

Improved Polynomial Detectors for Side-Informed Watermarking

Jonathan Delhumeau^a, Teddy Furon^a, Neil Hurley^b, and Guenole Silvestre^b

^aIRISA/INRIA, Campus de Beaulieu, Rennes – France

^b University College Dublin, Belfield, Dublin 4 – Ireland

ABSTRACT

In spread-spectrum watermarking, the watermarked document is obtained from the addition of an attenuated watermark signal to a cover multimedia document. A traditional strategy consists of optimising the detector for a given embedding function. In general, this leads to sub-optimal detection and much improvement can be obtained by exploiting side-information available at the embedder. In some prior art, the authors showed that for blind detection of small signals, maximum detection power is obtained to first order by setting the watermark signal to the gradient of the detector. Recently, Malvar *et al.* improved the performance of direct-sequence spread-spectrum watermarking by using a signal dependent modulation. In the first part of the paper, we develop this idea further and extend Costa's decoding theory to the problem of watermarking detection. In the second part, we propose a practical implementation of this work using non-linear detectors based on our family of polynomial functions. We show some improved performance of the technique.

Keywords: Digital watermarking, Detection, Side Information

1. INTRODUCTION

This paper deals with the problem of detecting the presence of a watermark signal in a digital content. Recently, side-informed embedding strategies have been shown to greatly improve watermark detection. These strategies exploit knowledge of the original signal in the construction of the watermark vector. Such knowledge can be used to set the watermark vector direction and the watermark strength. In particular, this paper presents strategies to modulate the watermarking strength depending on the original content to be protected. The document is structured as follows. Section 2 presents a general watermarking framework, adopting the notation first introduced by Cox *et al.*¹ Section 3 reviews the prior art while Section 4 presents the watermark detection problem from a theoretical point-of-view, exploring different watermarking schemes and evaluating them using information theory tools. Section 5 describes some practical strategies that can be applied to some well-known watermarking schemes. Section 6 finally gives some experimental results, that illustrate the trade-off between robustness and embedding distortion.

2. FRAMEWORK

This section describes the mathematical framework of the problem we are dealing with and sets the notation used thereafter.

— Authors' names appear in alphabetical order.

— Jonathan Delhumeau's research is supported by the BUSMAN IST European project.

— This work was funded by Entreprise Ireland and CNRS under the Ulysses Collaboration Programme.

— Send correspondence to G. Silvestre. E-mail: guenole.silvestre@ucd.ie, Telephone: +353-1-7162852.

2.1. Modelling Contents

Various different types of content can be watermarked: sounds, images, movies, software codes. All of them are a digital representation of a work that can be handled by computers. The object of this paper is to focus on a general versatile framework applicable to all watermarking techniques, rather than to detail a specific method designed for one particular kind of contents. For this reason, we will assume that there exists a suitable feature extraction method that maps each kind of content in space \mathcal{C} to a point in the watermark space isomorphic to \mathbb{R}^N . The watermark embedding and detection take place in this space and we assume that is possible to extract N real features from an original content and to map them back to produce a watermarked content. Formally, we define the extraction process $X(\cdot)$ as:

$$\begin{aligned} X(\cdot) : \mathcal{C} &\rightarrow \mathbb{R}^N \\ C &\rightarrow \mathbf{r} = X(C) \end{aligned} \quad (1)$$

The original content and its extracted vector satisfy $\mathbf{r}_o = X(C_o)$. The inverse extraction process maps the modified features back into the content:

$$\begin{aligned} X^{-1}(\cdot) : \{\mathcal{C}, \mathbb{R}^N\} &\rightarrow \mathcal{C} \\ \{C, \mathbf{y}\} &\rightarrow C' = X^{-1}(C, \mathbf{y}) \end{aligned} \quad (2)$$

At the embedding stage \mathbf{r}_o is mapped to \mathbf{r}_w and the watermarked content is produced as $C_w = X^{-1}(C_o, \mathbf{r}_w)$.

2.2. Embedding

As mentioned earlier, the embedding process is restricted to a function $E(\cdot)$ whose domain of definition and range are \mathbb{R}^N . It is defined as follows:

$$\begin{aligned} E(\cdot) : \mathbb{R}^N &\rightarrow \mathbb{R}^N \\ \mathbf{r}_w &\rightarrow \mathbf{r}_w = E(\mathbf{r}_o) \end{aligned} \quad (3)$$

To reflect the change in \mathbf{r}_o , we specify $E(\cdot)$ as:

$$\mathbf{r}_w = E(\mathbf{r}_o) = \mathbf{r}_o + g(\mathbf{r}_o)\mathbf{w}(\mathbf{r}_o) \quad (4)$$

where $\mathbf{w}(\mathbf{r}_o)$ is a vector whose norm is set to unity, and $g(\mathbf{r}_o)$ is a scalar controlling the embedding strength. The embedding distortion is defined in expectation as:

$$\mathcal{D}_E = E\{\|\mathbf{r}_w - \mathbf{r}_o\|^2\} = E\{g(\mathbf{r}_o)^2\} \quad (5)$$

where $E\{\cdot\}$ is the statistical expectation. In addition, we assume that the extracted vectors \mathbf{r}_o define a white gaussian noise of power σ_o^2 . Watermark to Content power Ratio is then defined as $WCR = \mathcal{D}_E/N\sigma_o^2$. This ratio is extremely low (typically -26 dB), so that it is assumed that the watermarked vectors \mathbf{r}_w have a power of $\sim \sigma_o^2$.

