An Asymmetric Watermarking Method

Teddy Furon, Pierre Duhamel

An Asymmetric Watermarking Method

T. Furon* and P. Duhamel

Abstract

This article presents an asymmetric watermarking method as an alternative to classical Direct Sequence Spread Spectrum and Watermarking Costa Schemes techniques. This new method provides a higher security level against malicious attacks threatening watermarking techniques used for a copy protection purpose. This application, which is quite different from the classical copyright enforcement issue, is extremely challenging as no public algorithm is so far known to be secure enough and some proposed proprietary techniques have been already hacked. Our method is thus a try towards the proof that the Kerckhoffs principle can be stated in the copy protection framework.

Index Terms

Copy protection system, watermarking, security, asymmetric methods.

I. Introduction

DIGITAL watermarking is the art of hiding information in digital contents such as images, audio clips and videos. This art is useful in many various applications, but this paper only deals with the copy protection scenario. Although the role of the watermark is very simple in this application, some particular constraints render the subject extremely challenging. The main concern is the assessment of a security level according to the Kerckhoffs principle.

A. A challenge for the watermarking community

To build a copy protection system for consumer electronic devices, we are looking for a technique, which could hide in an original content a signal commonly called watermark. Compliant devices such as players or recorders are able to detect the presence of this watermark. In this particular case, its presence means that the content is protected and, hence, it is illegal to copy it. Watermarking is used as a mean to distinguish personal or copy-free data from protected copyrighted contents. Real world copy protection systems are indeed more complex, but this simple description brings enough severe issues. The technical requirements are well known for this kind of application in the watermarking community:

- No perceptual distortion. Watermarking should not spoil the entertainment of the contents.
- One bit payload. Detectors only check for a presence of a watermark.
- Robustness to common content processing like coarser source coding, D/A + A/D transformation, low pass filtering...
- Low complexity of the detection algorithm.
- Relative security.

The fifth criterion must be detailed. What is the difference between security and robustness? In one hand, robustness measures the impact on the detection capability of common transformations applied non-intentionally or intentionally to the protected contents. If intentional, they are qualified as *blind attacks* because it is a hopeless attempt from the pirates to remove the watermark. They know nothing about watermarking and they hope such transformations could remove its protection. On the other hand, security measures the impact on the detection capability of intentional processing dedicated to a certain class of watermarking techniques. They are sometimes called *malicious attacks* in the sense that the pirates know perfectly the watermark embedding and detection algorithms and they seek for flaws in this targeted technique.

This criterion is somewhat subjective. For instance, the motto of the CPTWG[1], one of the first industrial forums dealing with copy protection system, is 'Keep Honest People Honest'. It means that the watermarking technique doesn't provide a high level of security. Its hack must not be obvious so that it stays out of the reach of 'honest people'. Nevertheless, another industrial forum, the SDMI[2], was more concerned about security as this requirement clearly appeared in the call for proposal. The rationale under this concern might be the following scenario: imagine a well experimented pirate has disclosed a reproducible hack ; he has implemented it in a user friendly software, and he distributes it on the Internet, so that, now, removing watermark is on the reach of a mouse click for the not so 'honest people'.

SDMI has launched a challenge to evaluate four proposed audio watermarking techniques. The watermarking community performed very well as, at least, two research teams broke all proposed techniques [1], [2]. It shows that some companies have

* Corresponding author
T. Furon is now with the TEMICS project of IRISA/INRIA, Rennes, France. E-mail: tfuron@irisa.fr
P. Duhamel works in the LSS lab of SUPELEC, Gif-sur-Yvette, France. E-mail: pierre.duhamel@lss.supelec.fr
This work was supported by the Security lab of THOMSON multimedia R&D France.
[1]Copy Protection Technical Working Group
[2]Secure Digital Music Initiative

proposed solutions with no assessment of their security level. These techniques were robust but not secure. They though that their algorithms would be kept secret so that the pirates would have no clue to hack them. This rationale is called 'Secrecy by Obscurity'. The SDMI hacks revealed that this motto may lead to very hazardous solutions. But, pulling down techniques and pointing out possible flaws is not the only role of the academic community. It has to prove whether better solutions are achievable or not.

B. Kerckhoffs principle

Cryptography is an old science compared to watermarking. An impressive amount of works has been done in this field to clarify the basic concepts, primitives and proofs of a security level. The same tasks have to be done in watermarking.

The fundamental basement of cryptography has been established in 1874 by A. Kerckhoffs [3]. He stated that the designer of a crypto-system must suppose that the opponent knows his algorithms in details except a parameter called the secret key. Hence, the security of the crypto-system must only stem from storing the secret key in a safe place, the rest of the system being public. The Kerckhoffs principle is a heuristic defended by two facts:

- There are proprietary algorithms (i.e. violating the Kerckhoffs principle) that have been hacked. The book [4] gives numerous examples. The most famous is the hack of the Enigma encryption machine during the Second World War.
- There are public encryption algorithms still unbroken (e.g. RSA, DES).

Our work is only motivated by this simple question: Is the Kerckhoffs principle a valid concept in watermarking for copy protection? The episode of the SDMI challenge illustrates the first item towards a positive answer. But, watermarking is too young a science to get any assessment of the second item. So far, there is no public and secure watermarking algorithm for the copy protection application.

The issue about the validity of this principle is extremely important. Let us analyse the two alternatives:

- The Kerckhoffs principle is not valid. Every watermarking technique will be developed in secrecy. The academic research on this topic and its publications, conferences, is no more justified. Patenting new techniques is also no more possible. As no public study about security assessment is available, nobody can compare or trust the proposed techniques. In cryptography, one always claims that no algorithm can be kept secret during more than two years in the consumer electronic industry. These watermarking techniques are then likely to be hacked periodically.
- The Kerckhoffs principle is valid. This gives credits to the academic research. This community can analyse techniques to assess their security level. These results are public. It shares more and more experiences so that the security level increases. Industries will then launch a fair and reliable business once the techniques are mature.

