the smallest element which is not covered by any of the lines. Then subtract it from each entry which is not covered by the lines and add it to each entry which is covered by a vertical and a horizontal line. Go back to 3.

So we have:

$$M^* = \{[a_i, b_j] : \mathbf{D}(i,j) = 0\}. \tag{16}$$

3.3 Application for NW and CW embedding

Construction of bipartite graphs: We want to create 2^{N_c} bipartite graphs which contain, for the host partition, points of distribution of several host signals (correlations). We construct points of target distributions by selecting only points in codeword of the desired message. The goal is to find the minimal cost (euclidean distance) perfect matching between the two partitions. Formally, we use N_m host signals $\{\mathbf{x}_i\}_{i=0,\ldots,N_m-1}$.

For each host signal $\mathbf{x}_i$ we construct the correlation host vector:

$$\mathbf{z}_{\mathbf{x}_i} = \left(z_{\mathbf{x}_i,\mathbf{u}_0}, \ldots, z_{\mathbf{x}_i,\mathbf{u}_{N_c-1}}\right).$$

We want to construct 2^{N_c} target distributions. As we have seen before, NW and CW modulations can permit to construct distributions we want to obtain in order to match the security class we want (stego-secure, key-secure, ...).

We proceed by watermarking for each message $\mathbf{m}$, host signals $\mathbf{x}_i$ with the chosen modulation to obtain $\mathbf{y}_i$. Finally, we construct the correlation marked vector $\mathbf{z}_{\mathbf{y}_i} = \left(z_{\mathbf{y}_i,\mathbf{u}_0}, ..., z_{\mathbf{y}_i,\mathbf{u}_{N_c-1}}\right)$. We obtain, for each message, the weighted bipartite graph $G = \{X \sqcup Y, A, P\}$ with:

- $X = \{\mathbf{z}_{\mathbf{x}_i}\}_{i=0,...,N_m-1}$,
- $Y = \{\mathbf{z}_{\mathbf{y}_j}\}_{j=0,...,N_m-1}$ (depends on $\mathbf{m}$),
- A defines the set of edges $[\mathbf{z}_{\mathbf{x}_i}, \mathbf{z}_{\mathbf{y}_j}]$ of the graph G,
- P is the weight function of the edges of G.

$$P([\mathbf{z}_{\mathbf{x}_i}, \mathbf{z}_{\mathbf{y}_j}]) = \|\mathbf{z}_{\mathbf{x}_i} - \mathbf{z}_{\mathbf{y}_j}\|_2.$$

We obtain 2^{N_c} minimal cost perfect matchings M_k^* between host correlations and marked correlations by using Hungarian method.

Mapping reduction: In previous section, we construct one bipartite graph for one message to embed. In order to reduce complexity of bijections calculus, we can use property of symmetry of our distributions (the axis of symmetry are the carriers); points of target distributions are computed in order to embed a constant message (for example the message $(1, 1, ..., 1)$). For the rest of this article, we notate N_m-map a triplet (X, Y, M^*) constructed with N_m host signals. Fig. 5 shows a correlation bipartite graph with $N_m = 3$, $N_c = 2$ and its minimal cost perfect matching found by Hungarian method.

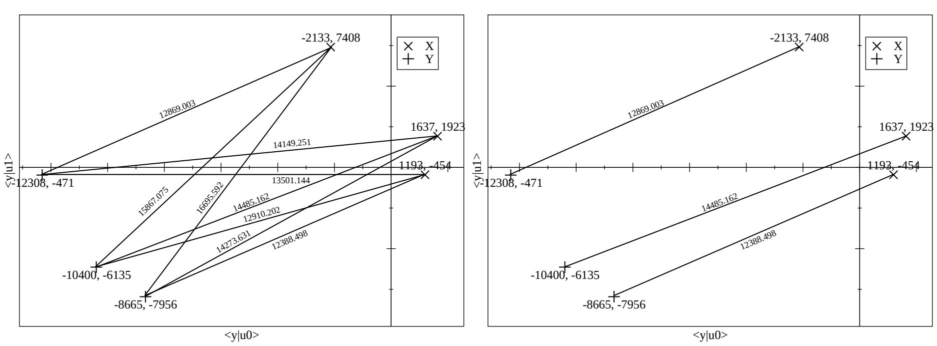


Fig. 5. Projection over two carriers ($N_c = 2$): correlation bipartite graph construction and minimal cost perfect matching found by Hungarian method with two partitions of three vertices, host correlations and marked correlations ($N_m = 3$). Euclidean distances (weights) between two vertices are noted on each corresponding edge. The minimal cost perfect matching associates elements of each vertex partition while minimizing the summation of the distances.

