Natural Watermarking: a secure spread spectrum technique for WOA

Patrick Bas, François Cayre

Natural Watermarking: a secure spread spectrum technique for WOA

Patrick Bas[1,2] and François Cayre[2]

[1] CIS / Helsinki University of Technology
P.O. Box 5400
FI-02015 HUT FINLAND
[2] LIS/INPG
961, rue de la Houille Blanche BP 46
F-38042 St. Martin d'Hères Cedex, France

Abstract. This paper presents a spread spectrum (SS) watermarking technique that is secure against carriers estimation in a Watermark Only Attack framework. After reviewing the sufficient conditions to design secure algorithms for watermarking and steganography, we present a setup based on Blind Source Separation (BSS) theory to assess the lack of security of classical SS techniques such as classical SS or ISS. We motivate a new SS watermarking algorithm called Natural Watermarking (NW) where the estimation of the secret carriers is impossible and which achieves perfect secrecy thanks to unchanged Gaussian distributions of the secret carriers. The theoretical evaluation of the NW security is carried out and the case of multi-bit embedding is addressed. Finally, a robust extension of NW is presented and the properties of NW and Robust-NW are both practically verified.

1 Introduction

Robustness, *capacity* and *imperceptibility* have always been considered, since the very beginning of watermarking, as the main three constraints to respect in order to build a valuable watermarking scheme. Recently the watermarking community has thrown light on the problem of *security* which appears also to be a fundamental constraint to respect in order to guaranty the usability of a watermarking technology. Several authors [1][2][3] showed that some information about the secret key may leak from several observations of watermarked pieces of content. Using this information, it may be possible to estimate the secret key, and then to destroy the security of the considered scheme by removing, copying or altering the embedded messages. Several studies address also the security of practical watermarking techniques for digital images [4][5].

In this paper, we tackle the problem of security for the well-known class of spread spectrum (SS) watermarking schemes. In this case, the secret key which practically is the seed of a random generator, corresponds to the set of secret carriers that is used to convey the information. It is important to note that an

attacker does not need the seed used to initialiaze the random number generator: the secret carriers are good enough to attack the watermark. We propose a watermarking scheme that is secure (e.g. it does not offer information leakage of the secret key) for the class of Watermark Only Attacks (WOA). This class of attacks, proposed by [1], considers an attack that is based on the observation of watermarked contents, watermarked with the same key but conveying different messages. We named the proposed scheme *natural spread spectrum watermarking* because embedding is achieved without altering the natural distribution of each secret carrier before and after embedding. As shown in the paper, this characteristic enables to achieve perfect secrecy. Moreover, when embedding several bits, we show that if each carrier is embedded in the contents with an amplitude following a Gaussian distribution, it is impossible to individually estimate the carriers.

The rest of the paper is divided into five sections. First, the security of classical SS techniques for WOA are analysed as a Blind Source Separation (BSS) problem: in section 2 we show that the characterisation of the distributions of each carrier for the observed contents enables to estimate the different carriers. Section 3 presents the constraints, principles and characteristics of Natural Watermarking (NW). The embedding, decoding and distortion related to NW are presented and the link with the Scalar Costa's Scheme, another scheme preserving perfect secrecy, is outlined. An extension of NW to increase the robustness is presented in section 4, the implications in term of security are also mentioned. Section 5 presents a comparison between the estimations of the secret carriers for different SS watermarking schemes including NW. We show that for NW it is impossible to estimate the carriers. For Robust-NW only the estimation of the watermark subspace is possible. Finally section 6 concludes this paper and presents open research lines for future works.

2 Assessing the security of spread-spectrum techniques using BSS techniques

2.1 Notations

Vectors are denoted in bold face ($\mathbf{v}$) and coefficients of vectors with parenthesis ($\mathbf{v}(i)$ is the coefficient number i in vector $\mathbf{v}$). Matrices are denoted in capital bold face and are generally composed of several realizations of vectors of the same name, column-wise: the columns of $\mathbf{V}$ are several realizations $\mathbf{v}_1 \dots \mathbf{v}_N$ of a "template" vector $\mathbf{v}$.

Let us denote $\mathbf{x}$ the host vector of N_v coefficients into which we want to hide a binary message vector $\mathbf{m}$ of N_c bits. The resulting watermarked vector is denoted $\mathbf{y}$. To this aim, we use $\mathbf{u}_i$ orthogonal carriers, $1 \leq i \leq N_c$. The decoded message is denoted $\hat{\mathbf{m}}$. It is to be estimated from $\mathbf{y}'$, a potentially degraded version of $\mathbf{y}$. Let us further denote $z_{\mathbf{v},\mathbf{u}_i}$ the correlation between a vector $\mathbf{v}$ and a carrier $\mathbf{u}_i$:

$$z_{\mathbf{v},\mathbf{u}_i} = < \mathbf{v} | \mathbf{u}_i > = \frac{1}{N_v} \sum_{k=1}^{N_v} \mathbf{v}(k)\mathbf{u}_i(k) \tag{1}$$

