Achieving subspace or key security for WOA using Natural or Circular Watermarking

Patrick Bas, François Cayre

Achieving Subspace or Key Security for WOA using Natural or Circular Watermarking

Patrick Bas [1],[2]
[1] CIS Helsinki University of Technology
P.O. Box 5400
FI-02015 HUT FINLAND
patrick.bas@hut.fi

François Cayre [2]
[2] LIS/INPG
961, rue de la Houille Blanche BP 46
38042 St. Martin d'H res Cedex, France
cayre@inpg.fr

ABSTRACT

This paper presents two watermarking schemes that are secure when considering the Watermarked content Only Attack (WOA) framework. The definition of watermark security is first recalled and the distinction between key-security and subspace-security classes for Spread Spectrum (SS) watermarking schemes is presented afterwards. Blind source separation techniques are also recalled as a tool to assess the security of a SS watermarking scheme and prove that classical SS and Improved SS are not secure within the WOA framework. To further illustrate these security issues, we next build on a new watermarking scheme called Natural Watermarking (NW). We prove it to be subspace-secure under specific hypotheses. Natural watermarking does not change the Gaussian natural distributions of the projection of each carriers. Furthermore NW prevents estimations both of the watermark subspace (subspace-security) and the different carriers (key-security). We then extend the nice properties of NW to derive the more general family of Circular Watermarking schemes which are key-secure and may offer a better robustness to AWGN attack than NW. An implementation of CW based on ISS is next proposed and comparison of bit error rates for NW, CW, SS and ISS finally draws some conclusions on the robustness cost to achieve security.

Categories and Subject Descriptors

K.4.4 [**Computers and society**]: Electronic CommerceSecurity

General Terms

Security, Algorithms, Theory

Keywords

Watermarking, key-security, subspace-security, natural watermarking, circular watermarking, WOA.

MM&Sec'06, September 26–27, 2006, Geneva, Switzerland.

1. INTRODUCTION

Beside robustness, imperceptibility and capacity, the somewhat new constraint of security shall be considered as one important requirement for watermarking schemes. The notions of security and robustness are close but still very different in essence. They can be better understood when considering the attacker point of view. We would like to emphasize on the accurate definitions proposed in [1]:

- *"attacks to robustness are those whose target is to increase the probability of error of the data-hiding channel."*
- *"attacks to security are those aimed at gaining knowledge about the secrets of the system (e.g. the embedding and/or the detection keys)."*

This last definition is coherent with the definition proposed in [2]: *"watermark security refers to the inability by unauthorised users to have access to the raw watermarking channel"*. It implies that it is not possible to either modify the embedded information or to copy it to another content if the watermarking scheme is secure. Performing an attack that estimates the secret key used for embedding and then copy the embedded message to another content using the estimated key is a threat on the security of a watermarking scheme. On the other hand, robustness is concerned by specific processes that prevent the detection of a watermark or alter randomly the embedded message. For example, noise addition or JPEG compression are processes that affect the robustness of a watermarking scheme: they may forbid the detection of the watermark or increase the bit error rate, possibly to a given level.
Additionally, it is a common mistake to claim that a watermarking technique is secure because it relies on the use of a secret key during the message embedding process. When several contents are watermarked using the same secret key, it is important to guarantee that it does not exist a way to estimate this secret key. Exhaustive search is not the only solution to perform such an estimation. Several works [3, 4, 5, 1] recently showed how to disclose keys based on observations of watermarked contents. It gave rise to the notion of watermarking security *levels*: the number of observations required for improving the estimation of the secret by an order of magnitude.

This paper deals with the two aspects presented above. Section 2 focusses on the security of Spread-Spectrum water-

marking schemes and emphasizes on the difference between the notions of key-security and subspace-security. Section 3 and 4 present the principles of Natural Watermarking (NW) and Circular Watermarking (CW) that are respectively subspace-secure and key-secure. Implementations of NW and CW are also presented and the robustness of each scheme is compared with the robustness of classical SS and ISS.

2. SECURITY OF SPREAD SPECTRUM WATERMARKING SCHEMES

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 $\mathbf{u}_i$:

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

Moreover, we define the different matrices:

- $\mathbf{Y}$ is the matrix of observed watermarked signals. Its size is N_v N_o. Each column $\mathbf{y}_i$ of $\mathbf{Y}$ represents one observed signal,
- $\mathbf{X}$ is the matrix of host signals. Its size is N_v N_o. Each column $\mathbf{x}_i$ of $\mathbf{X}$ represents one host signal,
- $\mathbf{U}$ is the matrix of secret carriers. Its size is N_v N_c. Each column $\mathbf{u}_i$ of $\mathbf{U}$ represents one secret carrier,
- $\mathbf{S}$ is the matrix of carriers modulation. Its size is N_c N_o. Each row $\mathbf{s}_i$ of $\mathbf{S}$ represents the modulation of the i^{th} bit for each watermarked content.

Because all the schemes that are presented in this paper belong to the class of SS watermarking schemes, the decoding rule is given by:

$$\hat{\mathbf{m}}(i) = 1 \quad \text{if} \quad z_{\mathbf{y}',\mathbf{u}_i} > 0 \quad (2)$$
$$\hat{\mathbf{m}}(i) = 0 \quad \text{if} \quad z_{\mathbf{y}',\mathbf{u}_i} < 0 \quad (3)$$

2.2 Security classes: subspace-secure, key-secure, insecure

In this paper, we shall emphasise on security *classes*: a watermarking scheme either belongs to the subspace-secure, the key-secure or the insecure class (by decreasing security order). Security is to be assessed within a *framework* (WOA, KMA or KOA [3]). If, within the given framework, the watermarking scheme belongs to the unsecure class, one may then compute its security *level*. WOA (Watermark Only Attack) is when the pirate has only access to watermarked contents, KMA (Known Message Attack) is when the pirate has access to watermarked contents and the corresponding hidden message, and finally KOA (Known Original Attack) is when the pirate has access to pairs of watermarked and corresponding original contents.