Model-based embedding: We consider (X, Y, M^*) the N_m-map constructed in the previous section. Now, we want to mark a signal $\mathbf{x}$ with any message $\mathbf{m}$ using the map. First, we compute the correlation host vector $\mathbf{z_x} = \left(z_{\mathbf{x},\mathbf{u}_0}, ..., z_{\mathbf{x},\mathbf{u}_{N_c-1}}\right)$. We want to associate $\mathbf{z_x}$ with a point of Y by using M^*.

We have seen that elements of Y have been constructed in order to embed the message $(1, ..., 1)$. We note indices of $\mathbf{m}$ where the bit is different from 1. Note that a matching from an original content to a given codeword enables to generate matchings to any codewords by symmetrising both points along appropriated axes. Consequently sign changes must be made on the coefficients of $\mathbf{z_x}$ in the indices that have undergone symmetries. Afterwards inverse symmetry must be performed after watermarking in order to embed the correct message $\mathbf{m}$. Formally, we construct $\mathbf{Rz_x}$,

$$\mathbf{R} \in \mathcal{M}_{N_c,N_c}(\mathbb{R}), \mathbf{R}(i,j) = \begin{cases} 0 \text{ if } i \neq j, \\ (-1)^{\mathbf{m}(i)+1} \text{ if } i = j. \end{cases}$$

Next, we find the nearest neighboor (minimal euclidean distance) of $\mathbf{Rz_x}$ in X, for example $\mathbf{z}_{\mathbf{x}_{i_0}}$. Thanks to the perfect matching M^* we find $\mathbf{z}_{\mathbf{y}_{j_0}}$, the correspondance of $\mathbf{z}_{\mathbf{x}_{i_0}}$. Next, we apply inverse symmetry to compute the correlation marked vector $\mathbf{z_y}$:

$$\mathbf{z_y} = \mathbf{R^{-1}} \mathbf{z}_{\mathbf{y}_{j_0}} = \mathbf{R} \mathbf{z}_{\mathbf{y}_{j_0}}.$$

So, we obtain the correlation vector of our marked signal. By a difference between $\mathbf{z_y}$ and $\mathbf{z_x}$, we have $\mathbf{z_w}$, the watermark correlation vector. Proper retro-projection of this signal in the N_vD-space is assured by :

$$\mathbf{w} = \sum_{i=0}^{N_c-1} \frac{\mathbf{z_w}(i)}{\langle \mathbf{u}_i | \mathbf{u}_i \rangle} \mathbf{u}_i.$$

And finally, we compute the watermarked signal $\mathbf{y} = \mathbf{x} + \mathbf{w}$. Fig. 6 shows this process by using the 3-map constructed, see Fig. 5.

3.4 Application for χ^2 Watermarking

Construction of bipartite graphs: We want to create 2^{N_c} bipartite graphs which contain, for the host distribution, norms of several host signals. We construct points of target distribution by selecting only real points in codewords of the desired message. We generate N_m Gaussian vectors $\mathbf{x}_i$ and, for each message, N_m Gaussian vectors $\mathbf{y}_j$ with $\|\mathbf{y}_j\|^2$ in the right codewords. We can construct, for each $k = 0, ..., 2^{N_c} - 1$ a bipartite graph $G = \{X \sqcup Y, A, P\}$ with:

- $X = \{\|\mathbf{x}_i\|^2\}_{i=0,...,N_m-1}$,
- $Y = \{\|\mathbf{y}_j\|^2\}_{j=0,...,N_m-1}$,
- A defines the set of edges $[\|\mathbf{x}_i\|^2, \|\mathbf{y}_j\|^2]$ of the graph G,
- P is the weight function of the edges of G. We use the absolute value of the differences of norms.

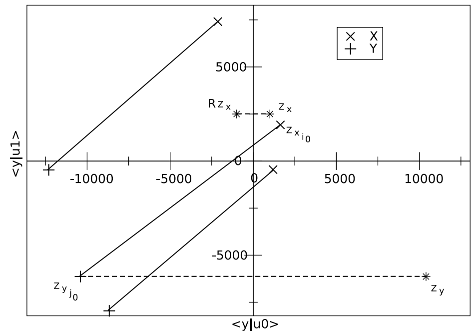


Fig. 6. Model-based watermarking scheme: illustration of Sec. 3.3. After calculating the correlation vector $\mathbf{z_x}$ of a host signal, we compute the watermarked correlations $\mathbf{z_y}$ by using the 3-map of Fig. 5 ($N_m = 3$) with the constant message $\mathbf{m} = (0, 1)$ ($N_c = 2$).