2.3. Attack

The attacks are blind as the attacker may not have access to the extracted vectors. The impact of the attack is modelled as the addition of white gaussian noise of power is σ_n^2 followed by Wiener filtering. This is equivalent to multiplication by a factor $\rho = \sqrt{\sigma_o^2/(\sigma_o^2 + \sigma_n^2)}$ such that:

$$\mathbf{r}_p = \rho(\mathbf{r}_w + \mathbf{n}) \quad (6)$$

The effect of the ρ factor is to set the power of the attacked signal to the same level as the watermarked vector, while maintaining the Noise to Content power Ratio to $NCR = \sigma_n^2/\sigma_o^2$. Moreover, the distortion due to this attack, measured by

$$\mathcal{D}_A = E\{\|\mathbf{r}_p - \mathbf{r}_w\|^2\} = N((1 - \rho)^2\sigma_o^2 + \rho^2\sigma_n^2), \quad (7)$$

is indeed lower than the distortion without factor ρ . Clearly the attacker should perform such Wiener filtering in order to decrease the impact of the attack*.

*Thanks to Stéphane Pateux for pointing out this fact.

2.4. Detection

The detector receives an unknown content whose extracted vector is denoted \mathbf{r}_u . In the detection process, we distinguish two hypotheses: namely hypothesis H_0 that \mathbf{r}_u is an original non-watermarked vector and hypothesis H_1 that \mathbf{r}_u is an attacked watermarked vector. Under both hypotheses, the power of \mathbf{r}_u is equal to σ_o^2 .

To distinguish between H_0 and H_1 , the detector applies a function $D(\cdot)$, called the detection function, to the received vector. It yields a kind of likelihood that hypothesis H_1 is true (versus hypothesis H_0).

$$\begin{aligned} D(\cdot) : \mathbb{R}^N &\rightarrow \mathbb{R} \\ \mathbf{r}_u &\rightarrow d = D(\mathbf{r}_u) \end{aligned} \quad (8)$$

The output of the detection function is compared to a threshold T , which is set to achieve a given probability of false alarm P_{fa} . The detector declares \mathbf{r}_u is not watermarked if $d < T$. It decides it is watermarked if $d > T$.

2.5. Robustness

The idea of robustness was first introduced by Cox *et al.*,⁶ and depending on the watermarking method, there may exist a function measuring the robustness of a given point in the space. The robustness is measured as the power of independent noise to be added to lower, in expectation, $D(\mathbf{r}_w)$ below the threshold value T and defined as:

$$\begin{aligned} R(\cdot) : \mathbb{R}^N &\rightarrow \mathbb{R} \\ \mathbf{r} &\rightarrow R(\mathbf{r}) = \|\mathbf{n}\|^2 \end{aligned} \quad (9)$$

2.6. Performance

Various measures can be used to compare different watermarking methods and evaluate the performance for a given vector of length N :

1. \mathcal{D}_E , the embedding distortion (or equivalently the ratio WCR).
2. $P_{fa} = E\{(d > T)|H_0\}$, the probability of false alarm.
3. $P_p = E\{(d > T)|H_1\}$, the power of the detection test.
4. D_{KL} , the Kullback-Leibler distance between the pdf of the original vector and the watermarked vector. Section 4 details the importance of this criterion.
5. $\epsilon = \mu_{d|H_1}/\sigma_{d|H_1}$, the deflection factor of the tested statistic.
6. $\eta = (T - \mu_{d|H_1})/\sigma_{d|H_1}$, the argument of the cdf of $\mathcal{N}(0, 1)$ to calculate the power of the test in the gaussian case: $P_p = 1 - Q(\eta)$.

Criteria 1 to 3 directly reflect the performance of the test. They can be estimated experimentally by averaging a huge number of trials. Criteria 4 to 6 are not direct outputs of the test but can help in interpreting how parameters are related to yield criteria 1 to 3. Criteria 5 and 6 are more deeply concerned with the case of the tested statistic being gaussian distributed. An important issue is to determine how criteria 2 to 6 depend on the parameters $\{N, \text{WCR}, \text{NCR}\}$.

3. DESCRIPTION OF PRIOR ART

This section presents a brief overview of five methods for watermark detection; namely Direct-Sequence Spread-Spectrum (DSSS), Just Another N-order Informed Scheme (Janis), Zero Attraction (ZATT), Peaking DSSS (PEAK) and Improved Spread-Spectrum (ISS).

3.1. DSSS: Direct Sequence Spread Spectrum

In this well known method,² the watermark signal is constant in its direction and in its norm,

$$g(\mathbf{r})\mathbf{w}(\mathbf{r}) = g\mathbf{w} \quad (10)$$

The detection function is a linear correlation between \mathbf{r}_u and \mathbf{w} , given by,

$$D(\mathbf{r}_o) = \sum_{i=1}^N \frac{w[i]r_o[i]}{\|\mathbf{r}\|} \quad (11)$$

3.2. JANIS: Just Another N-order Informed Scheme

In this method,^{3,4} the watermarked vector, $\mathbf{w}(\mathbf{r}_o)$, is equal to the normalised gradient of the detection function at the point \mathbf{r}_o . Thus, this vector points in the direction $\mathbf{r}_o + \mathbf{h}$ where $D(\mathbf{r}_o + \mathbf{h})$ increases at the highest rate. The embedding strength is constant.

$$g(\mathbf{r}_o)\mathbf{w}(\mathbf{r}_o) = g \frac{\nabla D(\mathbf{r}_o)}{\|\nabla D(\mathbf{r}_o)\|} \quad (12)$$

This strategy has been analysed on a simple n^{th} -order polynomial detection function:

$$D(\mathbf{r}_o) = \sum_{i=1}^{N/n} \prod_{j=1}^n r_o[i_j] \quad (13)$$

For small orders n , $D(\mathbf{r}_o)$ is gaussian distributed under both hypotheses with different means and variances i.e. $\mathcal{N}(0, \sigma_{d|H_0}^2)$ and $\mathcal{N}(\mu_{d|H_1}, \sigma_{d|H_1}^2)$.