In the sequel, we assume that the attacker have access to a set of contents that have been watermarked using the same key but different messages. This framework has been called WOA (Watermarked content Only Attack) [3].

For SS schemes the secret key is defined by the set of carriers that are used to convey the secret message. Consequently, breaking the security of such schemes implies the estimation of the different carriers that are used to modulate each bit[1]. Note that, because of the embedding equiprobability and unknown a-priori bit ordering, the estimation of the set of carrier may lead to the estimation of any matrices $\mathbf{U}_\pi = \pi(\{\pm\mathbf{u}_i\})$ and $\mathbf{S}_\pi = \pi(\{\pm\mathbf{s}_i\})$ where $\pi()$ is a permutation function.
An important question naturally arise: whether a scheme can be considered secure if it is not possible to estimate a matrix $\mathbf{U}_\pi$, but if it is still possible to estimate the secret subspace which corresponds to the span of all vectors $\mathbf{u_i}$?
To answer this question we need to distinguish between several classes of security for Spread Spectrum watermarking schemes: *subspace-security*, *key-security* and *insecurity* as defined below:

- Definition of subspace-security:
  A watermarking scheme is *subspace-secure* if it is impossible to estimate, from the observations $\mathbf{Y}$, the secret subspace, of which one orthogonal basis is represented by the columns of the matrix $\mathbf{B}$.

  Corollary:
  A watermarking scheme is not subspace-secure if it is possible to decompose $\mathbf{Y}$ into:
  $$\mathbf{Y} = \hat{\mathbf{X}} + \hat{\mathbf{B}}\hat{\mathbf{W}} \quad (4)$$
  where the estimated matrices satisfy: $\lim_{N_o\to+\infty} \hat{\mathbf{X}} = \mathbf{X}$, $\lim_{N_o\to+\infty} \hat{\mathbf{B}} = \mathbf{B}$. The estimated matrix $\hat{\mathbf{W}}$ represents the projections of the watermark signals on the basis associated to $\hat{\mathbf{B}}$. The true carriers $\mathbf{u}_i$ therefore lie somewhere in $\text{span}(\mathbf{u}_i) = \text{span}(\mathbf{b}_i)$.

- Definition of key-security:
  A watermarking scheme is *key-secure* if it is impossible to estimate the secret key, *e.g.* the set of carriers represented by a matrix $\mathbf{U}_\pi$, from the observations $\mathbf{Y}$.

  Corollary:
  A watermarking scheme is not key-secure if it is possible to decompose $\mathbf{Y}$ into:

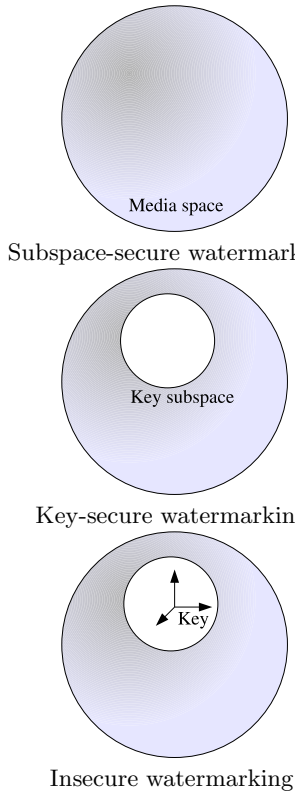
[1]Each carrier $\mathbf{u}_i$ is supposed to be associated with the bit $\mathbf{b}_i$ that is transmitted. In particular, informed coding strategies are not considered in this paper.

$$\mathbf{Y} = \hat{\mathbf{X}} + \hat{\mathbf{U}}_\pi \hat{\mathbf{S}}_\pi \quad (5)$$

where the estimated matrices satisfy: $\lim_{N_o \to +\infty} \hat{\mathbf{X}} = \mathbf{X}$, $\lim_{N_o \to +\infty} \hat{\mathbf{U}}_\pi = \mathbf{U}_\pi$, $\lim_{N_o \to +\infty} \hat{\mathbf{S}}_\pi = \mathbf{S}_\pi$.

- Definition of insecurity:
  A watermarking scheme is *insecure* if it is not key-secure.

Note that subspace-secure schemes are also key-secure. Subspace-security differs from key-security in the way that if only key-security is achieved, it can still be possible to estimate basis vectors that span the subspace in which are included the secret carriers. If a scheme satisfies subspace-security, it implies that the watermarking algorithm is totally secure: according to the definition of security, it is not possible to obtain information on the secret key. If an algorithm is key-secure but not subspace-secure, it means that it is possible to copy or erase the watermark, but that it is not possible to decode the embedded message or to modify targeted bits. Note that the identification of the watermark subspace can also be used to improve the efficiency of a specific attack affecting the robustness of the algorithm by focusing only on the most useful information. The definitions of key-security, subspace-security and insecurity classes are illustrated on Fig. 1.



Subspace-secure watermarking.

Key-secure watermarking.

Insecure watermarking.

Figure 1: Representations of the possible observations for the three different classes of security.