The watermarking community is about to approve this fundamental principle for some applications. For instance, steganography is an application where, thanks to works from Cachin, one knows how to measure a security level [5], the stego-system being public. In the same way, the community has made a lot of improvements in copyright protection: some malicious attacks (e.g. the deadlock problem [6], the 'copy attack' [7]) have been spotted and counter attacks have been proposed. It seems that, in the copy protection application, this 'attacks / counter attacks' virtuous mechanism is not engaged due to the 'Secrecy by Obscurity' argument.

This article aims to propose an alternative solution to the classical watermarking schemes. This stronger solution would argue for the use of the Kerckhoffs principle for copy protection applications. The document is structured as follows. A brief overview of the watermarking processes introduces the basic notation. It is followed by a threat analysis in the copy protection framework in section II. Asymmetric schemes are presented as a counter-attack to the spotted threats. We describe in section III one of these methods and especially its unusual detection algorithm whose design requires some skills in testing hypothesis in spectral analysis. The security level provided by this asymmetric method is assessed in section IV. The simulations of section V confirm the previous analysis.

II. Conventional watermarking

This section details two different watermarking methods. Assuming that they fulfil the robustness and perceptibility constraints, attention is focused on the security criterion in the context of copy protection.

A. Notation

We set in this paragraph the usual structure of the watermarking scheme. It is described with the notation of [8], [9].

The goal of the embedding algorithm is to select the most important perceptual features of the covert content and to add a very small amount of watermark energy to each of them. From a cover content C_o belonging to the 'media space', an extraction function $X(.)$ maps the cover data into a feature vector of the 'watermark space': $\mathbf{r_o} = X(C_o)$. Denote N the length of this extracted vector. The role of the embedding process is to modify $\mathbf{r_o}$ into a vector $\mathbf{r_w}$ belonging to the critical region $\mathcal{R}$ of the 'watermark space'. This region is composed with extracted vectors considered as watermarked by the detector. The modification is performed by the 'mixing function' $F(.)$, which mixes the desired watermark signal with the extracted vector: $\mathbf{r_w} = F(\mathbf{r_o}, \mathbf{w})$. The 'inverse extraction' function $Y(.)$ finishes the embedding process. It maps back from the 'watermark

space' to the 'media space': $C_w = Y(\mathbf{r_w}, C_o)$. This modification is made under the constraint that the resulting watermarked content C_w is perceptually close to C_o.

The detection process gets an unknown received content C_u. It extracts the vector $\mathbf{r_u} = X(C_u)$ and checks whether it belongs to critical region $\mathcal{R}$. Let denote H_0 the hypothesis when the received content is not watermarked and H_1 the alternative hypothesis. The detection is a decision rule whose output $\check{D}(\mathbf{r_u})$ equals 1 if C_u is considered as watermarked and 0 else. Usually, this hard decision is the comparison of the likelihood function $D(\mathbf{r_u})$ with a positive threshold.

$$\check{D}(\mathbf{r_u}) = \begin{cases} 1 & \text{if} \quad D(\mathbf{r_u}) > T \\ 0 & \text{else} \end{cases} \tag{1}$$

The performance of the detection is measured by the probability of false alarm P_{fa} which is the probability that the detector claims a received content is protected whereas it was not watermarked, and the power of the test P_{de} which is the probability that the detector correctly detects a watermarked content. They are mathematically defined by $P_{fa} = P(\check{D}(\mathbf{r_u}) = 1|\mathrm{H}_0) = E\{\check{D}(\mathbf{r_u})|\mathrm{H}_0\}$ and $P_{de} = P(\check{D}(\mathbf{r_u}) = 1|\mathrm{H}_1) = E\{\check{D}(\mathbf{r_u})|\mathrm{H}_1\}$, where $E\{x\}$ is the mathematical expectation of random variable x.

The benefit of this model is that we tackle the contents as vectors of length N. Moreover, we assume that the vectors represent central stationary random processes whose autocorrelation functions are absolutely summable. In the simplest assumptions, vector $\mathbf{r_o}$ is modelled as central Gaussian white noise of variance $\sigma^2_{\mathbf{r_o}}$.

B. Direct Sequence Spread Spectrum

Direct Sequence Spread Spectrum (DSSS) is a modulation by a pseudo-random carrier, invented during the World War II [10]. The term "spread" comes from the fact that the information to transmit (here in this article, the presence of a watermark) has been embedded into a huge number of features. Hence, if the pirate succeeds to remove the watermark for some selected coefficients, there is enough energy left in the other features to detect the watermark. The carrier is generated by a pseudo random generator seeded with a secret key, as usually done in military digital communications. It provides the following well-known qualities [11]:

- The presence of the transmitted signal is hidden (low interception). This helps to respect the perceptual constraint.
- The ability to fight against interferences due to intentional scrambling, to other communications or to selective channels (e.g. multi-path channels). This brings robustness and it helps to fight against collusion attacks.
- Security is ensured as soon as the pseudo-random carrier is kept secret. Receivers ignoring this carrier can neither decode the message, nor change it. Jamming the communications then needs far more power.

These properties explain the success of the DSSS modulation in the watermarking community. This article clearly focuses on this technique as we hide one bit of information (presence or absence of watermark), through the embedding a long pseudo random sequence in the content.

Mathematically, the 'mixing function' is the addition of the original extracted vector with a normalised pseudo random sequence $\mathbf{w}$, as defined in Eq. (2), where gain g fixes the watermark strength, linked to a relative watermark to original content power ratio: $G = \frac{g^2}{\sigma^2_{\mathbf{r_o}}}$.

$$\mathbf{r_w} = \mathbf{r_o} + g\mathbf{w}. \tag{2}$$

At the detection side, the likelihood function is usually defined by the linear correlation of Eq. (3), which is the optimal tested statistic following the Neyman-Pearson strategy, if $\mathbf{r_o}$ is a white Gaussian vector..

$$D(\mathbf{r_u}) = {\mathbf{r_u}}^T\mathbf{w} \tag{3}$$

The tested statistic $D(\mathbf{r_u})$ being then Gaussian distributed, its pdf is fully described by its mean and the variance $(\mu_{\mathrm{H}_i}, \sigma^2_{\mathrm{H}_i})$ under hypothesis H_i, whose value are known to be

$$\mu_{\mathrm{H}_0} = 0 \qquad \mu_{\mathrm{H}_1} = gN \tag{4}$$

$$\sigma^2_{\mathrm{H}_0} = \quad \sigma^2_{\mathrm{H}_1} \quad = N\sigma^2_{\mathbf{r_o}}. \tag{5}$$