3.3. ZATT: Zero Attraction

In this method,⁵ $g(\mathbf{r})\mathbf{w}(\mathbf{r})$ is not constant. Basically, the embedding resets a small number of secret projections of the original vector to zero. The embedding is defined as:

$$\mathbf{r}_w = (\mathbf{I} - \mathbf{P})\mathbf{r}_o$$

where \mathbf{P} is a secret projection matrix of rank $k = \lfloor \mathcal{D}_E / \sigma_o^2 \rfloor$. Watermark embedding is achieved by removing some part of the signal. With the additive attack, the k projections are not set to zero but they are distributed as $\mathcal{N}(0, \rho^2 \sigma_n^2)$.

The detection function calculates the energy of the k projections and decides that the received vector is watermarked if this energy is below a given threshold.

3.4. PEAK: Peaking DSSS

In this method,⁶ $\mathbf{w}(\mathbf{r})$ is constant and set to \mathbf{w} , a normalised vector, but its amplitude $g(\mathbf{r})$ varies. In fact, the embedding perfectly sets the projection of the vector onto \mathbf{w} to a discrete value. It is an ‘erase and write’ strategy, as referred to by Costa.⁷ The embedding is defined as:

$$\mathbf{r}_w = \mathbf{r}_o + (\alpha - \mathbf{w}^T \mathbf{r}_o)\mathbf{w}$$

The parameter α is given by:

$$\mathcal{D}_E = \alpha^2 + \sigma_o^2$$

Clearly, to enforce this method, \mathcal{D}_E should be greater than σ_o^2 , which the energy required to reset the projection to zero. This can be stated as $N \text{WCR} > 1$. With the attack, the projection onto \mathbf{w} is not set to α but it is distributed as $\mathcal{N}(\alpha, \rho^2 \sigma_n^2)$.

The detection function is the same as DSSS, a linear correlation between \mathbf{r}_u and \mathbf{w} .

3.5. ISS: Improved Spread Spectrum

ISS is a generalisation of PEAK.⁸ In this case, $\mathbf{w}(\mathbf{r})$ is also constant and set to \mathbf{w} , a normalized vector, but its amplitude $g(\mathbf{r})$ varies. The embedding sets the projection of the vector onto \mathbf{w} to a value depending on the original vector. It makes a trade-off between the embedding distortion and the robustness. It is a ‘writing on dirty paper’ strategy and the embedding process is defined as:

$$\mathbf{r}_{\mathbf{w}} = \mathbf{r}_{\mathbf{o}} + (\alpha - \lambda \mathbf{w}^T \mathbf{r}_{\mathbf{o}}) \mathbf{w}$$

Parameters $\{\alpha, \lambda\}$ satisfy the equation:

$$\mathcal{D}_E = \alpha^2 + \lambda^2 \sigma_o^2$$

The detection function is the same as DSSS, a linear correlation between $\mathbf{r}_{\mathbf{u}}$ and \mathbf{w} . For $\alpha^2 = \mathcal{D}_E$ and $\lambda = 0$, the scheme reduces to DSSS. For $\alpha^2 = \mathcal{D}_E - \sigma_x^2$ and $\lambda = 1$, it reduces to PEAK.

4. THEORETICAL COMPARISON

4.1. Wiping out a Cliché

At a first glance, it would seem that the problem of watermark detection is simpler than the decoding of hidden symbols, because the final output of a decoding system belongs to a message space which is larger than the detection range $\{0, 1\}$. However, the authors believe this is not true for the following reasons.

Firstly, no theoretical limit has been shown for watermark detection. In the decoding problem, Costa has shown that the capacity of the watermarking channel is bounded above by the optimal capacity given by:⁷

$$C^* = \frac{1}{2} \log\left(1 + \frac{\mathcal{D}_E}{N\sigma_n^2}\right).$$

The fundamental insight is that the optimal capacity does not depend on the power of the original vector. It is equal to the capacity when the decoder is not blind. An interesting problem is to find the equivalent of this bound for the detection problem. In detection, a binary decision is required and we are not concerned with channel capacity. The detection goal is to distinguish from which of two probability distributions a received vector is drawn. A measure of the difference between these two distributions is required. We argue that the statistical Kullback-Leibler ‘distance’⁹ (D_{KL}) can play the role of capacity for detection.

What is the upper bound in terms of this measure? Copying Costa’s solution, we might believe that this bound is given when the decoder knows the original vector. In this case, $D_{\text{KL}} = +\infty$ which implies that it is possible to build a perfect test ($P_p \rightarrow 1 \quad \forall P_{fa} \in (0, 1]$) regardless of the strength of the attack. This is obvious for a non-blind detector as it already knows $\mathbf{r}_{\mathbf{o}}$. Yet, this is not possible when the detector is blind.

The fundamental difference between detection and decoding is the fact that the decoder is a multiple hypothesis test (one per hidden symbol) where the hypothesis to receive an ‘untouched’ and ‘raw’ original vector is not considered. Nevertheless, it might be possible to have a bound that only depends on $\{\mathcal{D}_E, N, \sigma_n^2\}$. However, it has not been found yet.

Secondly, Costa not only gives the theoretical bound, but also shows a way to achieve it. This method is not possible in practice, but, at least, it gives some clues to derive some good ‘dirty paper codes’.^{10, 11} Basically, the game of watermarking decoding now is to easily build, partition, and quickly browse a set of codewords in \mathbb{R}^N .¹² In watermarking detection, so far, we have absolutely no clue how to derive a good solution.