2.2 Information theoretical constraints for perfect secrecy

Perfect secrecy has different meanings according to the domain of application. For steganography, perfect secrecy means the impossibility to distinguish between an original content ($\mathbf{x}$) and a stego content ($\mathbf{y}$). Cachin studied the necessary conditions to obtain a secure steganographic scheme and claims that a scheme is secure if the Kullback-Leibler divergence D_{KL} between the distributions $P_{\mathbf{x}}$ and $P_{\mathbf{y}}$ of $\mathbf{x}$ and $\mathbf{y}$ is null. The quantity D_{KL} is defined by:

$$D_{KL}(P_{\mathbf{x}}||P_{\mathbf{y}}) = \sum_i P_{\mathbf{x}}(i) \log \frac{P_{\mathbf{x}}(i)}{P_{\mathbf{y}}(i)} \tag{2}$$

which means that perfect secrecy may be achieved if and only if the distributions of $\mathbf{x}$ and $\mathbf{y}$ are identical. A practical implementation of a steganographic scheme satisfying the perfect secrecy constraint has been proposed in [6].
For robust watermarking, the problem does not concern a possible distinction between the original and the watermarked content: it is not important to know wether a content is watermarked or not, but is it important not to disclose the secret carriers based on observations of pieces of watermarked contents. The concept of information leakage in the context of robust watermarking has been proposed in [1] and developed in [2][3]. The notion of information leakage stems from the definition of the mutual information between N_o watermarked contents $\mathbf{Y}$ and the secret carriers $\mathbf{U}$ (where the $\mathbf{u}_i$ are the columns of $\mathbf{U}$ and the N_o observed $\mathbf{y}$ are the columns of $\mathbf{Y}$):

$$I(\mathbf{U}, \mathbf{Y}) = H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{U}) = H(\mathbf{U}) - H(\mathbf{U}|\mathbf{Y}) \tag{3}$$

then a watermarking scheme is secure if the mutual information between $\mathbf{U}$ and $\mathbf{Y}$ is null: in this case there is no information leakage.

2.3 Spread spectrum carriers estimation

As mentioned previously, a SS watermarking scheme is secure if it is impossible to estimate the secret carriers using observed signals. On the contrary, if a given technique enables to estimate the secret carriers $\mathbf{u}_i$ based only on the observations of $\mathbf{y}$, then the security of the watermarking scheme is greatly reduced. Such a tool can be provided using BSS theory. The goal of BSS is to decompose the observations as a mixture of signals having special statistical properties.

For example, a Principal Component Analysis decomposes observations into orthogonal components according to their variances, and an Independent Component Analysis decomposes the observations into independent signals. Making connections between BSS and SS watermarking is straightforward. A noteworthy property of the class of spread spectrum watermarking schemes is the fact that the embedding part of a SS scheme can be formulated exactly as a blind source separation problem:

$$\mathbf{Y} = \mathbf{X} + \mathbf{U}\mathbf{S}_{\mathbf{m}}. \tag{4}$$

In this equation, the matrix $\mathbf{U}$ is an $N_v \times N_c$ matrix called the mixing matrix (in a BSS framework) and which represents in our case the different carriers $\mathbf{u}_i$ column-wise. The matrix $\mathbf{S_m}$ denotes the different sources that have to be extracted and represents the modulation signal for each carrier that is a function of the embedded message $\mathbf{m}$. The matrix $\mathbf{X}$ represents the host signals (column-wise) and shall be considered as noise in a BSS framework. The goal of BSS is to estimate the matrices $\mathbf{U}$ (the secret carriers) and $\mathbf{S_m}$ according to the observation matrix $\mathbf{Y}$.

In the case of WOA, this decomposition is possible thanks to the fact that each message is embedded independently from another one. It is consequently possible to use Independent Component Analysis techniques to break the security of many spread spectrum watermarking schemes.

The principle of ICA techniques is to find directions in the observed data space whose projections give singular distributions [7]. Based on the fact that the sum of independent variables tends to a Gaussian law, independent components are defined as the most "non-gaussian" components. Moreover, if the watermark components have a different variance than the host data, principal component analysis can be used to perform a reduction of dimension that makes the search of independent components easier.

We now focus on the estimation process of the secret carriers for two popular SS schemes (classical SS and ISS). The ability to estimate their secret carriers is presented.

Classical Spread Spectrum watermarking: In the case of classical SS, the embedding is given for each vector by:

$$\mathbf{y} = \mathbf{x} + \sum_{i=1}^{N_c} \mathbf{b}(i)\mathbf{u}_i \tag{5}$$

where $\mathbf{b} \in \{-1; +1\}^{N_c}$ is the BPSK modulation of the embedded message $\mathbf{m}$.

Improved Spread Spectrum watermarking (ISS): ISS was proposed in [8], it can be considered an informed-embedding variation of classical SS. The embedding is given in this case by:

$$\mathbf{y} = \mathbf{x} + \sum_{i=1}^{N_c} (\alpha \mathbf{b}(i) - \lambda \frac{z_{\mathbf{x},\mathbf{u}_i}}{||\mathbf{u}_i||})\mathbf{u}_i \tag{6}$$

where α and λ are respectively calculated to respect the targeted distortion and to achieve the most little error probability after addition of white Gaussian noise.