Noting $Q(.)$ the cumulative distribution function of $\mathcal{N}(0,1)$, the following relations between P_{fa}, P_{de} and T easily come:

$$T = T(P_{fa}) = \mu_{\mathrm{H}_0} + \sigma_{\mathrm{H}_0}Q^{-1}(1-P_{fa}) \tag{6}$$

$$P_{de} = P_{de}(P_{fa}) = 1 - Q\left(\frac{\sigma_{\mathrm{H}_0}}{\sigma_{\mathrm{H}_1}}Q^{-1}(1-P_{fa}) - \frac{\mu_{\mathrm{H}_1} - \mu_{\mathrm{H}_0}}{\sigma_{\mathrm{H}_1}}\right) \tag{7}$$

P_{de} is an increasing function of the deflection coefficient $\epsilon = (\mu_{\mathrm{H}_1} - \mu_{\mathrm{H}_0})/\sigma_{\mathrm{H}_1}$, equaling $\epsilon_{DSSS} = \sqrt{GN}$ in the DSSS case:

$$P_{de,DSSS} = 1 - Q\left(Q^{-1}(1-P_{fa}) - \sqrt{GN}\right) \tag{8}$$

C. Threat Analysis of DSSS

The watermark signal $\mathbf{w}$ clearly appears in the detection process. If one knows this signal, one can easily change $\mathbf{r_w}$ to forge a pirated content C_p. For instance, a pirate is sure that the following action will fool the detection process as it sets the correlation of Eq. (3) to zero, deluding the detector:

$$\mathbf{r_p} = X(C_p) = \mathbf{r_w} - (\mathbf{r_w}^T.\mathbf{w})\mathbf{w} \tag{9}$$

This is the reason why $\mathbf{w}$ must remain secret.

The impact of the discovery of $\mathbf{w}$ depends on the application. For copyright protection systems, it is likely that this secret vector, presumably issued by a trusted registration party is different for each content. Hence, the knowledge of one signal $\mathbf{w}$ only helps the pirates to forge one illegal content. But, for the copy protection framework, this knowledge may pull down the whole system: As explained below, all protected contents are watermarked with a unique vector $\mathbf{w}$.

The main issue of our security analysis is to know whether $\mathbf{w}$ can remain secret in the copy protection framework. Here is a list of possible threats issued from [12], [13]:

To average a huge amount of protected content: Due to a low cost constraint, the memory and computing power available at the detection side for extracting the vectors and calculating the correlation is very small. Hence, the secret vector is not very long. Due to the perceptual constraint, the watermark to content power ratio is very low. In many real-world techniques, the watermark signal is then tiled (i.e. repeated) all along the content. At the detection side, an average process increases the power ratio and consequently the efficiency of the test. This results in the fact that all protected contents are watermarked with the same secret, and moreover, each of them contains it several times. As the 'mixing function' is often linear or nearly linear, an average of $O(G^{-1})$ extracted vectors of independent protected contents gives an accurate estimation of this secret. To make a comparison with cryptography and its ciphertext only attack [14], this average attack belongs to a strategy called *watermarked content only attack*.

To defeat the embedding process: Designers of watermarking technique must be careful that their embedder is not leaking some information about the secret, while it watermarks special contents. For instance, watermarking uniform (or smooth gradient) areas of pictures is not a good idea from a perceptual point of view. But, it is also a real security threat: Pirates will easily isolate the components coming from the watermark signal. Another big issue is to watermark contents that have already been publicly disclosed. We can think of the following scenario: The content owners prefer to watermark their contents just before they are pressed on DVD or broadcast on the air. Trailers released several months ago when the movie was on screen, could be a source of the cover contents. Making a difference, in the 'watermark space', between watermarked vectors $\mathbf{r_w}$ and their corresponding original vector $\mathbf{r_o}$ gives pirates a lot of information about the secret signal. Virtually, one pair of content is enough to reveal the secret sequence. This is a *known cover content attack* (in comparison to the cryptographic known plaintext attack [14]).

To analyse the behaviour of the detector: I. Cox and J.P. Linnartz pointed out that the pirates are able to estimate the secret vector just observing how the chip behaves for a given number of faked chosen contents [12]. Although dishonest users have only access to the binary decision (i.e. the content is watermarked or not), the detection output leaks sufficiently information about the secret key to achieve its accurate estimation. This attack takes $O(N)$ tries as proved by T. Kalker [15]. We call it the *oracle attack*.

To steal the secret by reverse engineering: The secret is located in all embedders and detectors. The embedders are not publicly available. They are placed in the preparation stage of the contents, e.g. in the movies studio. We assumed that this place is safe from secret leakage. In the copy protection framework, this assumption does not hold for the detectors, since they will be implemented in hardware and placed in consumer electronic devices (e.g. DVD recorder). The reverse engineering of these devices is a real threat because tamper-proof hardware might be too expensive for the consumer electronic industry.

The reader not versed in security might be doubtful about these threats. There are numerous news about reverse engineering of security functions: game station (copy protection 'mod' chip), DVD players (dezoning, encryption hack [16]), mobile phone (authentication breaks), set top boxes (conditional access hacks). Indeed, no security can be provided if a safe place where keys are stored does not exist. In this article, we do not address the issue of how to avoid reverse engineering in hardware implementations, and we assume it is possible to do so. The other threats are somewhat more theoretical because no watermarking technique has been really deployed up to now. But, the SDMI challenge and its hacks provide us a good idea of the power of these 'signal processing' hacks [1], [2].

We discard in this article attacks like desynchronisation via geometric transformations (still images) or sample frequency wobbling of audio contents because they belongs more to the robustness issue which should be tackled by the design of the extraction function. Watermarking technique will be more and more robust against these blind attacks. Forged contents' quality will be lower and lower, so that, the pirates will give up this strategy, preferring the use of more powerful malicious attacks. Whereas robustness was so far the weak point focusing all the research efforts, the improvements achieved recently let us foresee that security will be the main issue in the future.