By using the Hungarian algorithm, we find 2^{N_c} minimal cost perfect matching M_k^* between host norms and marked norms (functions of embedded message). For the rest of this article, we denote N_m-smap the set $\{(X, Y_k, M_k^*)\}_{k=0,\ldots,2^{N_c}-1}$ constructed with N_m signals.

Embedding: Process is similar to SS model-based embedding. To embed a vector $\mathbf{x}$ with message $\mathbf{m}$ and a N_m-smap; we calculate $\|\mathbf{x}\|^2$. Next we find the nearest neighboor of $\|\mathbf{x}\|^2$ in X. By M_k^* (k depends on $\mathbf{m}$), we find $\|\mathbf{y}\|^2$. Finally, we obtain $\mathbf{y} = \sqrt{\frac{\|\mathbf{y}\|^2}{\|\mathbf{x}\|^2}}\mathbf{x}$.

4 Experiments

The goal of this section is to assess the preservation of the distributions after the Hungarian method, the impact of this method on distortion and the general robustness of the three secure embedding schemes we presented.

4.1 Numerical values and assessments

In practice, $N_c = 2$. For NW and CW, we use $N_v = 256$ and for χ^2W, we use $N_v = 55$ (in order to have the same distortion for the three schemes, $WCR = -18dB$), $N_w = 8$. Tests are made with 2000 host Gaussian signals. For χ^2W, in order to have the equiprobable condition, we use an estimator of a fractile function of χ^2 distribution given in [13] which uses an estimator of the repartition function of a normal distribution in [14], we use the partition of the real-positive axis defined on Tab. 1. We have constructed a 10000-map and a 10000-smap with our signals database by using CW, NW and χ^2W and we have marked the initial 2000 signals by using on the one hand the secure modulation, and on the other hand the corresponding model-based found by Hungarian method (HCW, HNW and Hχ^2W).

message	probability	real-interval
(0 0)	0.1	[0;42.06]
(0 1)	0.2	[42.06;49.055]
(1 0)	0.15	[49.055;53.037]
(1 1)	0.15	[53.037;57.016]
(0 0)	0.15	[57.016;61.665]
(0 1)	0.05	[61.665;63.577]
(1 0)	0.1	[63.577;68.796]
(1 1)	0.1	[68.796;$+\infty$[

Table 1. χ^2W: secret key used for $N_c = 2$, $N_w = 8$, $N_v = 55$. This table shows the real-interval functions of probability of messages. Each message appears with the same probability (0.25). This partition is the same than in Fig. 2.

4.2 Distribution preservation after the Hungarian method

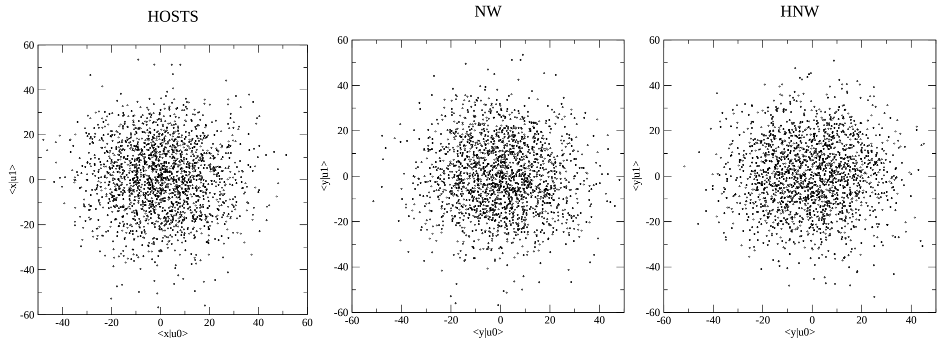
Fig. 7 shows host, NW and HNW distributions on two carriers. As we can see, distribution of correlations after Natural Watermarking is the same than distribution of host correlations. It is consistent with the definition of stego-security. Moreover, correlations are the same after using Hungarian Method. Our model-based doesn't impair security. Distribution of CW and HCW on two carriers is shown on Fig. 8, we can see that the distribution is circular and we can conclude that for all bases $(\widehat{\mathbf{u}}_\mathbf{0}, \widehat{\mathbf{u}}_\mathbf{1})$ of $vect(\mathbf{u}_0, \mathbf{u}_1)$ the distribution $p(\mathbf{y}_0, ..., \mathbf{y}_{1999}|\widehat{\mathbf{u}}_\mathbf{0}, \widehat{\mathbf{u}}_\mathbf{1})$ will be the same (rotations of the secret subspace). It is consistent with the definition of key-security, the pirate can access the subspace of the codewords but has no information about the decoding regions. As HNW, HCW doesn't impair security. Fig. 9 shows projections of host, χ^2W and Hχ^2W signals over the two first components. The distribution do not change with the two methods. As NW, χ^2W is stego-secure.