4.2. Detection Theory

The Kullback-Leibler distance between the two random processes $\mathbf{r}_{\mathbf{o}}$ and $\mathbf{r}_{\mathbf{p}}$ is defined as an integral over \mathbb{R}^N :

$$D_{\text{KL}}(\mathbb{R}^N) = \int_{\mathbb{R}^N} p_{\mathbf{H}_0}(\mathbf{r}) \log \frac{p_{\mathbf{H}_0}(\mathbf{r})}{p_{\mathbf{H}_1}(\mathbf{r})} d\mathbf{r}$$

According to the *data processing* theorem,⁹ we have:

$$D_{\text{KL}}(\mathbb{R}^N) \geq D_{\text{KL}}(\mathbb{R}) \geq D_{\text{KL}}(\{0, 1\}) = P_{fa} \log \frac{P_{fa}}{P_p} + (1 - P_{fa}) \log \frac{(1 - P_{fa})}{(1 - P_p)} \quad (14)$$

where $D_{\text{KL}}(\mathbb{R})$ ($D_{\text{KL}}(\{0, 1\})$) is the KL distance for the random variable d (resp. $(d > T)$). Hence, the performance of the detectors, in terms of $\{P_{fa}, P_p\}$ are limited by $D_{\text{KL}}(\mathbb{R}^N)$ via (14). For instance, if we are looking for a perfect test where $P_{fa} = 0$, then its power is bounded above by:

$$P_p \leq 1 - e^{-D_{\text{KL}}(\mathbb{R}^N)}$$

The power of the test tends to one only if $D_{\text{KL}}(\mathbb{R}^N)$ goes to $+\infty$. This results in a new way of considering the watermarking detection problem. The goal is to find an embedding function $E(\cdot)$ that maximises $D_{\text{KL}}(\mathbb{R}^N)$ under the distortion constraint given by (5).

4.3. Comparing Kullback-Leibler distances

Table 1 gives the KL distances for the watermarking methods considered in Section 3.

Table 1. Comparing DKL

Method	Space	D_{KL} without noise	D_{KL} with noise
DSSS	\mathbb{R}^N	$\frac{1}{2}N \text{ WCR}$	$\frac{1}{2}N \text{ WCR} \rho^2$
JANIS	\mathbb{R}	$\frac{1}{2}nN \text{ WCR}$	$\frac{1}{2} \left(\log \frac{\sigma_{d H_1}^2}{\sigma_{d H_0}^2} - 1 + \frac{\sigma_{d H_0}^2}{\sigma_{d H_1}^2} + \frac{\mu_{d H_1}^2}{\sigma_{d H_1}^2} \right)$
ZATT	\mathbb{R}^N	$+\infty$	$\frac{1}{2}N \text{ WCR} \left(\log(\text{NCR} \rho^2) - 1 + \frac{1}{\text{NCR} \rho^2} \right)$
PEAK	\mathbb{R}	$+\infty$	$\frac{1}{2} \left(\log(\text{NCR} \rho^2) - 1 + \frac{1}{\text{NCR} \rho^2} + \frac{N \text{ WCR} - 1}{\text{NCR}} \right)$

Note that ZATT, PEAK and ISS have an infinite measure of D_{KL} when there is no attack. It means that a perfect test is theoretically possible. For instance, with the ZATT method, under H_0 , d is a continuous random variable whose pdf's range is $(0, +\infty]$, whereas under H_1 , it becomes a deterministic value. The probability $P(d = 0|H_0)$ is null as the singleton 0 does not belong to the Borel extension of \mathbb{R} . The test considers the received vector to be watermarked if $d = 0$. Hence, $(P_{fa}, P_p) = (0, 1)$ when there is no attack. The ZATT and PEAK methods are particular cases of the strategy of using a quantisation process as the embedder. Roughly, this transforms a continuous random variable into a discrete random variable. Since in this case P_{fa} is the integral of a pdf over a set of discrete values, $P_{fa} = 0$. Under H_1 , it is easily shown that $P_p = 1$. Note that no requirement has been made about the quantisation grain, so that in theory, the quantisation (or embedding) distortion can tend to zero whilst maintaining the same performance.

When some noise is added, the discrete random variable is transformed into a continuous random variable which is the convolution of the pmf by the pdf of the noise, i.e. a multimodal gaussian. In the ZATT method, the test yields the binary output ($d < T$) with the threshold T set by the probability of false alarm. The performance of the test decreases with the amount of noise.

Figure 1 gives the plot of the Kullback-Leibler distances for various schemes. Whereas, for ZATT, PEAK, and ISS, D_{KL} tends to $+\infty$ when there is no attack, it quickly decreases in the presence of noise, contrary to DSSS which is very robust (D_{KL} decreases very slowly).