D. Watermarking Costa's Schemes

Watermarking Costa's Schemes (WCS) constitutes another class of extremely popular watermarking processes based on Costa's article [17]. A partition divides a constellation $\mathcal{U}$ of $L_{\mathcal{U}}$ independent reference vectors called codewords into $L_{\mathcal{M}}$ disjoint codebooks: $\mathcal{U} = \bigcup_{m\in\mathcal{M}} \mathcal{U}_m$. Each codebook is associated with a symbol to be transmitted. This allows transmitting one out of $L_{\mathcal{M}}$ symbols per cover content. The mixing function acts like an attraction where the extracted vector is pushed towards the nearest codeword belonging to the codebook $\mathcal{U}_m$ associated with the symbol m to be transmitted. This codeword is denoted $\mathbf{q}(\mathbf{r_o}, m)$ as it is the quantization of $\mathbf{r_o}$ on the codebook $\mathcal{U}_m$.

$$\mathbf{r_w} = \mathbf{r_o} + \alpha(\mathbf{q}(\mathbf{r_o}, m) - \mathbf{r_o}) \tag{10}$$

The detection finds the closest codeword to the received vector. The decoded symbol $\hat{m}$ is the index of the codebook, which this codeword belongs to. The reader can find more explanations on possible constructions of the codebooks, partitions and performances in [18], [19], [20].

Eq. (10) clearly shows that the watermark signal is created with the knowledge of the original extracted vector. This is called the side-information at the embedding stage and it yields schemes achieving (capacity / robustness against additive noise) characteristics far better than the ones resulting from DSSS. Side-information is absolutely recommended for applications requiring a large capacity where the additive noise attack model is suitable like steganography, copyright protection, fingerprinting or content integrity verification.

Yet, no work has been done about the application of such methods to copy protection. A priori, Costa's idea is not suitable as it efficiently decodes hidden messages in watermarked contents (that are structured in distinct ways at the embedding stage), whereas it is not designed to detect watermarked from original contents that are not at all structured. Only two research works tackle this issue.

The first idea is based on hard decision decoding [21]. A decoder fed with an original content retrieves a random message depending on the nearest codeword to $\mathbf{r_o}$. These random messages are assumed to be uniformly distributed: $P(\hat{m} = m|\mathrm{H}_0) = L_{\mathcal{M}}^{-1} \quad \forall m \in \mathcal{M}$. Then, one protects the contents hiding the message m_0. Hence, the test hypotheses becomes $\mathrm{H}_0 : \hat{m} \neq m_0$ vs. $\mathrm{H}_1 : \hat{m} = m_0$. Under hypothesis H_0, the relation $P_{fa} = P(\hat{m} = m_0|\mathrm{H}_0) = L_{\mathcal{M}}^{-1}$ sets the size of the alphabet. Under hypothesis H_1, the use of channel codes studied in [20] improves power of the test P_{de} at a given watermark to noise power ratio.

The second idea is based on soft information, where the goal is to produce a measure of reliability of the decoding process. Knowing the embedded symbol m_0, the pdf of the original and the watermarked vectors are significantly different so that a test based on likelihood is relevant: $D(\mathbf{r_u}) = p(\mathbf{r_u}|\mathrm{H}_1, m_0)/p(\mathbf{r_u}|\mathrm{H}_0)$. Performances are given in [20].

But, as these methods rely on quantization on codebooks, their Achilles' knee is the amplitude scaling and offset attack. The critical region is a set of small cells centred on the codewords of $\mathcal{U}_{m_0}$. The watermarked vector can easily go out of the sphere by a change of amplitude. This imposes first to recover the equivalently scaled and offset constellation observing possibly noisy version of watermarked vectors before being able to measure the likelihood. An estimation algorithm of scaled and offset codebooks for Scalar Costa Scheme (SCS) is proposed in [20]. Note that DSSS is inherently more robust to these attacks as its critical region is compact and in one piece.

E. Threat Analysis of WCS

Although no work is reported on the application of WCS to the copy protection framework, it seems possible to do so regarding its detectability and robustness. Its security in this framework is now analysed.

Let analyse the watermarked content only attack. Dithered Modulation (DM), Quantized Index Modulation (QIM) and SCS are competitive practical implementations of WCS where the constellation is structured as the product, component by component, of a uniform scalar quantizer: $\mathcal{U} = \Delta\mathbb{Z}^N$. It is believed that a secret dithering vector $\mathbf{k}$ would 'secure' DM, QIM and SCS such that $\mathcal{U} = \Delta\mathbb{Z}^N + \mathbf{k}$ is now a secret constellation. In the copy protection framework, there is only one instance of parameters Δ and $\mathbf{k}$ embedded in all consumer electronics devices. The challenge for pirates is to estimate the codebook $\mathcal{U}_{m_0}$ (whence the critical region) observing watermarked vectors. This is exactly the goal of the estimation method we mentioned above! The dishonest users proceed it on almost noiseless vectors coming from good quality watermarked contents, whereas the detector runs the same algorithm on possibly heavily attacked contents. Referring to [20], in this condition, $O(250)$ watermarked contents are necessary to estimate Δ and $\mathbf{k}$ within 1% of relative error. Once the codebook is discovered, the pirates forge good quality pirated contents thanks to, for instance, an inverse SCS algorithm with reduced distortion also available in [20].

Things are slightly easier with a known cover content attack. Moreover, according to M. Mansour and A. Tewfik, an oracle attack is also possible against WCS [22]. Up to now, the actual practical implementations of WCS are not designed for the copy protection scenario. It is technically possible to adapt them to such a context where they will achieve better performances than DSSS. Yet, they do not provide better security levels than DSSS.

F. Conclusion

Section II sets the basic notation and it makes an account of two widely used watermarking methods. We explain the reasons why attacks based on reverse engineering and on desynchronisation are discarded in this article. The main analysis focuses on three attacks that are typical from the copy protection scenario. It turns out that the classical watermarking methods give low security levels in this context.