Eq. 5 and Eq. 6 can be easily transposed in the multidimensional case to obtain a formulation similar to Eq. 4: the matrix $\mathbf{U}$ still contains the carriers and the matrix $\mathbf{S_m}$ contains the modulation signals for each carrier.

To assess the security of a SS-based technique, we have decided to adopt the following methodology which is generally used in BSS benchmarks:

1. We generate N_o observations of watermarked contents and generate the matrix of observations $\mathbf{Y}$.
2. We whiten the observed signals using principal component analysis. To reduce the searching time, a reduction of dimension is therefore performed. If we consider that the host signal is generated from an i.i.d. process, the subspace containing the watermark generated by N_c carriers will be included into a N_c-dimensional space of different variance [9]. We consequently select the subspace generated by eigenvectors presenting singular (lower or higher) eigenvalues.
3. We run the FastICA algorithm [10] on this subspace to estimate the independent components and the independent basis vectors (e.g. the secret carriers).
4. We compute the normalized correlation c between each original and estimated carriers. A value of c close to 1 means that the estimation of the component is accurate. An estimation close to 0 means that the estimation is erroneous. If $N_c = 2$, we evaluate the estimation accuracy by plotting a 2D constellation of points of coordinates $(c_1; c_2)$. A successful estimation will then provide a point close to one of the four cardinal points $(0, 1)$, $(0, -1)$, $(1, 0)$, $(-1, 0)$[3].

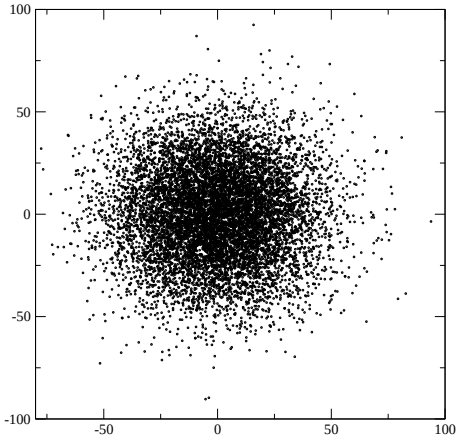

Fig. 1. Joint distributions of two carriers for original contents. $N_o = 10000$, $N_v = 512$, $\sigma^2_{\mathbf{x}} = 1$.

We have depicted the empirical joint distributions of two carriers in the observed watermarked signals (see Fig. 2) and applied our estimation setup for $N_c = 2$ secret carriers (see Fig. 3). In both cases the host vectors are Gaussian

[3] We use $N_c = 2$ for illustration purposes, hiding more bits would require to use the Hungarian method [11] to assign original and estimated carriers prior to the computation of the normalized correlation c.

i.i.d. signals of law $\mathcal{N}(0,1)$ and two carriers where used during the embedding. The Watermark to Content Ratio (WCR) was fixed in both case to $-21dB$. For each SS scheme, the joint distribution of the carriers in the observed content is the sum of four bi-dimensional Gaussian distributions. Note that the variance of each distribution is less important for ISS embedding than for SS because of the embedding optimization performed by ISS. Note also that the global variance of the distribution for SS is more important than for ISS. For each distribution, the directions of the two carriers (horizontal and vertical axis) are easily identifiable by the ICA algorithm.

Additionally we see that the Cachin criterion which considers the distributions of original and watermarked contents is not fulfilled for this two schemes. Fig. 1 depicts the distribution of the two same carriers for original contents. This distribution is rather different than the distributions after SS watermarking.

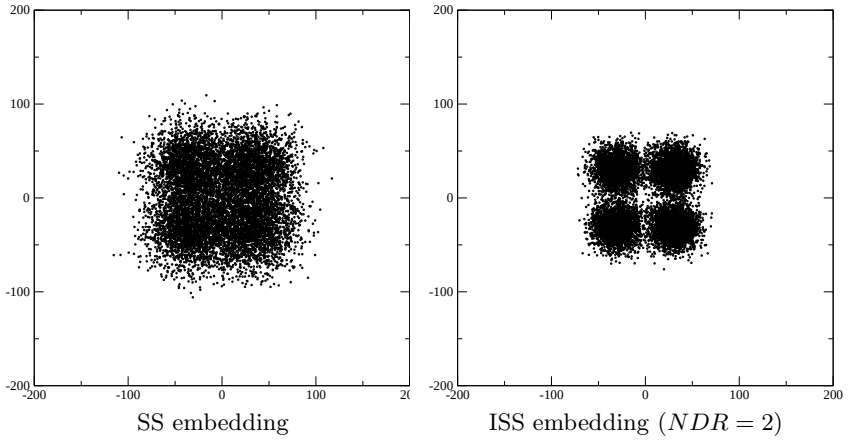

Fig. 2. Joint distributions of two carriers for SS and ISS schemes. For both schemes $N_o = 10000$, $WCR = -21dB$ and $N_v = 512$, $\sigma^2_{\mathbf{x}} = 1$.