4.3 Distortion minimisation

Tab. 2 shows the impact on distortion obtained on average on our 2000 signals for NW, CW, χ^2W, HNW, HCW and Hχ^2W. We can see that we gain 2.7dB of distortion for NW, 1.1dB for CW and 3.6dB for χ^2W. This last result is due to the fact that there are two codewords for on message and these codewords are away in the real-positive axis.

4.4 Robustness

Beside distortion and security, the last constraint to assess is the general robustness of the presented schemes. We measure robustness of these secure modulations with and without distortion optimisation. Fig. 10 shows Bit Error Rate functions of chosen $WCNR$ over 2000 signals and we can verify that distortion optimisation does not modify robustness of our schemes. As we can see, CW is more robust than NW which is more robust than χ^2W. An insight of the poor



(a) (b) (c)

Fig. 7. (a): distribution of the projections of the host signals over two carriers. (b) and (c): distributions of the projection of the marked signals over two carriers for NW and HNW.

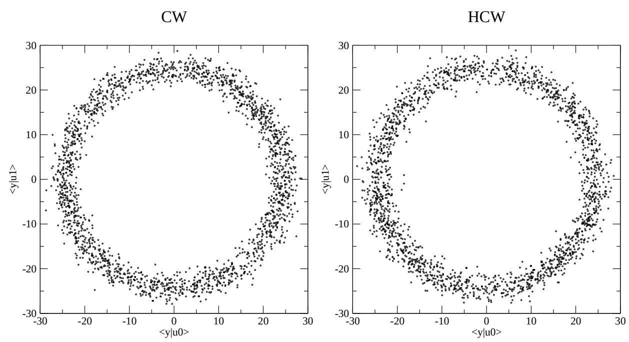
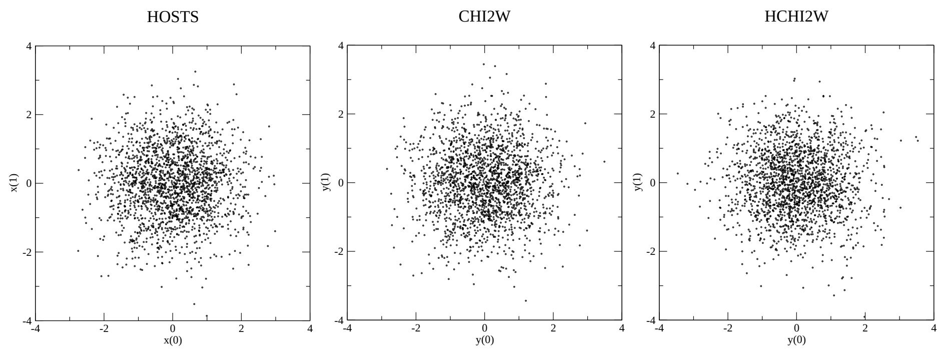


Fig. 8. Distribution of the projections of the marked signals over two carriers for CW and HCW.

robustness of χ^2W can be given by the fact that for this embedding, the decoding regions are always very close to each other (see Fig. 3). Consequently one watermarked vector corrupted by noise will have a higher probability to change of coding regions for χ^2W than for NW or CW. Note that we cannot show robustness of the six schemes with the same distortion because NW modulation does not allow to set a target distortion.

5 Conclusion

The goals of this paper are twofold: to propose and compare secure embedding schemes for data-hiding and to propose a general method to minimise the global embedding distortion for each scheme. The first point is addressed by proposing the χ^2 embedding scheme which is more fragile than other scheme like NW or



(a) (b) (c)

Fig. 9. (a): projection of host signals over the two first components. (b) and (c): projections of marked signals distributions over the two first components for χ^2W and Hχ^2W.