III. Asymmetric watermarking methods

F. Hartung and B. Girod's article is the very first paper to introduce the notion of public-key (and thus asymmetric) watermarking [23]. Although the proposed solution was not at all secure, it stressed the issue that the would-be secret key of DSSS can not remain a secret in some applications. Recently, new asymmetric watermarking schemes have been proposed as an interesting alternative especially for the copy protection framework [24], [25], [26], [27], [28], [29]. Although invented independently, the authors showed that these methods shared the same detection algorithm (i.e. the correlation is replaced by a quadratic form), and that they have similar performances [30]. The only exception is the work from and J. Picard and A. Robert based on neural networks [31]. Moreover, we would like to point out the very interesting work from L. Gomes and *al.* who improved our scheme so that decoding of hidden messages and not only detection of a watermark signal is achieved [32].

The concept of asymmetry must be understood as a method and not as a complete technique. This method is based on the same breakdown structure explained in section II-A. The only changes are the way watermark signals are created and the definition of the critical region, i.e. the decision rule. This allows us to derive an asymmetric version from any classical watermarking techniques. The article [33] is the implementation of an asymmetric method to the well-known Boney Tewfik audio watermarking technique [34]. The article [35] is the performances comparison of an extremely robust still images watermarking technique [36] and its asymmetric version.

A. The basic idea: randomized embedding

In the classical watermarking methods described in II, the creation of the watermark signal is deterministic. In DSSS, $\mathbf{w}$ plays the role of the secret key, hence it is constant in the copy protection framework. In WCS, $\mathbf{w}$ is a fixed function of $\mathbf{r_o}$ given by Eq. (10). The main idea of asymmetric methods is to transform the watermark embedding into a random processing. The watermark signal depends on a secret key and on a random variable. Hence, two watermarked versions of the same original content are different as the random variable changes each time the embedding is run.

This idea is not new in security. This kind of variable is called a *random* in cryptography. Goldwasser-Micali, El Gamal, Blum-Goldwasser encryptions are examples of such probabilistic encryption schemes [37]. The random is here to cast a given property of the signals emitted by the embedding stage without revealing their exact nature.

The challenge resides on how to enforce an idea coming from cryptography in the signal processing field. The difficulty is that the detector ignores the random used at the embedding stage. It only knows the secret key, whence the asymmetry between the embedding and the detection. The detector can not check whether a precise signal has been hidden in the content. It must detect presence of watermark signals without knowing their exact values. The technical solution found so far by watermarkers cited above is to use second order statistics. The detector verifies whether the received content has a particular statistical property (related to a secret key) which is due to the presence of a random watermark signal. This property can not be expected from usual original contents. The way we implemented this idea in our method is described in the following subsection.

B. Example of an asymmetric method

The creation of watermark signals is as sketched in Fig. 1. First, a pseudo-random generator fed by the random gives a white Gaussian noise $\mathbf{v}$ with variance unity. This signal goes through a filter h whose frequency response module matches with a given spectrum's template $|H(f)|^2$. The filter is normalised so that $\int_{-0.5}^{0.5} |H(f)|^2 \, df = 1$. The resulting coloured noise is interleaved by a pseudo-random permutation ϖ of the vectors indexes to give the final watermark signal.

$$\mathbf{w} = \varpi(h \otimes \mathbf{v}) \tag{11}$$

$\otimes$ is the convolution product. Finally, this watermark sequence $\mathbf{w}$ will be embedded into sequence $\mathbf{r_o}$ thanks to the 'mixing function' $F(.)$ of Eq. (2) and the 'inverse extraction function' $Y(.)$ concludes the watermarking process. Compared to DSSS, the embedding process remains unchanged except the creation of the sequence $\mathbf{w}$.

The detection process first extracts a received vector $\mathbf{r_u} = X(C_u)$ which goes through the de-interleaver $\widetilde{\mathbf{r_u}} = \varpi^{-1}(\mathbf{r_u})$. This vector is then a mix of the interleaved extracted vector coming from the cover content and, if the received content is watermarked, a coloured noise whose spectrum is shaped like $|H(f)|^2$. The permutation is assumed to have a perfect whitening action. The detection does not know the watermark signal but only the de-interleaver and the spectrum's template.

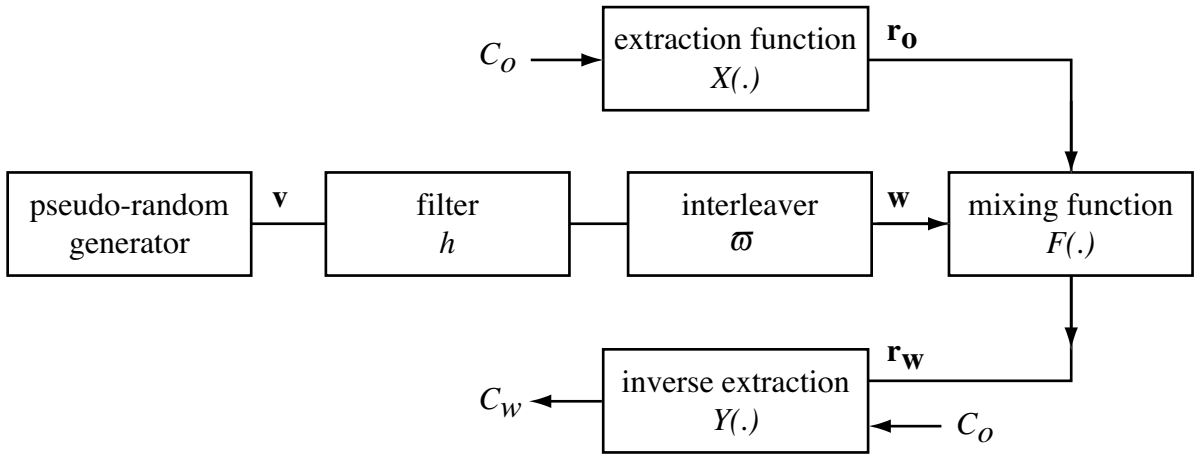

Fig. 1. The watermark embedding process.

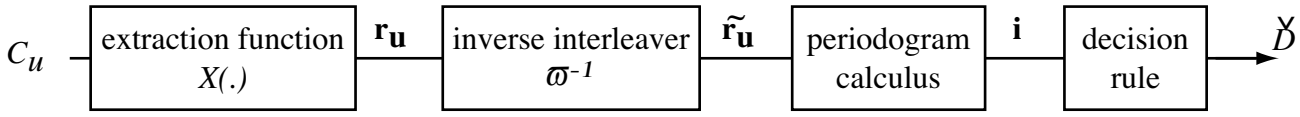

Fig. 2. The watermark detection process.