Fig. 3 depicts the normalized correlation between the original and estimated carriers for 100 trials considering every 1000 watermarked vectors. We can notice that the estimations are globally more accurate for SS than for ISS. In this case, this is mainly due to the fact that the variance of the embedding for ISS is lower than for SS and consequently the estimation of the subspace relative to the watermark is less accurate in the second case.

3 Natural Spread Spectrum Watermarking

In this section, we show how to build a SS-based watermarking scheme in such a way that the identification of the watermarking subspace spanned by the secret carriers is impossible.

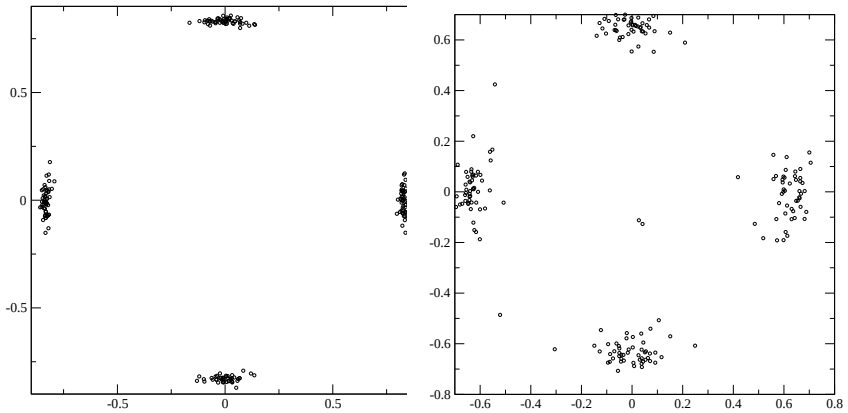

Fig. 3. Normalized correlations between the two estimated carriers and the real ones. For both schemes $N_o = 1000$, $WCR = -21dB$ and $N_v = 512$.

3.1 Embedding and decoding

If we consider a carrier $\mathbf{u}_i$ such that its coefficients follows a Gaussian model ($\mathbf{u}_i \sim \mathcal{N}(0, \sigma^2_{\mathbf{u}_i})$), and host signals $\mathbf{x} \sim \mathcal{N}(0, \sigma^2_{\mathbf{x}})$ then the random variable $z_{\mathbf{x},\mathbf{u}_i}$ follows a Gaussian law of parameters $z_{\mathbf{x},\mathbf{u}_i} \sim \mathcal{N}(0, \sigma^2_{\mathbf{u}_i}\sigma^2_{\mathbf{x}})$. Note that this property is true because of the Central limit theorem as far as N_v is important. Moreover, this property is still valid if $\mathbf{x}$ does not follow a Gaussian model. The goal of NW is to design the embedding in such a way that the distribution of $z_{\mathbf{x},\mathbf{u}_i}$ before and after embedding will remain identical. Moreover, as it is shown at the end of this section, the fact that each carrier follows a Gaussian distribution prevents the estimation of the different carriers and guarantees the security of the scheme.

The decoding rule remains the same than for usual SS schemes:

$$\hat{\mathbf{m}}(i) = 1 \text{ if } z_{\mathbf{y}',\mathbf{u}_i} > 0 \tag{7}$$

$$\hat{\mathbf{m}}(i) = 0 \text{ if } z_{\mathbf{y}',\mathbf{u}_i} < 0 \tag{8}$$

The watermarked vector $\mathbf{y}$ is:

$$\mathbf{y} = \mathbf{x} + \mathbf{w} \tag{9}$$

where the watermark signal $\mathbf{w}$ is computed as follows:

$$\mathbf{w} = -\sum_{i=1}^{N_c} \left(1 + (-1)^{\mathbf{m}(i)} \text{sign}(z_{\mathbf{x},\mathbf{u}_i})\right) \frac{z_{\mathbf{x},\mathbf{u}_i}}{\|\mathbf{u}_i\|^2} \mathbf{u}_i \tag{10}$$

which means that the watermark $\mathbf{w}_i$ associated to each carrier $\mathbf{u}_i$ follows this simple embedding rule:

$$\mathbf{w}_i = 0 \quad \text{if} \quad \mathbf{m}(i) = 1 \ ; \ z_{\mathbf{x},\mathbf{u}_i} > 0 \tag{11}$$

$$\text{or} \quad \mathbf{m}(i) = -1 \ ; \ z_{\mathbf{x},\mathbf{u}_i} < 0 \tag{12}$$

$$\mathbf{w}_i = -2\frac{z_{\mathbf{x},\mathbf{u}_i}}{\|\mathbf{u}_i\|^2}\mathbf{u}_i \quad \text{if} \quad \mathbf{m}(i) = 1 \ ; \ z_{\mathbf{x},\mathbf{u}_i} < 0 \tag{13}$$

$$\text{or} \quad \mathbf{m}(i) = -1 \ ; \ z_{\mathbf{x},\mathbf{u}_i} > 0 \tag{14}$$

Eq. 10 states that $\mathbf{x}$ is symetrically modified iff $\text{sign}(z_{\mathbf{x},\mathbf{u}_i}) \neq (-1)^{\mathbf{m}(i)}$. This embedding rule is depicted on Fig. 4.