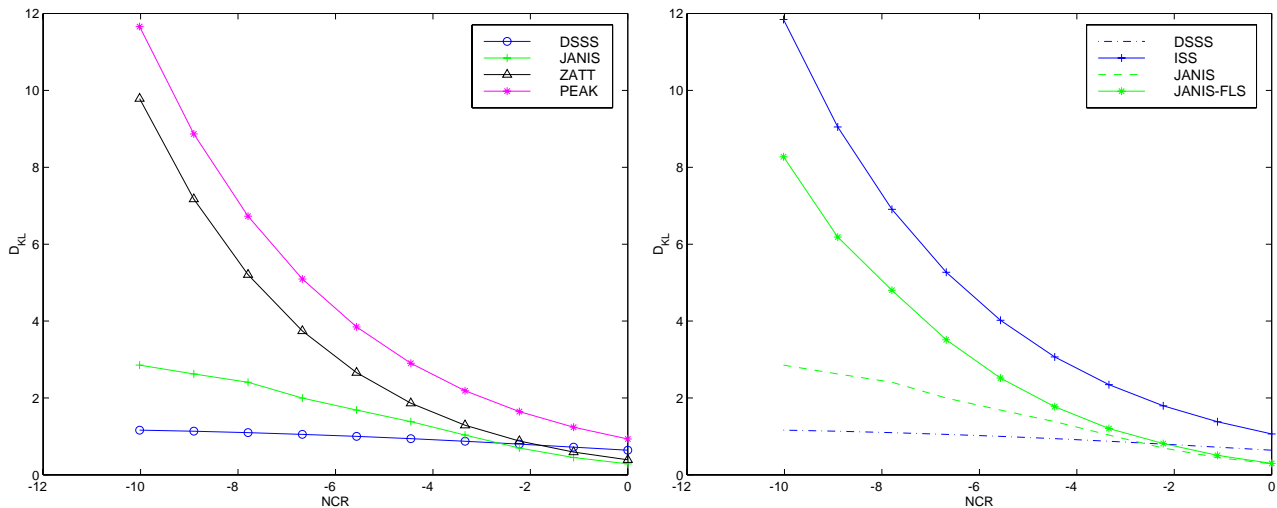


Figure 1. D_{KL} against NCR in dB, with $N \text{ WCR} = 2.56$ ($N = 1024$, $\text{WCR} = -26$ dB). (a) D_{KL} for methods of prior art (cf. Section 3). (b) Improvement due to the Florencio-Malvar embedding strategy, for DSSS and JANIS.

5. EMBEDDING STRATEGIES

This section endeavors to explore different strategies concerning the design of the function $g(\cdot)$. They were introduced by Miller *et al.* for their scheme.¹³ We aim to apply these strategies to others schemes.

It is clear that in practice, the embedding strength of the watermark signal varies and typically depends on a perceptual model, which acts like a local gain controller. In having selected an extremely simple content model free from perceptual considerations in Section 2, we would like to study various performance-related reasons for varying the embedding strength depending on the content. Importantly it should be noted that the strategies analysed here are used at the embedding stage under hypothesis H_1 and do not change anything at the detection side. The fundamental reason upon which this statement is based is the Neyman-Pearson detection strategy, where threshold T is only related to P_{fa} , i.e. what happens under hypothesis H_0 .

5.1. Maximising Detection for Fixed Distortion

The fixed distortion strategy is the most common strategy in watermarking. Both DSSS and JANIS follow this strategy as explained in Sections 3.1 and 3.2. The constraint on the embedding distortion is not only fulfilled in average but for all content:

$$g(\mathbf{r}_o) = g = \sqrt{\mathcal{D}_E} \quad \forall \mathbf{r}_o \in \mathbb{R}^N \quad (15)$$

This does not mean that the watermarking vector is fixed. In fact for JANIS, $\mathbf{w}(\mathbf{r}_o)$ is directed to the point on the N -dimensional hypersphere of radius g and centre \mathbf{r}_o , which maximises $D(\mathbf{r}_o + g\mathbf{w}(\mathbf{r}_o))$. This latter method has better performance than DSSS because, for the same embedding distortion \mathcal{D}_E , the detection yields more separated pdf's under both hypotheses. In other words, the tested statistic in the JANIS method is more sensitive to the embedding distortion than in DSSS. Two very simple ideas⁸ can be used to improve this strategy and consist of pre-processing (before the embedding process) and a post-processing (after the embedding process) steps based on the detection function:

- *pre-processing*: if $D(\mathbf{r}_o) > T$ then $\mathbf{r}_w = \mathbf{r}_o$, as there is no need to run the embedding.
- *post-processing*: if $D(E(\mathbf{r}_o)) < T$ then $\mathbf{r}_w = \mathbf{r}_o$ as the embedding process has been useless.

These ideas have an impact on the global embedding distortion which becomes smaller $\mathcal{D}_E' = \mathcal{D}_E(P_p - P_{fa})$. The pre-processing saves a small amount of distortion as it concerns original vectors yielding a probability of false alarm, which is supposed to be small. The post-processing might save a dramatic amount of distortion in the case where the power of the test is not so high (i.e. when $N \text{ WCR}$ is small). Note that both processing do not change (P_p, P_{fa}) .

5.2. Minimizing Distortion for Fixed Detection

This strategy aims to find the closest point \mathbf{r}_w to \mathbf{r}_o such that $D(\mathbf{r}_w) = T + \delta$ for some $\delta > 0$. ZATT and PEAK methods follow this strategy as explained in Sections 3.3 and 3.4.

$$\begin{aligned} E(.) : \mathbb{R}^N &\rightarrow \mathbb{R}^N \\ \mathbf{r}_o &\rightarrow \mathbf{r}_w = \arg_{D(\mathbf{r})=T+\delta} \min \|\mathbf{r} - \mathbf{r}_o\| \end{aligned} \quad (16)$$

This strategy might not be possible as there are vectors where the embedding distortion is very high. The probability to have such vectors is small. In theory, only the average embedding distortion counts. In practice, there will certainly be an upper limit on the embedding distortion. Hence, the following pre- and post-processing steps apply:

- *pre-processing*: if $D(\mathbf{r}_o) > T + \delta$ then $\mathbf{r}_w = \mathbf{r}_o$.
- *post-processing*: if $\|E(\mathbf{r}_o) - \mathbf{r}_o\|^2 > \mathcal{D}_{\max}$ then $\mathbf{r}_w = \mathbf{r}_o$