C. Testing hypothesis

The goal of this subsection is to define the decision rule for this asymmetric method. The structure of the detector is sketched in Fig.III-C.

1) Simple hypothesis test: It is extremely difficult to define a cost when the detector takes wrong decisions, so that Bayes or minimax strategies are not really suited for our purpose. Yet, the usual constraint on the detection (see SDMI or CPTWG calls for proposal) is to fix an upper level ω of the probability of false alarms P_{fa} This clearly references to a Neyman-Pearson test, which is the most powerful among all the tests with respect to the constraint $P_{fa} < \omega$ [38]. It is based on the comparison of the *log likelihood ratio* (or any monotone function of the likelihood ratio) to a threshold, so that the decision rule is defined by Eq. (12).

$$\check{D}(\mathbf{r_u}) = \begin{cases} 1 & \text{if} \quad \log \frac{p(\widetilde{\mathbf{r_u}}|\mathrm{H}_1)}{p(\widetilde{\mathbf{r_u}}|\mathrm{H}_0)} > T(\omega) \\ 0 & \text{else} \end{cases} \tag{12}$$

The threshold $T(\omega)$ is chosen such that $P_{fa} = E\{\check{D}(\mathbf{r_u})|\mathrm{H}_0\} < \omega$.

The main difficulty now is to measure the log-likelihood under each hypothesis. According to the above simple assumptions, $\widetilde{\mathbf{r_u}}$ is a central Gaussian vector of length N, so that:

$$\log p(\widetilde{\mathbf{r_u}}|\mathrm{H}_i) = -\frac{1}{2}\left(N \log 2\pi + \log\det(\mathbf{R}_i) + \widetilde{\mathbf{r_u}}'\mathbf{R}_i^{-1}\widetilde{\mathbf{r_u}}\right) \quad i \in \{0,1\} \tag{13}$$

where $\mathbf{R}_i$ is the covariance matrix of $\widetilde{\mathbf{r_u}}$ under the hypothesis H_i. For $i = 0$, $\widetilde{\mathbf{r_u}}$ is a white noise, hence, $\mathbf{R}_0 = \sigma^2_{\mathbf{r}_u}\mathbf{I}_N$. Yet, for $i = 1$, the calculus of $\det(\mathbf{R}_1)$ and of $\mathbf{R}_1^{-1}$ are really cumbersome, even for simple expressions of the filter h.

2) Principal part of the likelihood: Since the likelihood function is often used in parameters estimation or in hypothesis tests, many works have been done in order to render its use more practical. Whittle suggested to overcome the above-mentioned difficulty by replacing the log likelihood by its principal part $\widehat{L}_N(\mathbf{r}|\mathrm{H}_i)$, which satisfies Eq. (14).

$$\frac{\widehat{L}_N(\mathbf{r}|\mathrm{H}_i) - \log p(\mathbf{r}|\mathrm{H}_i)}{\sqrt{N}} \to 0 \quad \text{as } N \to \infty \quad \text{(conv. in prob.)} \tag{14}$$

The expression of $\widehat{L}_N(\mathbf{r}|\mathrm{H}_i)$ is simpler than the log likelihood's one, and at the same time, estimators and hypothesis tests based on principal parts ratio are proved to be asymptotically equivalent to those based ons log likelihood ratios.

Theorem 1 (Whittle): Let the expected power spectral density S_i under hypothesis H_i and the corresponding covariance function $\mathbf{R_i}$ of a stationary random process $\mathbf{r}$ satisfy the following conditions: $\exists \eta > 0 \mid S_i(f) \geq \eta \quad \forall f \in (-\frac{1}{2}, \frac{1}{2}]$ and $\sum_{k=1}^{\infty} k\,|R_i(k)|^2 < \infty$

Then relation (14) holds for

$$\widehat{L}_N(\mathbf{r}|\mathrm{H}_i) = -\frac{N}{2}\left(\log 2\pi + \int_{-\frac{1}{2}}^{\frac{1}{2}} \log S_i(f)df + \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{I_N(f)}{S_i(f)}df\right) \tag{15}$$

$$\text{with } I_N(f) = \frac{1}{N}\left|\sum_{k=0}^{N-1} r[k]e^{2\pi\jmath kf}\right|^2 \tag{16}$$

$I_N(f)$ is the periodogram of the vector $\mathbf{r}$.

For practical use, the integral forms in Eq. (15) are replaced by their Riemann sums sampled at the Fourier frequencies $\mathbf{f} = (1/N, \ldots, \bar{N}/N)^T$ with $\bar{N} = N/2 - 1$. N is a power of 2 to speed up the periodogram ordinates calculus. This stems in a new interpretation of the Whittle's approximation. Let us recall that the random variables $\{I_N(f[k])\}$ are distributed as a central χ_2 with 2 degrees of freedom with expectation $S_i(f[k])$ and variance $S_i^2(f[k])$:

$$p_{I_N(f[k])}(x) = \frac{1}{S_i(f[k])}e^{-\frac{x}{S_i(f[k])}} \quad \forall x \in [0, \infty) \tag{17}$$

Moreover, if the components of $\mathbf{r}$ are Gaussian distributed, then the periodograms sampled at Fourier frequencies are independent.

$$E\left\{\sum_{k=0}^{N-1} r[k]e^{2\pi\jmath f[a]k}.\sum_{l=0}^{N-1} r[l]e^{2\pi\jmath f[b]l}\right\} = 0 \text{ for } f[a] \neq f[b] \tag{18}$$

Eq.(18) implies the non-correlation of random variables $\{\sum_{l=0}^{N-1} r[l]e^{2\pi\jmath f[k]l}\}$, hence independence in the Gaussian case. Let us consider the vector $\mathbf{i} = (I_N(f[1]), \ldots, I_N(f[\bar{N}]))^T$. Its log likelihood under hypothesis H_i is the sum of the components' log likelihood thanks to their independence:

$$\log p(\mathbf{i}|\mathrm{H}_i) = -\sum_{k=1}^{\bar{N}} \log S_i(f[k]) + \frac{I_N(f[k])}{S_i(f[k])} \tag{19}$$