Note that the embedding rule used for NW is quite similar to the rule used for ISS. In both cases, each carrier $\mathbf{u}_i$ is modulated according to the correlation $z_{\mathbf{x},\mathbf{u}_i}$. For ISS, $z_{\mathbf{x},\mathbf{u}_i}$ is used to increase the distance between the different codewords and increase the robustness. For NW, $z_{\mathbf{x},\mathbf{u}_i}$ is used to not modify the natural distribution of the carriers and consequently to increase the security.

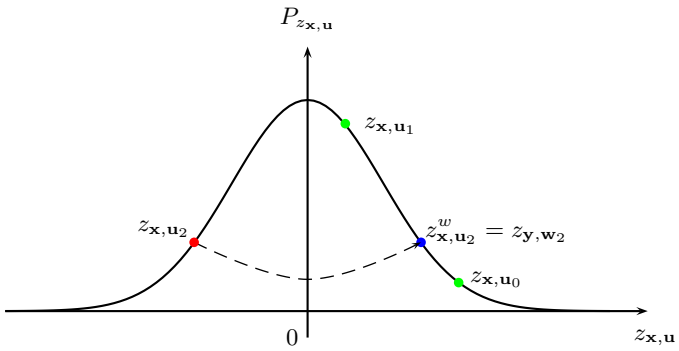

Fig. 4. Natural watermarking for $\mathbf{m} = \{1, 1, 1\}$ ($N_c = 3$). Only the third bit calls for a model-based symmetry.

3.2 Distortion

Distortion is usually expressed by means of the WCR (Watermark to Content Ratio), in dB. Under the assumption that, on average, one bit out of two calls for a symmetry, we are able to compute the MSE.

The pdf of the watermark (on each sample) is the following:

$$\text{if} \ \ x < 0 \ f(x) \sim \mathcal{N}(0, 4\sigma_x^2\sigma_u^2/N_v) \tag{15}$$

$$\text{if} \ \ x \geq 0 \ f(x) = \delta(x)/2 \tag{16}$$

Then:

$$\sigma_W^2 = \frac{2\sigma_x^2\sigma_u^2}{N_v} \tag{17}$$

and the Watermark-to-Content-Ratio is expressed as follows:

$$WCR = 10 \log \sigma_W^2 / \sigma_x^2 = 10 \log \frac{2\sigma_u^2}{N_v} \tag{18}$$

We shall stay with this last approximation, since Fig. 5 shows no difference between this theoretical approximation and the practical measurements.

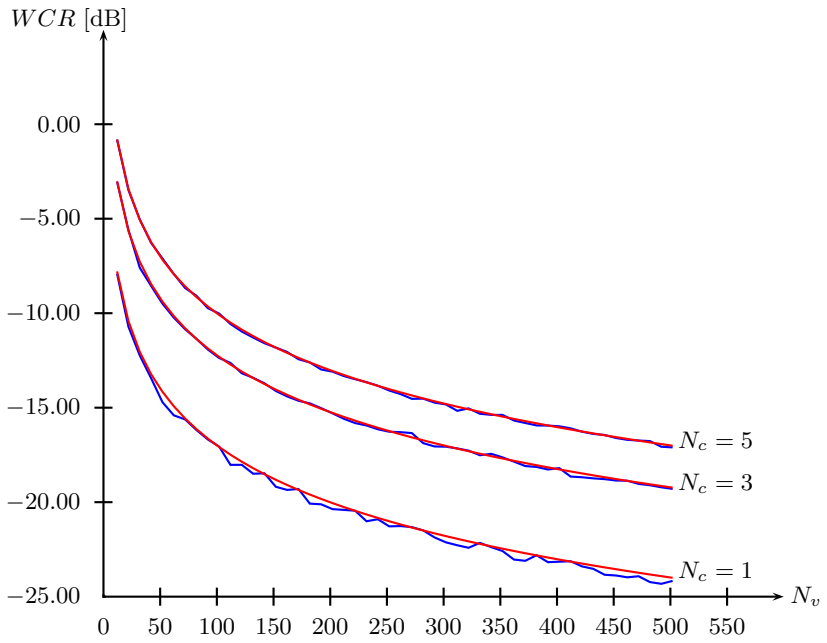

Fig. 5. Comparison between theoretical and practical WCR.

If targeting a classical $WCR = -20dB$, Eq. 18 leads to the trivial relation $N_v = 200N_c$ if we set $\mathbf{u} \sim \mathcal{N}(0,1)$ and $\mathbf{x} \sim \mathcal{N}(0,1)$.

3.3 Theoretical evaluation of the security of Natural Watermarking

The goal of this section is to show that NW enables to have no information leakage in the case of a WOA setup. We first compute the mutual information for a scheme using only one carrier $\mathbf{u}$ ($N_c = 1$: therefore the message m is a scalar binary value).