	WCR (classical)	WCR (with the Hungarian method)
NW	-18.07	-20.76
CW	-17.97	-19.11
χ^2W	-18.02	-21.65

Table 2. Distortion for NW, CW and χ^2W on initial embedding schemes and after using the Hungarian optimisation scheme.

CW. The optimisation regarding robustness appears not to be straightforward and future works will be devoted to find more robust schemes and to find coding regions that will improve the robustness of χ^2W.
Moreover, we found that the Hungarian method is the ideal practical tool to minimize distortion while guaranteeing a given class of security.

6 Acknowledgments

Benjamin Mathon, François Cayre and Patrick Bas are supported (in part) by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT and the National French projects Nebbiano ANR-06-SETIN-009, RIAM Estivale and ARA TSAR.

References

1. Kalker, T.: Considerations on watermarking security. Proc. MMSP (October 2001) 201–206
2. Kerckhoffs, A.: La cryptographie militaire. Journal des Sciences militaires **IX** (January 1883) 5–38

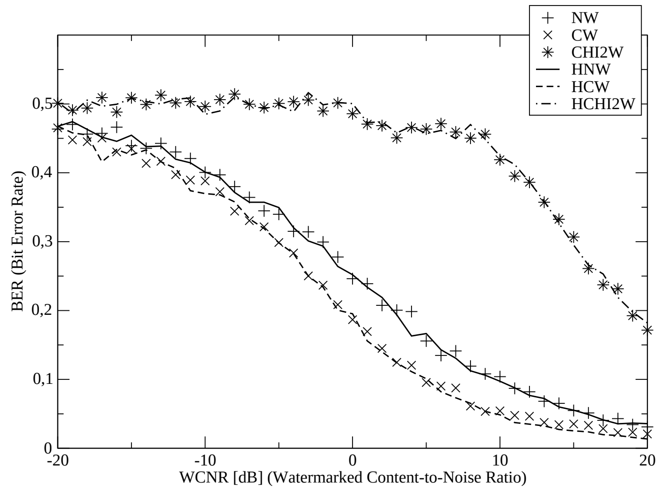


Fig. 10. BER vs WCNR for NW, CW and χ^2W. For NW, CW and χ^2W, we have $WCR = -18dB$. For HNW, HCW and Hχ^2W, we obtain respectively WCR = -20.76dB, -19.11dB and -21.65dB.

3. Comesaña, P., Pérez-Freire, L., Pérez-González, F.: Fundamentals of data hiding security and their application to spread-spectrum analysis. In: 7th Information Hiding Workshop, IH05. Lecture Notes in Computer Science, Barcelona, Spain, Springer Verlag (June 2005)
4. Kutter, M., Voloshynovskiy, S., Herrigel, A.: The watermark copy attack. In: Electronic Imaging 2000, Security and Watermarking of Multimedia Content II. Volume 3971. (2000)
5. Cayre, F., Furon, T., Fontaine, C.: Watermarking security: Theory and practice. IEEE Trans. Signal Process. **53**(10) (October 2005) 3976–3987
6. Cayre, F., Bas, P.: Kerckhoffs based embedding security classes. IEEE Trans. Inf. Forensics Security (2007)
7. Mathon, B., Bas, P., Cayre, F.: Practical performance analysis of secure modulations for woa spread-spectrum based image watermarking. Proc. ACM MM&Sec'07 (September 2007)
8. Sallee, P.: Model-based steganography. In: International Workshop on Digital Watermarking (IWDW), LNCS. Volume 2. (2003) 154–167
9. Shannon, C.E.: Communication theory of secrecy systems. Bell System Technical Journal **28** (October 1949) 656–715
10. Malvar, H.S., Flôrencio, D.: Improved spread spectrum: a new modulation technique for robust watermarking. IEEE Trans. Signal Process. **53** (April 2003) 898–905
11. Moulin, P., Briassouli, A.: A stochastic qim algorithm for robust, undetectable image watermarking. In: ICIP. (2004) 1173–1176
12. Kuhn, H. W.: The Hungarian method of solving the assignment problem. Naval Res. Logistics Quart. **2** (1955) 83–97
13. Hill, I.D., Pike, M.C.: Algorithm 299. ACM TOMS (June 1985) 185
14. Ibbetson, D.: Algorithm 209. Collected Algorithms of the CACM (1963) 616