5.3. Maximising Robustness for Fixed Distortion

This strategy has been proposed by Cox *et al.*¹³ The constraint on the embedding distortion is not only fulfilled on average but for all content as in (15). This does not mean that the watermarking vector is fixed: $\mathbf{w}(\mathbf{r}_o)$ is directed to the point on the N -dimensional hypersphere of radius g and centre \mathbf{r}_o , which maximises $R(\mathbf{r}_o + g\mathbf{w}(\mathbf{r}_o))$. An expression of the robustness function is needed. For correlation based detectors such as DSSS, PEAK, ISS and also the ZATT method, the following expression holds⁶:

$$R(\mathbf{r}) = \|\mathbf{r}\|^2 \left(\frac{D(\mathbf{r})^2}{T^2} - 1 \right) \quad (17)$$

For the Janis detector, we proved that:

$$R(\mathbf{r}) = \|\mathbf{r}\|^2 \left(\left(\frac{D(\mathbf{r})}{nT} \right)^{2/n} - 1 \right) \quad (18)$$

Note these expressions are only valid for vectors belonging to the critical region, i.e. $D(\mathbf{r}) > T$.

5.4. Minimising Distortion for Fixed Robustness

This strategy aims to find the closest point \mathbf{r}_w to \mathbf{r}_o such that $R(\mathbf{r}_w) = r$.

$$\begin{aligned} E(.) : \mathbb{R}^N &\rightarrow \mathbb{R}^N \\ \mathbf{r}_o &\rightarrow \mathbf{r}_w = \arg_{R(\mathbf{r})=r} \min \|\mathbf{r} - \mathbf{r}_o\| \end{aligned} \quad (19)$$

This strategy might not be possible as there are vectors where the embedding distortion is very high. An upper limit might also be necessary as in Section 5.2.

5.5. Florencio-Malvar strategy

Florencio and Malvar give a method to find a trade-off between distortion, performance and robustness. From Section 3.5, $g(\mathbf{r})$ is modulated so that $D(\mathbf{r}_w) = \alpha - \lambda D(\mathbf{r}_o)$. For a set of parameters $\{\mathcal{D}_E, \text{WCR}, \text{NCR}\}$, λ is optimised to maximise a given criteria such as the Kullback-Leibler distance, the power of the test or the global robustness.

6. EXPERIMENTAL RESULTS

To illustrate the impact of the adopted strategy, we present results dealing with the trade-off between embedding distortion and robustness. We have developed for this purpose a Matlab toolbox available at <http://www.irisa.fr/temics/Equipe/Delhumeau/> or <http://ihl.ucd.ie/>.

6.1. Maximizing Detection for Fixed Distortion

Figure 2 gives the power plot for fixed distortion using the maximising detection strategy of Section 5.1, with pre- and post-processing for DSSS and JANIS at $n = 2, 4$. Saving distortion when the embedding fails allows us to increase the distortion for vectors whose embedding succeeds. This strategy has a higher benefit on less efficient methods such as DSSS since approximately half of the original vectors remain untouched and half of them bear an embedding distortion of ~ -23 dB, leading to a global distortion of -26 dB. As a result, the robustness has increased because the watermarked vectors are more deeply pushed inside the critical region.

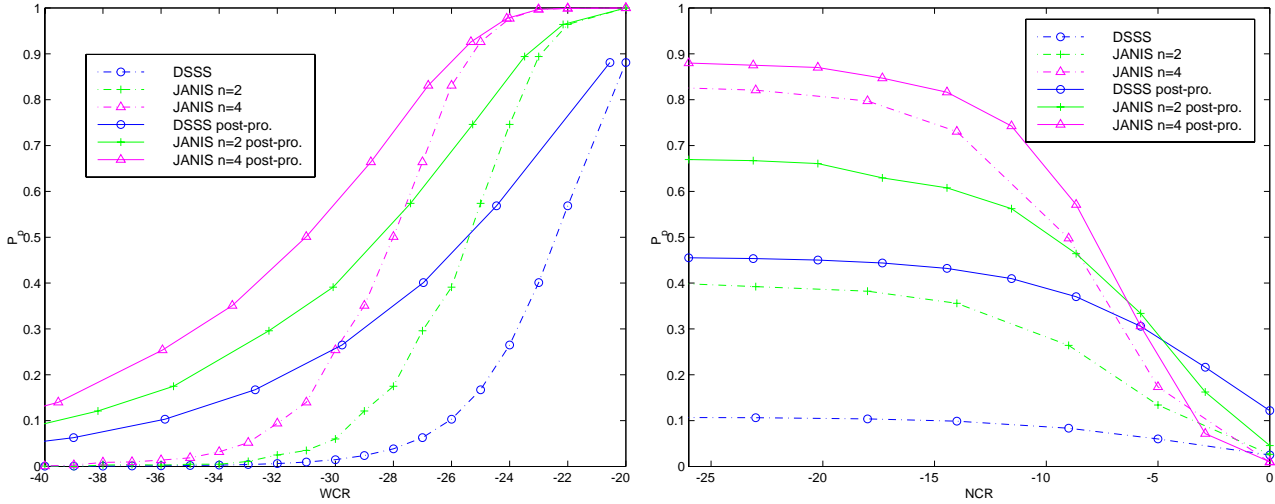


Figure 2. Maximizing detection for fixed distortion: Impact of pre- and post-processing (plain line) on prior art method (dashed line). **(a)** P_p against WCR in dB, with $N = 2400$, $P_{fa} = 10^{-4}$. **(b)** P_p against NCR in dB for an average distortion of -26 dB, and a maximum distortion of -22.7 dB (DSSS), -24.4 dB (JANIS $n=2$), -25.6 dB (JANIS $n=4$).