We consider the case $z_{\mathbf{x},\mathbf{u}} > 0$ but the equations are similar if $z_{\mathbf{x},\mathbf{u}} < 0$. The embedding formula can be expressed as:

$$\text{If } m = 1 \text{ then } \mathbf{y} = \mathbf{x} \tag{19}$$

$$\text{If } m = 0 \text{ then } \mathbf{y} = \mathbf{x} - 2\frac{\mathbf{u}\mathbf{u}^\top}{N_v}\mathbf{x}. \tag{20}$$

By construction, $\mathbf{y}$ can also be modelled as an i.i.d. Gaussian process: $\mathbf{y} \sim \mathcal{N}(0,1)$. Then:

$$H(\mathbf{y}) = H(\mathbf{x}) = \frac{N_v}{2}\log(2\pi e). \tag{21}$$

We recall that the mutual information between one observation $\mathbf{y}$ and a secret carrier $\mathbf{u}$ is given by:

$$I(\mathbf{y};\mathbf{u}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{u}). \tag{22}$$

Because the pdfs $f_{\mathbf{y}|\mathbf{u},m=1}(x)$ and $f_{\mathbf{y}|\mathbf{u},m=0}(x)$ are disjoint, and supposing that $\Pr[m=1] = \Pr[m=0] = 1/2$, the conditional $H(\mathbf{y}|\mathbf{u})$ is given by:

$$H(\mathbf{y}|\mathbf{u}) = \frac{1}{2}(H(\mathbf{y}|\mathbf{u}, m=1) + H(\mathbf{y}|\mathbf{u}, m=0)) + \log(2). \tag{23}$$

Using Eq.19, Eq.20 and the definition of differential entropy, it is easy to show that $H(\mathbf{y}|\mathbf{u}, m=1) = H(\mathbf{x}) - \log 2$ and $H(\mathbf{y}|\mathbf{u}, m=0) = H(\mathbf{x} - 2\mathbf{u}\mathbf{u}^\top\mathbf{x}/N_v) - \log 2$.
We obtain then the following expression of conditional entropy for NW:

$$H(\mathbf{y}|\mathbf{u}) = \frac{1}{2}H(\mathbf{x}) + \frac{1}{2}H(\mathbf{x} - 2\frac{\mathbf{u}\mathbf{u}^\top}{N_v}\mathbf{x}) \tag{24}$$

where

$$H(\mathbf{x} - 2\frac{\mathbf{u}\mathbf{u}^\top}{N_v}\mathbf{x}) = H((\mathbf{I} - 2\frac{\mathbf{u}\mathbf{u}^\top}{N_v})\mathbf{x}) = H(\mathbf{x}) + \ln|\det(\mathbf{I} - 2\frac{\mathbf{u}\mathbf{u}^\top}{N_v})|. \tag{25}$$

Because $||\mathbf{u}||^2 = N_v$, the matrix $\mathbf{I} - 2\mathbf{u}\mathbf{u}^\top/N_v$ is an elementary reflection which corresponds to an orthogonal matrix:

$$(\mathbf{I} - 2\frac{\mathbf{u}\mathbf{u}^\top}{N_v})(\mathbf{I} - 2\frac{\mathbf{u}\mathbf{u}^\top}{N_v})^\top = \mathbf{I} - 4\frac{\mathbf{u}\mathbf{u}^\top}{N_v^2} + 4\frac{\mathbf{u}\mathbf{u}^\top\mathbf{u}\mathbf{u}^\top}{N_v^4} = \mathbf{I}$$

since $\mathbf{u}^\top\mathbf{u}/N_v^2$=1. Consequently $|\det(\mathbf{I} - 2\mathbf{u}\mathbf{u}^\top/N_v)| = 1$ and we obtain:

$$H(\mathbf{y}|\mathbf{u}, m=-1) = H(\mathbf{x}) \tag{26}$$

and:

$$I(\mathbf{y};\mathbf{u}) = H(\mathbf{y}) - H(\mathbf{x}) = H(\mathbf{x}) - H(\mathbf{x}) = 0. \tag{27}$$

If we now consider N_o observations, and because $\mathbf{x}_1, ..., \mathbf{x}_{N_o}$ are independent random vectors, then due to Eq.19 and Eq.20, $\mathbf{y}_1, ..., \mathbf{y}_{N_o}$ and $\mathbf{y}_1|\mathbf{u}, ..., \mathbf{y}_{N_0}|\mathbf{u}$ are also independent vectors. Consequently we have the following properties:

$$H(\mathbf{y}_{N_o}|\mathbf{y}_{N_o-1}, ..., \mathbf{y}_1) = H(\mathbf{y}) \tag{28}$$

$$H(\mathbf{y}_1, .., \mathbf{y}_{N_o}) = N_o H(\mathbf{y}) = N_o H(\mathbf{x}) \tag{29}$$

$$H(\mathbf{y}_1, .., \mathbf{y}_{N_o}|\mathbf{u}) = N_o H(\mathbf{y}|\mathbf{u}) = N_o H(\mathbf{x}) \tag{30}$$

and finally using Eq. 29 and Eq. 30:

$$I(\mathbf{Y};\mathbf{U}) = N_o H(\mathbf{X}) - N_o H(\mathbf{X}) = 0 \tag{31}$$

Note that having a mutual information between the secret key and the observed content being null has already been found for another watermarking scheme. In [3], authors show that for the Scalar Costa's Scheme [12] and an embedding parameter $\alpha = \frac{1}{2}$, there is no information leakage too. Such a similar result is not surprising because in both cases, the distributions of $\mathbf{x}$ and $\mathbf{y}$ are the same.