3) Final decision rule: Finally the decision rule is expressed in Eq. (20).

$$\check{D}(\mathbf{r_u}) = \begin{cases} 1 & \text{if} \quad D(\mathbf{i}) > T(\omega) \\ 0 & \text{else} \end{cases} \tag{20}$$

where $D(\mathbf{i})$ is the tested statistic, whose expression is

$$D(\mathbf{i}) = \sum_{k=1}^{\bar{N}} I_N(f[k])\left(\frac{S_1(f[k]) - S_0(f[k])}{S_1(f[k])S_0(f[k])}\right) + \sum_{k=1}^{n} \log\frac{S_0(f[k])}{S_1(f[k])} \tag{21}$$

The threshold $T(\omega)$ is estimated applying the Neyman-Pearson strategy with respect to the statistic $D(\mathbf{i})$. This statistic is distributed as a χ_2 with $N-2$ degrees of freedom. For large N, we assume it is a normal distribution invoking the central limit theorem so that $T(\omega)$ is calculated replacing P_{fa} by ω in Eq. (6). the mean and the variance under hypothesis H_i $i \in \{0, 1\}$, are as follows:

$$\mu_{\mathrm{H}_i} = \sum_{k=1}^{\bar{N}} S_i(f[k])\left(\frac{S_1(f[k]) - S_0(f[k])}{S_1(f[k])S_0(f[k])}\right) + \sum_{k=1}^{n} \log\frac{S_0(f[k])}{S_1(f[k])} \tag{22}$$

$$\sigma^2_{\mathrm{H}_i} = \sum_{k=1}^{\bar{N}} S_i^2(f[k])\left(\frac{S_1(f[k]) - S_0(f[k])}{S_1(f[k])S_0(f[k])}\right)^2 \tag{23}$$

We still have to estimate the power spectrum density under both hypotheses H_i.

- H_0: $\widetilde{\mathbf{r_u}} = \widetilde{\mathbf{r_o}}$ which is a white noise. Hence, $S_0(f[k]) = \sigma^2_{\mathbf{r_o}}$.
- H_1: $\widetilde{\mathbf{r_u}} = \widetilde{\mathbf{r_o}} + g(\mathbf{h} \otimes \mathbf{v})$ where $\mathbf{v}$ is a random process independent from $\mathbf{r_o}$. Hence, $S_1(f[k]) = \sigma^2_{\mathbf{r_o}} + g^2|H(f[k])|^2$. Moreover, the power of the marked vector is $\sigma^2_{r_u} = \sigma^2_{r_o} + g^2$.

Finally, the tested variable is:

$$D(\mathbf{i}) = \sum_{k=1}^{\bar{N}}\left(\frac{I_N(f[k])}{\sigma^2_{r_u}}\frac{g^2(|H(f[k])|^2 - 1)}{\sigma^2_{r_u} + g^2(|H(f[k])|^2 - 1)} + \log\frac{\sigma^2_{r_u}}{\sigma^2_{r_u} + g^2(|H(f[k])|^2 - 1)}\right) \tag{24}$$

4) Power of the test: Interesting performances are the receiver operating characteristic $P_{de} = P_{de}(P_{fa})$ plotted for a fixed G and the power function $P_{de} = P_{de}(G)$ plotted for a fixed level of false alarm ω. These curves help comparing different tests. When the sufficient statistic is Gaussian distributed, the power function is given by Eq. (7). To explicit how P_{de} depends on N and P_{fa}, we insert the expressions of the expected spectrum of III-C.3 in Eq. (22) and (23). Then, a Taylor development of the deflection coefficient ϵ and the ratio $\sigma_{\mathrm{H}_0}/\sigma_{\mathrm{H}_1}$ is made with respect to the variable $G \ll 1$. To estimate the behaviour of this expressions as N goes large, we assume that the Riemann sums converge to their corresponding integrals, i.e. $N^{-1}\sum_{k=1}^{N} \ldots = \int_0^1 \ldots df$. Here are the results:

$$\epsilon = G\sqrt{\frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} |H(f)|^4 df} + o(G\sqrt{N}) \tag{25}$$

$$\frac{\sigma_{\mathrm{H}_1}}{\sigma_{\mathrm{H}_0}} = 1 - G\frac{\int_{-\frac{1}{2}}^{\frac{1}{2}} |H(f)|^6 df}{\int_{-\frac{1}{2}}^{\frac{1}{2}} |H(f)|^4 df} + o(G) \tag{26}$$

The main conclusion is that the deflection coefficient ϵ is nearly proportional to $G\sqrt{N}$ which is lower than ϵ_{DSSS} in Eq. (8). Asymmetric detectors are thus less efficient than DSSS ones by a factor $\sqrt{G} < 1$. As G is a parameter fixed by the perceptual constraint, the only way to face this inconvenient is to increase the length of the extracted vectors. Usually, $G \sim -20$dB so that the vectors of an asymmetric scheme should be 10 times longer to reach the same efficiency than DSSS schemes.

Another strategy is to tile the same watermark signal in order to artificially increase the ratio G by an average process at the detection stage. In order to fairly analyse the gain of this tiling, we fix the total length of the embedded signals: $\mathbf{w}$ is now of length $N' = N/\tau$ but repeated τ times. The decision rule works on the vector $\mathbf{r_u}'$:

$$\mathbf{r_u}' = \frac{1}{\tau}\sum_{i=0}^{\tau-1} \mathbf{r_u}^i = \frac{1}{\tau}\sum_{i=0}^{\tau-1} \mathbf{r_o}^i + g\mathbf{w} = \mathbf{r_o}' + g\mathbf{w} \tag{27}$$

The power $\sigma^2_{r'_o}$ is τ times lower than $\sigma^2_{r_o}$, if we assume the averaged vectors are independent. Then, the new deflection coefficient of the asymmetric scheme is:

$$\epsilon' = G'\sqrt{N'} = G\sqrt{N\tau} = \sqrt{\tau}\epsilon \tag{28}$$

One notices that tiling does not increase the deflection coefficient of DSSS scheme: $\epsilon'_{DSSS} = \epsilon_{DSSS}$. In this case, the tiling only reduces the complexity of the correlation-based detection. Yet, the tiling has also some drawbacks. It compromises the assumption $N \to \infty$ used to build our decision rule. Moreover, it could introduce a security flaw in our scheme, as explained in section IV.