6.2. Minimising Distortion for Fixed Detection

To implement this strategy, the embedding is carried out using an iterative process: $\mathbf{r}_{i+1} = \mathbf{r}_i + g\nabla D(\mathbf{r}_i)$ where $\mathbf{r}_{i=0} = \mathbf{r}_o$. g is the iterative step of an equivalent WCR value of -46 dB. The iteration ends when $D(\mathbf{r}_i) > T$. Observing Figure 3a, the following comments apply: on one hand, JANIS $n = 4$ needs less embedding distortion than the others. This means that its critical area is more widely spread in \mathbb{R}^N . On the other hand, JANIS $n = 4$ is less robust to noise addition when the vectors are pushed just above the boundary of the critical region.

6.3. Maximising Robustness for Fixed Distortion

Figure 3b gives the P_p versus NCR for a fixed distortion of -26 dB using the strategy of Section 5.3 i.e. maximising the robustness by setting $\mathbf{w} = \nabla(R(\mathbf{r}))$.

6.4. Florencio-Malvar strategy

The Florencio-Malvar strategy was tested for JANIS and DSSS, i.e. ISS. Firstly, we choose λ maximising D_{KL} . For instance, in the ISS method:

$$\lambda = \lambda^* = \arg \max_{(0,1)} \frac{1}{2} \left(\log(\rho^2((1-\lambda)^2 + \text{NCR})) - 1 + \frac{1}{\rho^2((1-\lambda)^2 + \text{NCR})} + \frac{N \text{WCR} - \lambda^2}{(1-\lambda)^2 + \text{NCR}} \right)$$

The same strategy has been followed for JANIS and results are reported in Figure 1b. This strategy when applied to DSSS has a really good impact: the resulting D_{KL} is the highest we have achieved so far, for all NCR. Since it is well known that JANIS is more efficient than DSSS when the noise attack is light,⁴ it would have been expected that applying the Florencio-Malvar strategy to JANIS would yield even better results. In fact, this is not the case and it is really difficult to determine why this strategy is more beneficial to DSSS than JANIS.

The fact that $\lambda \neq 0$ implies $\sigma_{d|H_0} \neq \sigma_{d|H_1}$. Hence, assuming d is still gaussian distributed, it is not the best test statistic in the Neyman-Pearson sense.

Another important point to note is that D_{KL} is a theoretical criteria. It is masking some important realities of the detection problem, especially the existence of a fixed threshold T , related to P_{fa} . For this reason, we apply the Florencio-Malvar strategy with P_p as the criterion to be optimised instead of D_{KL} . For instance, DSSS gives:

$$P_p = 1 - Q(\eta(\lambda)) = 1 - Q\left(\frac{T - \rho\sigma_o\sqrt{N\text{WCR}} - \lambda^2}{\rho\sigma_o\sqrt{(1-\lambda)^2 + \text{NCR}}}\right)$$

As we work in very difficult conditions, i.e. low $N\text{WCR}$ for low P_{fa} , when $\lambda = 0$, $\eta(0) < 0$ because $T < \sqrt{N\text{WCR}}$. Hence, when increasing λ , η gets greater in amplitude but its sign is still negative, so that P_p is a decreasing function of λ as plotted in Figure 4b. In these conditions, the Florencio-Malvar strategy is useless for DSSS and JANIS with order $n = 2$. Yet, this is not the case for JANIS with order $n = 4$ or $n = 6$. Hence, there exists an optimal $\lambda > 0$ maximising P_p as shown in Figure 4b.

Figure 4a shows the power of the different tests and their improvement (for JANIS $n = 4$ and higher) with this strategy. Prior art methods were either robust but inefficient (e.g. DSSS, JANIS), or efficient but not robust (e.g. ZATT). Thanks to this strategy, JANIS schemes are now efficient and still robust. The optimal parameter λ is set to one when $\text{NCR} < -12$ dB (this reduces then to an ‘erase and write’ strategy) and set to zero (‘Maximising detection function for fixed distortion’ strategy) when $\text{NCR} > -10$ dB.

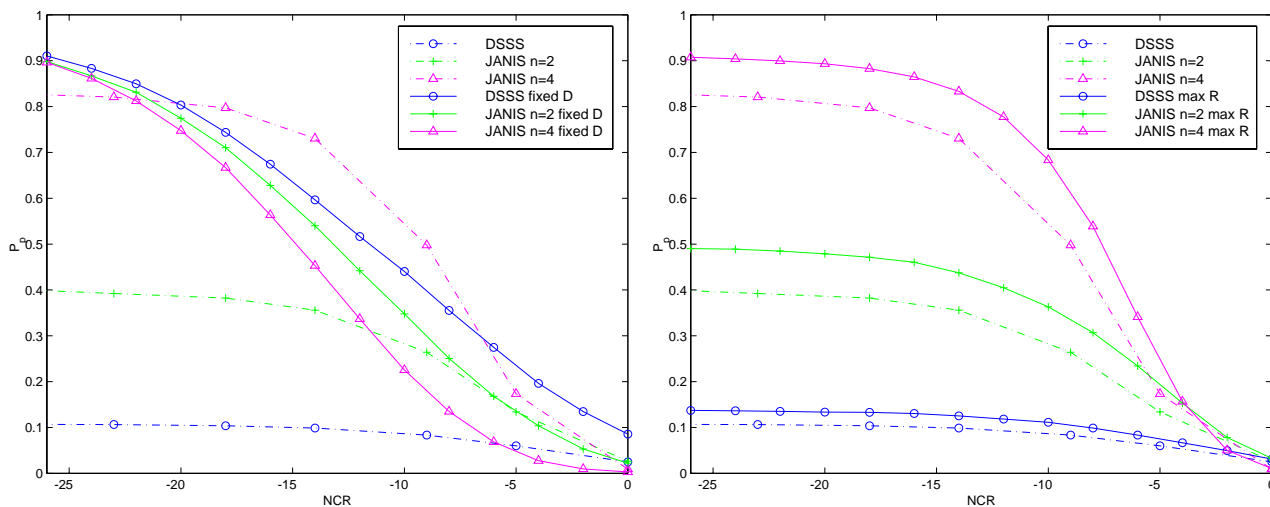


Figure 3. (a) Minimising distortion for fixed detection: P_p against WCR in dB, with $N = 2400$, $P_{fa} = 10^{-4}$. Achieved distortion are -22.3 dB (DSSS), -25.2 dB (JANIS $n=2$), -27.7 dB (JANIS $n=4$). **(b) Maximising robustness for fixed distortion:** P_p against NCR in dB for a fixed distortion of -26 dB.