2.3 Using BSS techniques for subspace and key estimation

In the case of spread spectrum watermarking, the secret key is equivalent to the set of secret sequences $\{\mathbf{u_i}\}$. Each secret sequence is added to the host signal (after a bit-wise modulation) to convey a message. Eventually, the secret key is only hidden by the host signal which suggests that denoising algorithms may be used to estimate the watermark.
Techniques that are used to prove that an algorithm is not key-secure are usually very similar to techniques proving that a scheme is not subspace-secure. Key-security of SS schemes was addressed in [6] and subspace-security was addressed in [4]. A general methodology to estimate both key and subspace security is described below.
Decomposition (5), that relates to key-security, and decomposition (4), that relates to subspace-security, can be performed using Blind Source Separation (BSS) techniques if we assume some statistical properties of the watermark signal. In the latter case, we will be able to estimate the true carriers, while in the former, we will only be able to disclose a basis of the secret subspace.

We first present techniques that can achieve decomposition (5). Because $\mathbf{s}_i$ of $\mathbf{S}$ represents the modulation of the i^{th} bit for each watermarked content, and because in the WOA setup each embedded bit is supposed to be independent from others, $\{\mathbf{s}_i\}_{i \in [1,\dots,N_c]}$ can be considered as independent signals. In this case, Independent Component Analysis (ICA) techniques can be used to estimate both $\hat{\mathbf{U}}_\pi$ and $\hat{\mathbf{S}}_\pi$. 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. In the ICA setup, $\hat{\mathbf{U}}_\pi$ is called the mixing matrix and $\hat{\mathbf{S}}_\pi$ the matrix of independent sources. $\mathbf{X}$ is considered as noise.

To assess the key-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. A reduction of dimension is therefore performed to reduce the searching time. If we consider that each 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 [4]. We consequently select the subspace generated by eigenvectors presenting singular (lower or higher) eigenvalues.

3. We run the FastICA algorithm [8] 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. For $N_c = 2$, we may 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 $(0,1)$, $(0,-1)$, $(1,0)$, $(-1,0)$[2].

ICA techniques will not be efficient if $\{\mathbf{s}_i\}_{i\in[1,\dots,N_c]}$ are dependent signals. In this case Principal Component Analysis can be used to estimate the watermark subspace. We can for example use the reduction of dimension of the initial observed space as presented in the second step to estimate the watermark subspace and one of its basis $\hat{\mathbf{B}}$.

2.4 Examples of insecure SS schemes

In this section we briefly describe two popular watermarking schemes (classical SS and improved SS) and test the proposed methodology to perform estimation of secret carriers. In the case of classical SS, the embedding is given for each vector by:

$$\mathbf{y} = \mathbf{x} + \sum_{i=1}^{N_c} \mathbf{b}(i)\mathbf{u}_i \tag{6}$$

where $\mathbf{b} \in \{-1;+1\}^{N_c}$ is the BPSK modulation of the embedded message $\mathbf{m}$.
For ISS, proposed in [10], which can be considered as an informed-embedding variation of classical SS, the embedding is given by:

$$\mathbf{y} = \mathbf{x} + \sum_{i=1}^{N_c} (\alpha\mathbf{b}(i) - \lambda\frac{z_{\mathbf{x},\mathbf{u}_i}}{||\mathbf{u}_i||})\mathbf{u}_i \tag{7}$$

where α and λ are respectively computed to minimise both the targeted average distortion and the error probability after addition of white Gaussian noise of a given variance.

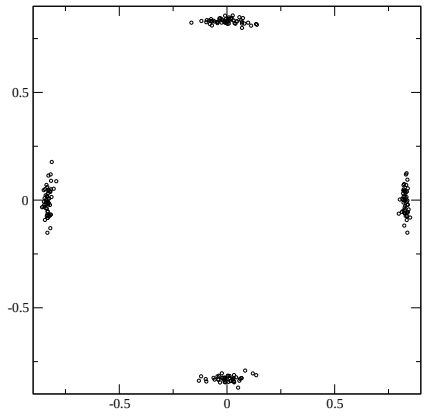
Eq. 6 and Eq. 7 can be easily transposed in the multi-dimensional case to obtain a formulation similar to Eq. 5: the matrix $\mathbf{U}$ still contains the carriers and the matrix $\mathbf{S_m}$ contains the modulation signals for each carrier.

We have applied the ICA-based carrier estimation method described in the previous section. Fig. 2 depicts the normalized correlation between the original and estimated carriers for 100 experiments considering each time 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. Both SS and ISS were experimented at the level of distorsion: the Watermark-to-Content Ratio (WCR) was set to $-21dB$.

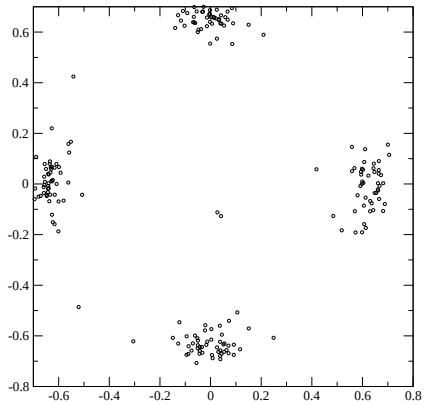
3. NATURAL WATERMARKING

Natural Watermarking is a spread-spectrum watermarking scheme that is designed to ensure subspace-security against WOA attacks for 0-centered i.i.d symmetric distributions.

[2]We use $N_c = 2$ for illustration purposes, dealing with more bits would require to use the Hungarian method [9] to assign original and estimated carriers prior to the computation of the normalized correlation c.



SS embedding



ISS embedding ($NDR = 3dB$, see [10] for definition of Document-to-Noise Ratio)

Figure 2: Normalised correlations between the two estimated carriers and $\mathbf{u}_i \sim \mathcal{N}(0,1)$ the real ones. For both schemes, $N_o = 1000$, $WCR = -21dB$ and $N_v = 512$.