Note also that natural watermarking also theoretically satisfies to the steganographic definition of security recalled in section 2.2. Because the distributions of $\mathbf{x}$ and $\mathbf{y}$ are the same, $D_{KL}(P_{\mathbf{x}}||P_{\mathbf{y}}) = 0$.

3.4 Multi-bits embedding

When $N_c > 1$ carriers are used, the security of the scheme can be seen as a problem of blind source separation: is it possible to separate a mixture of independent Gaussian variables with equal variance ?

ICA theory claims that it is impossible to perform such a separation because the joint distribution of such a mixture is rotationally symmetric [7]. Here, it is not possible for an BSS technique to find singular directions observing NW-watermarked vectors as it was possible with classical SS or ISS. Fig 6 illustrates the joint distribution of two original carriers obtained after Natural Watermarking. The distribution cannot be used to find the direction of the carriers.

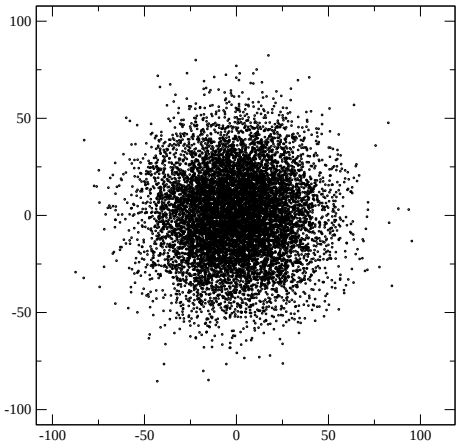

Fig. 6. Joint distributions of two carriers for NW. $N_o = 10000$, $WCR = -21dB$, $N_v = 512$, $\sigma^2_{\mathbf{x}} = 1$. The direction of the original carriers (horizontal and vertical axis) can not be estimated on a multivariate Gaussian distribution.

4 Toward robust natural watermarking

One solution to increase the robustness of the NW scheme is to increase the variance of the Gaussian distribution associated with each carrier. We can modify the embedding rule of NW (cf. Eq. 10) in such a way that the distribution of each carriers will still remain Gaussian, but will have a standard deviation proportional to a scale factor s. This modification enables to increase the average distance between codewords coding different symbols but also to increase the embedding distortion. The new modulation for Robust-NW is then given by:

$$\mathbf{w} = -\sum_{i=1}^{N_c} \left(1 + s(-1)^{\mathbf{m}(i)}\text{sign}(z_{\mathbf{x},\mathbf{u}_i})\right) \frac{z_{\mathbf{x},\mathbf{u}_i}}{\|\mathbf{u}_i\|^2}\mathbf{u}_i \tag{32}$$

and the distribution of the correlation $z_{\mathbf{x},\mathbf{u}_i}$ of one carrier $\mathbf{u}_i$ now follows the model $z_{\mathbf{x},\mathbf{u}_i} \sim \mathcal{N}(0, s^2\sigma^2_{\mathbf{u}_i}\sigma^2_{\mathbf{x}})$. If $s = 1$, the embedding rule corresponds to NW, if $s > 1$, the robustness is increased and this leads to Robust-NW. The expression of the WCR becomes:

$$WCR = 10\log\frac{(s^2+1)\sigma_u^2}{N_v} \tag{33}$$

Using Robust-NW leads also to another important consequence. Because distributions of carriers are now distinct from distributions of other components of watermarked contents, it is then possible to estimate the subspace related to the watermark using, for example, principal component analysis (PCA). But, if $N_c > 1$, it is still not possible to estimate each carrier separately. Consequently the attacker can remove the watermark by zeroing all the projections on every carriers but still he has no access to the hidden message itself: he cannot copy it to another content or modify the embedded message because he does not know the values of $z_{\mathbf{x},\mathbf{u}_i}$. Consequently the very security of the system is still preserved[4].

As a remark, we have also to point out that, even if this implementation enables to obtain robustness on one hand, on the other hand the degradation of the host signal become rather significant. For example, if $\sigma^2_{\mathbf{n}} = \sigma^2_{\mathbf{x}} = 1$ (which is equivalent to a $NDR = 1$ for ISS) and $N_v = 512$, for a same $BER = 12\%$ for both schemes, then the distortion is far more important for NW than for ISS ($WCR = -15dB$ for NW and $WCR = -21dB$ for ISS). For comparison, in the case of NW with $WCR = -21dB$, we have $BER = 25\%$.

5 Results

The aim of this section is to assess the theoretical properties of NW and Robust-NW. We have used the estimation setup proposed in section 2.3 for classical SS

[4] One may argue that the whole private subspace is already an important information about the secret. We would like to emphasise however that Robust-NW is proved to deliver better security since it explicitly does not allow for estimation of the carriers.

and ISS considering the same parameters ($N_c = 2$, $N_o = 1000$). The distortion for NW remains the same ($WCR = -21dB$) but is different for Robust-NW ($WCR = -14dB$, $s = 3$). Normalized correlations between the two estimated and original carriers are depicted on Fig. 7 for 100 different trials. For NW (left plot), the estimation of the secret carriers is unsuccessful because every point is very close to the origin for each trial. The right plot, obtained for Robust-NW, illustrates the fact that in this case the watermark subspace is estimated (the distance between each point and the origin is close to 1), but that the estimation of the two carriers is not possible because each trial leads to a point which is randomly chosen on the unitary circle.