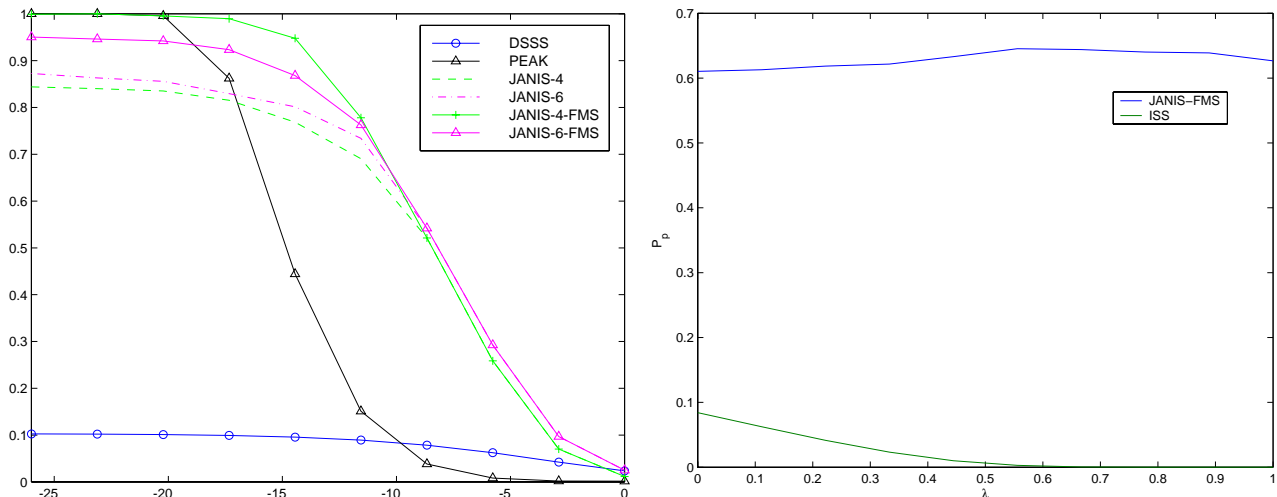


Figure 4. (a) P_p against NCR in dB, with N WCR = 6.0 ($N = 2400$, WCR = -26 dB). (b) The Florencio-Malvar embedding strategy does not improve P_p for DSSS. Here, NCR = -10 dB, there exists an optimal value for JANIS ($n=4$).

7. CONCLUSION

As far as robustness is concerned, the Florencio-Malvar strategy for maximising the power of the test yields one of the best results. It is particularly interesting as the distortion is more or less controlled. But a caveat is the fact that the embedder must expect the power of the noise attack. On the other hand, the ‘maximising robustness for a fixed distortion’ strategy yields as good results as the latter one, but no expectation is done about the strength of the attack.

REFERENCES

1. I. Cox, M. Miller, and A. McKellips, “Watermarking as communication with side information,” *Proc. of the IEEE* **87(7)**, pp. 1127–1141, July 1999.
2. I. Cox, J. Kilian, T. Leighton, and T. Shamoan, “Secure spread spectrum watermarking for multimedia,” *IEEE Transactions on Image Processing* **6**, pp. 1673–1687, December 1997.
3. T. Furon, *Use of watermarking techniques for copy protection*. PhD thesis, Ecole Nationale Supérieure des Télécommunications., 2002.
4. T. Furon, G. Silvestre, and N. Hurley, “JANIS: Just Another N-order side-Informed Scheme,” in *Proc. of Int. Conf. on Image Processing ICIP’02*, (Rochester, NY, USA), September 2002.
5. T. Furon, G. Silvestre, and N. Hurley, “Watermark detectors based on Nth order statistics,” in *Proc. of SPIE*, **4790**, SPIE, (Seattle, WA, USA), July 2002.
6. I. Cox, M. Miller, and J. Bloom, *Principles and Practice*, Morgan Kaufmann Publisher, 2001.
7. M. Costa, “Writing on dirty paper,” *IEEE Trans. on Information Theory* **29**, May 1983.
8. D. Florencio and H. Malvar, “An improved spread-spectrum technique for robust watermarking,” in *Proc of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 3301–04, (Orlando, FL, USA), May 2002.
9. R. Blahut, *Principles and practice of information theory*, Addison-Wesley, 1987.
10. J. Eggers and B. Girod, *Informed Watermarking*, Kluwer Academic Publishers, 2002.
11. M. Miller, G. Doerr, and I. Cox, “Dirty-paper trellis codes for watermarking,” in *Proc. of the IEEE Int. Conf. on Image Processing*, (Rochester, NY, USA), September 2002.
12. J. Chou and K. Ramchandran, “Robust turbo-based data hiding for image and video sources,” in *Proc. of the IEEE Int. Conf. on Image Processing*, (Rochester, NY, USA), September 2002.
13. M. Miller, I. Cox, and J. Bloom, “Informed embedding: exploiting image and detector information during watermark insertion,” in *Proc. of the IEEE Int. Conf. on Image Processing*, IEEE, (Vancouver, Canada), September 2000.