We consider here i.i.d. N_v-dimensional Gaussian host signals [11].
The principles that are the key concepts of NW are the following:

1. The fact that that the projection a each secret carrier on different host signals follows a Gaussian distribution (this is due to the Central Limit Theorem).
2. The proposed scheme is denoted "Natural" because the N_v-dimensional distributions of host and watermarked signals remain the same in the WOA context: embedding keeps the natural distribution of the host signal.
3. Point (1) (Gaussianity) prevents the estimation of each carrier and point (2) prevents the estimation of the watermark subspace.

Note that, in [11], the theoretical assessment of the security class of NW is done and it is proved that there is no information leakage after embedding for NW: $I(\mathbf{Y};\mathbf{U}) = 0$. NW can be seen as the Spread Spectrum equivalent of the Scalar Costa's Scheme [12] of a uniform host signal and a distortion parameter $\alpha = 0.5$ [5] because both schemes provide no information leakage about the secret key.

3.1 Embedding

For the sake of simplicity we consider N_c orthogonal carriers that are generated using a Gaussian generator following the law $\mathcal{N}(0, \sigma_{\mathbf{u}}^2)$ for each sample. Note that the distribution of each sample of the different carriers does not modify the properties of natural watermarking and that other distributions can be considered.
The host signals are also generated according to a Gaussian law: each sample of each observation follows $\sim \mathcal{N}(0, \sigma_{\mathbf{x}}^2)$. Consequenlty the random vector $\mathbf{x}$ follows a N_v-dimensional Gaussian law. The random variable $z_{\mathbf{x},\mathbf{u}_i}$ follows a Gaussian law of parameters $z_{\mathbf{x},\mathbf{u}_i} \sim \mathcal{N}(0, \frac{\sigma_{\mathbf{u}_i}^2 \sigma_{\mathbf{x}}^2}{N_v})$. This property is true because of the Central Limit Theorem as far as N_v is important. Moreover, this property is still valid if the various $\mathbf{x}$ are 0-centered and i.i.d.. The goal of NW is to design the embedding in such a way that the distribution of $z_{\mathbf{y},\mathbf{u}_i}$ will keep the same as the distribution of $z_{\mathbf{x},\mathbf{u}_i}$.

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

$$\mathbf{y} = \mathbf{x} + \mathbf{w} \tag{8}$$

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

$$\mathbf{w} = -\sum_{i=1}^{N_c} \left(1 + (-1)^{\mathbf{m}(i)} \text{sign}(z_{\mathbf{x},\mathbf{u}_i})\right) \frac{z_{\mathbf{x},\mathbf{u}_i}}{\|\mathbf{u}_i\|^2} \mathbf{u}_i \tag{9}$$

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

$$\mathbf{w}_i = 0 \quad \text{if} \quad \mathbf{m}(i) = 1 \ ; \ z_{\mathbf{x},\mathbf{u}_i} > 0 \tag{10}$$

$$\text{or} \quad \mathbf{m}(i) = 0 \ ; \ z_{\mathbf{x},\mathbf{u}_i} < 0 \tag{11}$$

$$\mathbf{w}_i = -2\frac{z_{\mathbf{x},\mathbf{u}_i}}{\|\mathbf{u}_i\|^2}\mathbf{u}_i \quad \text{if} \quad \mathbf{m}(i) = 1 \ ; \ z_{\mathbf{x},\mathbf{u}_i} < 0 \tag{12}$$

$$\text{or} \quad \mathbf{m}(i) = 0 \ ; \ z_{\mathbf{x},\mathbf{u}_i} > 0 \tag{13}$$

Eq. 9 states that $\mathbf{x}$ is symmetrically modified iff $\text{sign}(z_{\mathbf{x},\mathbf{u}_i}) \neq (-1)^{\mathbf{m}(i)}$. This embedding rule is depicted on Fig. 3.

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.

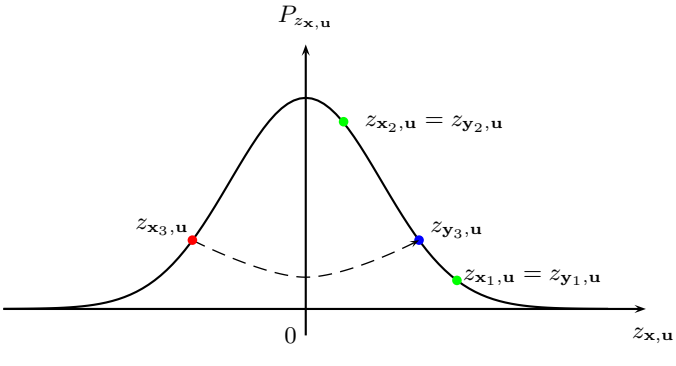


Figure 3: Natural watermarking for embedding of one bit equal to 1 on three different contents. Only the third bit calls for a model-based symmetry.

3.2 Advantages and drawbacks of NW

NW presents interesting properties:

- It is not possible to estimate the secret keys because the joint distribution of the carriers in the watermarked contents is circular. Fig. 4 presents the joint distribution of two carriers. As we can see, it is not possible to find the directions that are associated to each carrier.
- The host distribution is not modified after embedding and there is no information leakage. The ICA-based estimation attack described in Sec. 2.3 is not efficient to estimate the secret carriers. Fig. 5 presents the normalised correlation between the estimated carriers and the real ones after NW embedding for 100 experiments. The concentration of points close to the origin illustrates the fact that the estimation is not possible.