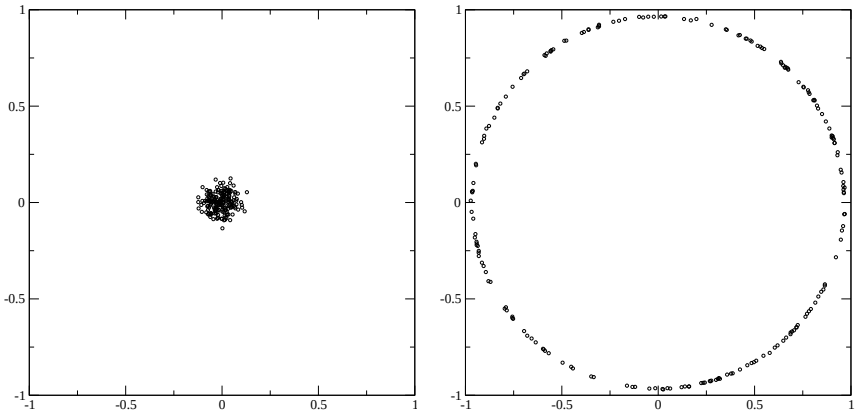

Fig. 7. Normalized correlations between the two estimated carriers and the original ones. For both schemes: $N_o = 1000$ and $N_v = 512$.

6 Conclusions and future works

This paper has presented a solution to obtain a secure spread spectrum watermarking scheme called Natural Spread Spectrum Watermarking for WOA. The security is guaranteed by two important properties:
- The distribution of each secret carrier is the same for the marked and original contents. From this property stems the fact that the mutual information between the secret carrier and the observed contents is null (condition of perfect secrecy for watermarking). Additionally, the embedding satisfies also the condition of secrecy for steganography. Note however that this first property is not true for Robust-NW ($s > 1$).
- The watermark subspace is equivalent to a mixture of Gaussian components having the same variance. BSS theory demonstrates that it is impossible to estimate each component (carrier) in this particular case.

We have also proposed a extension of NW to increase the robustness called Robust-NW which increases the variance of distributions of each carrier while preserving a multivariate Gaussian distribution. However, the compromises done to obtain a secure watermarking scheme have to balanced with the relative weak robustness of NW in comparison with SS or ISS. Our future works will concentrate on this point to discover if such a compromise is mandatory or if it is possible to obtain a more robust extension of NW. Anyway, Natural Watermarking should already be considered a secure alternative for fragile watermarking applications.

We would like also to propose a practical implementation of NW for real-life contents such as images or sounds. This will enable to evaluate specifically the perceptual distortion of the proposed system.

7 Acknowledgements

The work described in this paper has been supported (in part) by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT and the National French projects ACI-SI Fabriano and RIAM Estivale.

References

1. Cayre, F., Fontaine, C., Furon, T.: Watermarking security part I: Theory. In: Proceedings of SPIE, Security, Steganography and Watermarking of Multimedia Contents VII. Volume 5681., San Jose, USA (2005)
2. 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 (2005)
3. Pérez-Freire, L., Comesaña, P., Pérez-González, F.: Information-theoretic analysis of security in side-informed data hiding. In: 7th Information Hiding Workshop, IH05. Lecture Notes in Computer Science, Barcelona, Spain, Springer Verlag (2005)
4. Bas, P., Hurri, J.: Security of dm quantization watermarking scheme: a practical study for digital images. In: International Workshop on Digital Watermarking. Lecture notes in computer science, Sienna, Italy, Springer Verlag, Berlin, Germany (2005)
5. Cayre, F., Fontaine, C., Furon, T.: Watermarking security part II: Practice. In: Proceedings of SPIE, Security, Steganography and Watermarking of Multimedia Contents VII. Volume 5681., San Jose, USA (2005)
6. Sallee: Model-based steganography. In: International Workshop on Digital Watermarking (IWDW), LNCS. Volume 2. (2003)
7. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons (2001)
8. Malvar, Florencio: Improved spread spectrum: A new modulation technique for robust watermarking. IEEE Transactions on Signal Processing **51** (2003)
9. Doërr, G.J., Dugelay, J.L.: Danger of low-dimensional watermarking subspaces. In: ICASSP 2004, 29th IEEE International Conference on Acoustics, Speech, and Signal Processing, May 17-21, 2004, Montreal, Canada. (2004)

10. Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks **10**(3) (1999) 626–634
11. Kuhn, H. W.: The Hungarian method of solving the assignment problem. Naval Res. Logistics Quart. **2** (1955) 83–97
12. Eggers, J.J., Buml, R., Tzschoppe, R., Girod, B.: Scalar costa scheme for information embedding. IEEE Trans. on Signal Processing **51**(4) (2003) 1003–1019