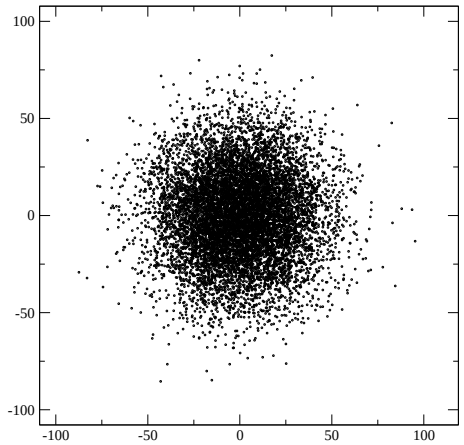


Figure 4: Joint distributions of two carriers for NW. $N_o = 10000$, $WCR = -21dB$, $N_v = 512$, $\sigma_{\mathbf{x}}^2 = 1$. The direction of the original carriers (horizontal and vertical axis) cannot be estimated on a multivariate Gaussian distribution.

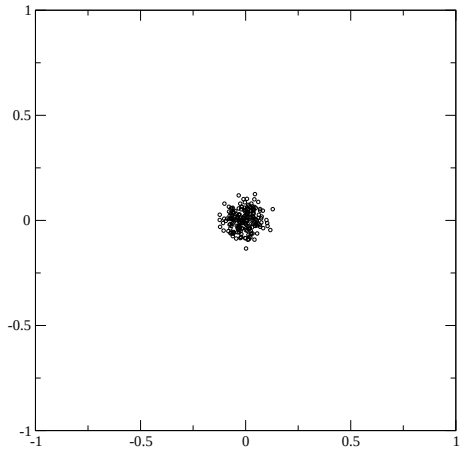


Figure 5: Normalised correlations between the two estimated carriers and the original ones. For both schemes: $N_o = 1000$ and $N_v = 512$, $WCR = -21dB$.

Beside these theoretical advantages, NW also unfortunately leads to several drawbacks:

- The robustness of NW to Additive White Gaussian Noise (AWGN) addition attack is poor for low Watermark-to-Noise ratios (WNR). Fig. 6 depicts the bit error rate (BER) for ISS, NW and SS with respect to σ_N^2 for $WCR = -21dB$ and the same N_c and N_v. In this case, NW is the less robust scheme for $\sigma_N^2 > 0.06$. For low noise addition however, NW is less sensitive to noise than SS. This is due to the fact that NW is built is such a way that the interference with the host signal is taken into account within the embedding function.

- The embedding for NW cannot guarantee that the power of the added carrier is in a given range. If the distortion is null for half of the bits; for the other half the distribution of the embedding strength for different watermarked contents follows a half-Gaussian distribution. Consequently, it is possible to have very important distortion due to the embedding on several contents (with very low probability however).

- The major drawback of NW is the fact that subspace-security class can only be achieved if we assume that the host signal is i.i.d. with Gaussian distribution. If it is not the case, ICA techniques will enable to estimate a N_c-dimensional subspace with multivariate Gaussian distribution which represents the watermark subspace. Consequently, if the host signal is not i.i.d. with N_c-dimensional Gaussian distributions, NW is only key-secure.

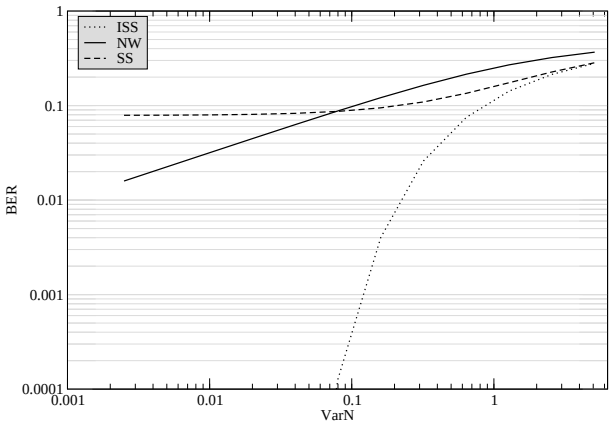


Figure 6: Comparison of BER for NW, classical SS ISS. $N_v = 512$, $WCR = -21dB$.

4. CIRCULAR WATERMARKING

4.1 Motivations

In this section, we present a new family of spread-spectrum watermarking schemes called Circular Watermarking (CW), of which NW is a special case. Looking back at NW with $N_c = 2$, it can be considered that the joint-distributions of the projection of the secret carriers considering watermaked contents are circular. The main characteristic of CW is that the joint-distributions of the secret carriers for CW keep the same property of circularity than the Gaussian joint-distributions for NW. Moreover the shape of the joint-distribution for CW has been designed to improve the robustness to AWGN addition and also to reduce the maximum possible distortion after embedding.
Another property of CW is the fact that CW schemes are no longer subspace-secure even for i.i.d. Gaussian host signals. The circular joint-distribution of the carriers after embedding offers a possibility to estimate the watermark subspace. However CW still offers key-security (the estimation of the secret carriers is not possible).

4.2 Definition of Circular Watermarking

The family of Circular Watermarking schemes is defined by the fact that the joint-distribution $f(z_1, \ldots, z_{N_c})$ of the secret carriers $\mathbf{u}_1, \ldots, \mathbf{u}_{N_c}$ is circular, which means that the distribution can be reduced to a distribution that depends only on one variable ρ:

$$f(z_1, \ldots, z_{N_c}) = f(\rho) \tag{14}$$

where $\rho = \sqrt{z_1^2 + z_2^2 + \ldots + z_{N_c}^2}$.

We can see that Natural Watermarking belongs to the family of Circular Watermarking schemes because the marginal densities of its carriers is Gaussian and then the joint-density can be written as:

$$f(z_1, \ldots, z_{N_c}) = \frac{\frac{1}{(\sigma\sqrt{2\pi})^{N_c}} \exp\left(\frac{-(z_1^2+\ldots+z_{N_c}^2)}{2\sigma^2}\right)}{\frac{1}{(\sigma\sqrt{2\pi})^{N_c}} \exp\left(\frac{-\rho^2}{2\sigma^2}\right)} \tag{15}$$

It can be shown that NW is the only scheme that provides a circular joint distribution while keeping independent modulations between carriers [13]. This means that, except for the case of NW, other CW schemes have to introduce dependency among modulation coefficients to achieve circularity.

4.3 A practical implementation based on ISS

We present here a possible solution to obtain one CW scheme which is based on Improved Spread Spectrum. The N_v-dimensional distribution of the projection of the carriers after ISS corresponds to sums of multivariate Gaussian distributions centred on cardinal points ($\alpha, \ldots, \alpha$). One solution to "circularise" this distribution after ISS is to randomly move each cardinal point on a specific region of a N_c-dimensional sphere.

To perform this, we construct a random vector $\mathbf{d}$ which is uniformly distributed on a N_c-dimensional sphere of unitary radius. One possible way to generate $\mathbf{d}$ is to first generate a N_c-dimensional Gaussian vector and to normalise it as proposed in [14]:

$$\mathbf{d} = (d_1, \ldots, d_{N_c}) = \left(\frac{g_1}{||\mathbf{g}||}, \ldots, \frac{g_{N_c}}{||\mathbf{g}||}\right) \tag{16}$$

where $g_i \sim \mathcal{N}(0, 1)$. We construct $\mathbf{d}^+ = (|d_1|, \ldots, |d_{N_c}|)$

which is a unitary random vector that is uniformly distributed on the portion of the N_c-sphere presenting positive coordinates. The embedding formula to achieve a circular implementation of ISS watermarking is:

$$\mathbf{y} = \mathbf{x} + \sqrt{N_c}\alpha \left(b(1)d_1^+\mathbf{u}_1 + \ldots + (b(N_c)d_{N_c}^+\mathbf{u}_{N_c}\right) \\ -\lambda(\tfrac{z_{\mathbf{x},\mathbf{u}_1}}{||\mathbf{u}_1||}\mathbf{u}_1 + \ldots + \tfrac{z_{\mathbf{x},\mathbf{u}_{N_c}}}{||\mathbf{u}_{N_c}||}\mathbf{u}_{N_c}) \quad (17)$$

Note that by analogy with dither modulation watermarking [15], vector $\mathbf{d}^+$ can be seen as a dither vector for the family of SS watermarking schemes : the modulation of each carrier $\mathbf{u}_i$ with d_i^+ hides the singularities of the initial distribution of the projections of the carriers as the dither vector in DM hides the location of the quantisation cells.

4.4 BER comparison

Fig. 7 depicts the joint-distributions for ISS and the circular extension of ISS for a 2-bit embedding scheme. As expected, the CW-ISS scheme produces a "ring-shaped" 2D circular distribution. Each codeword is associated with one quarter of the distribution and the shape of such a distribution enables to achieve a better robustness to AWGN than NW: it is due to the fact that there are statistically less codewords which are close to the decoding borders (X and Y axis) for CW-ISS than for NW (see Fig. 4).

Fig. 8 compares the robustness of the various proposed watermarking schemes to AWGN addition for the same hidden channel rate (same N_c and same N_v). Formulas for NW and CW-ISS are given in the appendices and those for SS and ISS are taken from [10]. For CW-ISS both practical and theoretical values are plotted. We can see in this figure that the improvement of CW-ISS over NW for $N_c = 2$ is significant. Moreover, if the performance CW-ISS is still far from the performance of ISS for low noise power, CW-ISS enables to have a key-secure watermarking scheme that equals or outperform classical SS. Also note that for the case of CW-ISS we used the same parameters α and λ than for ISS. No specific optimisation of these two parameters taking into account the specific shape of the distribution has been done.

Fig. 9 illustrates the behaviour of the robustness for CW according to the number of carriers when the distortion per carrier remains constant. We can notice that our implementation of natural watermarking has a drawback: contrary to other SS watermarking schemes, the robustness of CW decreases with the number of used carriers. Practically the BER seems to converge toward a limit when the dimension goes toward infinity (BER=9.4% for $N_c = 500$) which is still smaller than the BER achieved using NW.

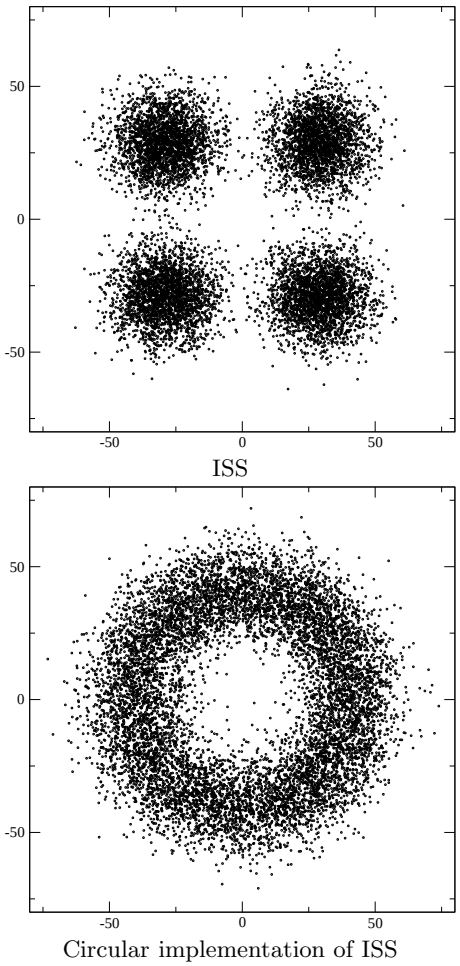


Figure 7: Joint distributions of two carriers for ISS and Circular implementation of ISS. $N_o = 10000$, $WCR = -21dB$, $\sigma_N^2 = 1$, $N_v = 512$, $N_c = 2$.

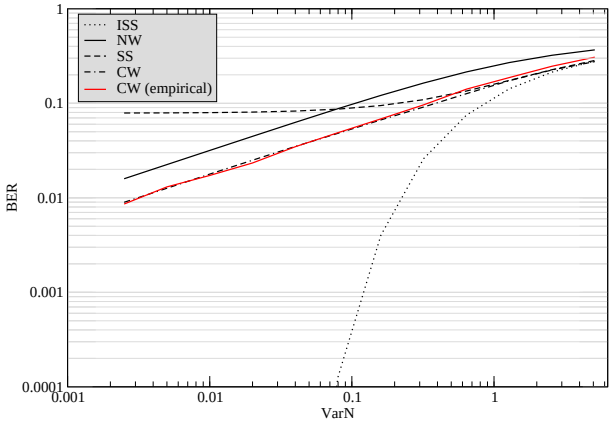


Figure 8: Comparison of BER for NW, classical SS ISS and CW-ISS. $N_v = 512$, $WCR = -21dB$ **and** $N_c = 2$.

5. CONCLUSION AND PERSPECTIVES

We have presented in this paper a classification of security as well as different ways to achieve key or subspace security for SS-based watermarking schemes. Natural Watermarking enables to obtain subspace security if the host signal is considered i.i.d. Gaussian. One very interesting property of NW is the fact that after embedding, the secret carriers have circular distributions which prevent their estimations. To obtain more robust watermarking scheme we have proposed to extend the property of circularity to other

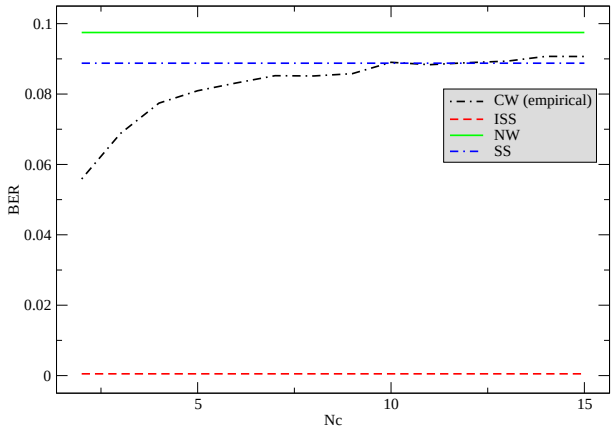


Figure 9: Evolution of BER according to the number of carriers for NW, classical SS, ISS and CW-ISS. $N_v = 512$, the distortion per carrier is constant ($WCR = -24dB$ for $N_c = 1$).

SS embedding schemes. It is important to notice that in this case, CW schemes are generally only key-secure. Our future works will focus on finding optimal embedding strategy for the family of CW schemes with respect to distortion and robustness criteria. We would like also to define the conditions to achieve security for more sophisticated SS-based watermarking schemes which use informed coding [16]. Finally, in this paper, we also wanted to outline what is the possible cost of watermarking security in terms of robustness.

6. 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, RIAM Estivale and ARA TSAR.

7. REFERENCES

[1] Pedro Comesaña, Luis Pérez-Freire, and Fernando Pérez-González. 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, June 2005. Springer Verlag.

[2] T. Kalker. Considerations on watermarking security. In *Proc. of MMSP*, pages 201–206, Cannes, France, October 2001.

[3] F. Cayre, C. Fontaine, and T. Furon. Watermarking security part I: Theory. In *Proceedings of SPIE, Security, Steganography and Watermarking of Multimedia Contents VII*, volume 5681, San Jose, USA, January 2005.

[4] Gwenaël J Doërr and Jean-Luc Dugelay. 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*, May 2004.

[5] Luis Pérez-Freire, Pedro Comesaña, and Fernando Pérez-González. Information-theoretic analysis of security in side-informed data hiding. In *7th Information Hiding Workshop, IH05*, Lecture Notes in Computer Science, Barcelona, Spain, June 2005. Springer Verlag.

[6] F. Cayre, C. Fontaine, and T. Furon. Watermarking security part II: Practice. In *Proceedings of SPIE, Security, Steganography and Watermarking of Multimedia Contents VII*, volume 5681, San Jose, USA, January 2005.

[7] A. Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[8] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[9] Kuhn, H. W. The Hungarian method of solving the assignment problem. *Naval Res. Logistics Quart.*, 2:83–97, 1955.

[10] Malvar and Florencio. Improved spread spectrum: A new modulation technique for robust watermarking. *IEEE Transactions on Signal Processing*, 51, 2003.

[11] P. Bas and F. Cayre. Natural watermarking: a secure spread spectrum technique for woa. In Neil F. Johnson, editor, *Information Hiding: 8th international workshop*, Lecture notes in computer science, Virginia, USA, 2006. Springer Verlag, Berlin, Germany.

[12] J. J. Eggers, R. Buml, R. Tzschoppe, and B. Girod. Scalar costa scheme for information embedding. *IEEE Trans. on Signal Processing*, 51(4):1003–1019, April 2003.

[13] A. Papoulis and U. Pillai. *Probability, Random Variables and Stochastic Processes*. Mac Graw Hill, 2002.

[14] Donald E. Knuth. *The Art of Computer Programming*. Vol. 2: *Seminumerical Algorithms*. Addison-Wesley, Reading, Mass., 1988. 2nd ed.

[15] Brian Chen and Gregory W. Wornell. Quantization index modulation : a class of provably good methods for digital watermarking and information embedding. In *IEEE Transaction on information theory, Vol. 47, N. 4*, pages 1423–1443, may 2001.

[16] M. L. Miller, G. J. Doerr, and I. J. Cox. Applying informed coding and embedding to design a robust, high capacity watermark. *IEEE Trans. on Image Processing*, 6(13):791–807, 2004.

APPENDIX

A. BER FOR NW

Let the noise being Gaussian with law $\mathcal{N}(0, \sigma_N^2)$ and corresponding pdf:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left(\frac{-x^2}{2\sigma_N^2}\right) \tag{18}$$

If a 0 is coded, the correlation value z has a pdf of a half-Gaussian function:

$$g(x) = \frac{2}{\sqrt{2\pi}\sigma_W} \exp\left(\frac{-x^2}{2\sigma_W^2}\right) \text{ if } x \leq 0 \tag{19}$$

$$g(x) = 0 \text{ if } x > 0 \tag{20}$$

We have the following relations (that are not used in the equations for the sake of clarity): $\sigma_N^2 = \sigma_u^2\sigma_n^2$ et $\sigma_W^2 = \sigma_u^2\sigma_x^2$.

The pdf of the watermarked signal which undergoes noise is:

$$(g * f)(t) = \int_{-\infty}^{+\infty} g(x)f(t-x)dx \tag{21}$$

which can be expressed as:

$$(g * f)(t) = A\int_{-\infty}^{0} \exp - \frac{\left(x - \frac{\sigma_W^2 t}{\sigma_N^2+\sigma_W^2}\right)^2}{2\left(\frac{\sigma_W\sigma_N}{\sqrt{\sigma_W^2+\sigma_N^2}}\right)^2} dx \tag{22}$$

with $A = 2\frac{\exp\left(\frac{-t^2}{2\sigma_N^2} + \frac{\sigma_W^2 t^2}{2\sigma_N^2(\sigma_N^2+\sigma_W^2)}\right)}{2\pi\sigma_W\sigma_N}$

after a variable substitution we can compute the integral part $I(t)$ as:

$$I(t) = \frac{\sqrt{2}\sigma_W\sigma_N}{\sqrt{\sigma_W^2+\sigma_N^2}}\frac{\sqrt{\pi}}{2}\ \text{erfc}\ \left(\frac{\sigma_W t}{\sqrt{2}\sigma_N\sqrt{\sigma_W^2+\sigma_N^2}}\right) \tag{23}$$

and finally :

$$P_e = \int_0^{+\infty} \mathcal{G}_{\sigma_W^2+\sigma_N^2}(t)\ \text{erfc}\ \left(\frac{\sigma_W t}{\sqrt{2}\sigma_N\sqrt{\sigma_W^2+\sigma_N^2}}\right) dt \tag{24}$$

with $\mathcal{G}_{\sigma_W^2+\sigma_N^2}(t) = \frac{1}{\sqrt{2\pi(\sigma_W^2+\sigma_N^2)}}\exp\left(\frac{-x^2}{2(\sigma_W^2+\sigma_N^2)}\right)$

B. DISTORTION FOR NW

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

$$\text{if} \quad x < 0 \quad f(x) \sim \mathcal{N}(0, 4\sigma_x^2\sigma_u^2/N_v) \tag{25}$$

$$\text{if} \quad x \geq 0 \quad f(x) = \delta(x)/2 \tag{26}$$

Then:

$$\sigma_b^2 = \frac{2\sigma_x^2\sigma_u^2}{N_v} \tag{27}$$

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

$$WCR = 10\log\sigma_b^2/\sigma_x^2 = 10\log\frac{2\sigma_u^2}{N_v} \tag{28}$$

C. DISTORTION FOR CW (ISS BASED IMPLEMENTATION - 2 CARRIERS)

The distortion for this implementation is the same than the distortion for classical ISS [10].

D. BER FOR CW (ISS BASED IMPLEMENTATION - 2 CARRIERS)

Let's suppose Gaussian noise with a 2D corresponding pdf:

$$f(x,y) = \frac{1}{2\pi\sigma_{Niss}^2}\exp\left(\frac{-(x^2+y^2)}{2\sigma_{Niss}^2}\right) \tag{29}$$

with $\sigma_{Niss}^2 = \sigma_N^2/(N_v\sigma_u^2)$.
The marginal pdf $f(y)$ is:

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma_{Niss}}\exp\left(\frac{-y^2}{2\sigma_{Niss}^2}\right) \tag{30}$$

After the Circular implementation of ISS we can assume that the pdf of the carriers for watermarked contents if the first coded bit corresponds to 0 is given by:

$$g(x,y) = \frac{2}{4\pi\sqrt{\pi}\sigma_{iss}\alpha_{iss}}\exp\left(-\frac{\left(\sqrt{x^2+y^2}-\sqrt{2}\alpha_{iss}\right)^2}{2\sigma_{iss}^2}\right) \quad \text{if } x \leq 0 \tag{31}$$

$$g(x,y) = 0 \text{ if } x > 0 \tag{32}$$

with α_{iss} calculated from [10] and:

$$\sigma_{iss}^2 = (1-\lambda_{iss})^2\sigma_X^2/(N_v\sigma_U^2).$$

The marginal pdf $g(y)$ is given by:

$$g(y) = \int_{-\infty}^{+\infty} g(x,y)dx \tag{33}$$

and must be computed numerically.

The pdf of the watermarked signal which undergoes noise is:

$$(g * f)(t) = \int_{-\infty}^{+\infty} g(y)f(t-y)dy \tag{34}$$

and the probability of error P_e is expressed as :

$$P_e = \int_0^{+\infty} (g * f)(t)dt \tag{35}$$

Also note that these last two expressions have to be computed numerically.