D. Asymmetric methods in the real world

We have made simple assumptions to properly study the structure of the test and its performances. These assumptions are not always realistic. This subsection focuses on changes implied by the practical use of asymmetric schemes.

1) Modulation of the perceptual constraint: The perceptual constraint usually leads to a modulation by a local gain control of the watermark strength. A more realistic embedding formula is then Eq. (29).

$$r_w[k] = r_o[k] + g[k].w[k] \text{ with } g = \frac{1}{N}\sum_{k=0}^{N-1} g[k] \tag{29}$$

This impacts on the expected spectrum under hypothesis H_1.

$$S_1(f) = \sigma^2_{\mathbf{r_o}} + S_{\widetilde{\mathbf{g}}}(f) \otimes |H(f)|^2 \quad \forall f \in (-\frac{1}{2}, \frac{1}{2}] \tag{30}$$

where $S_{\widetilde{\mathbf{g}}}(f)$ is the Fourier transform of the auto-correlation function of the vector $\widetilde{\mathbf{g}}$. Thanks to the whitening action of the pseudo-random permutation, this vector is assumed to be white, so that:

$$S_{\widetilde{\mathbf{g}}}(f) = \sigma^2_{\mathbf{g}} + g^2\delta(f) \quad \forall f \in (-\frac{1}{2}, \frac{1}{2}] \tag{31}$$

Finally, we have:

$$S_1(f) = \sigma^2_{\mathbf{r_o}} + \sigma^2_{\mathbf{g}} + g^2|H(f)|^2 \quad \forall f \in (-\frac{1}{2}, \frac{1}{2}] \tag{32}$$

The conclusion is that the pseudo random permutation boils out the impact of the perceptual modulation. We only have to know the average watermark strength g.

2) Composite alternative hypothesis: The knowledge of gain g is a very important issue that surprisingly received very little attention in the watermarking community. Indeed, up to know, the detection algorithm is always based on a simple hypothesis test as described in III-C.1. This assumption is not realistic: there exist contents which do not bear much watermark strength (e.g. the image 'peppers' with lot of uniform areas) and others which are less sensitive to watermark embedding (e.g. the image 'baboon' with lot of textured areas). Hence, it is not realistic to pretend that the average watermark strength is the same for all contents. Practical detection algorithms should thus be one sided tests: $\mathrm{H}_0 : \quad g = 0$ versus $\mathrm{H}_1 : \quad g > 0$.

This reality may not be a problem in the DSSS method. In Eq. (1), neither the sufficient statistic (a linear correlation), nor the threshold (fixed under H_0) depends on the parameter g. On the opposite, in the asymmetric method, the test is defined only for a given g. In other words, whereas in DSSS, a Uniform Most Powerful test may exist [38] ; in the asymmetric case, it is not possible to find one.

As an illustration of this issue, we distinguish, from now on, the real watermark strength g_e applied at the embedding stage from g_d expected at the detection side. Define $\gamma = (g_d/g_e)^2$. The functions $\mu_{\mathrm{H}_0}(\gamma)$, $\mu_{\mathrm{H}_1}(\gamma)$ and the calculated threshold $T(g_e, \omega)$ are drawn on Fig. 3. It illustrates what happens if we apply the simple hypothesis test strategy: a fixed threshold is calculated for a given g_d, e.g. $g_d = E\{g_e\}$. We note that when $g_e \neq g_d$, the test is likely to find a watermarked content except if $\gamma \gtrsim 2.4$, where the expectation of the tested statistic falls below the threshold. It means that contents that have received a small amount of watermark energy will never be detected. On the other hand, if the detector has access to the value of g_e, the statistic is compared to the adaptive threshold $T(g_e, \omega)$, and watermarked contents are more likely to be detected whatever the embedding strength.

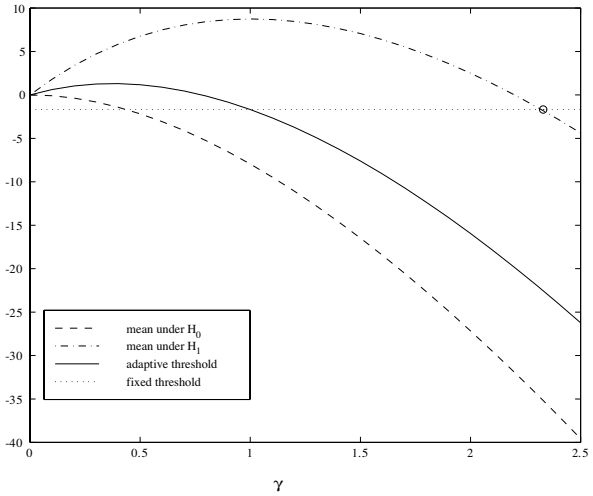

Fig. 3. Mismatch between the watermarking strength g_e used at the embedding side and its expectation g_d at the detection side. The mean of the tested statistic is plot for both hypotheses. The test is more powerful locally when the mean under H_1 is maximum, i.e. $g_d \sim g_e$. If $g_d > g_e$ the likelihood that the content is watermarked decreases.

Some solutions to this problem are proposed below:

• $\check{D}_1$: Estimate the pdf $p_{G_e}(.)$ of the parameter g_e from a huge amount of contents. The detection uses a Neyman-Pearson test with the parameter g_d that maximises the expectation of the power function:

$$g_d = \arg\max_{g>0} \overline{P_{de}}(g) \tag{33}$$

with

$$\overline{P_{de}}(g) = \int_0^\infty P_{de}(g_e, g) p_{G_e}(g_e) dg_e \tag{34}$$

This is not an easy task as the threshold also depends on the parameter g_d. This test is only optimal when $g_e = g_d$, i.e. there exist other tests more powerful locally. But, $\check{D}_1$ globally maximises the power of the test.

• $\check{D}_2$: Build a test not depending on this parameter and optimal when the two hypothesis are hardly distinguishable, i.e. $g \gtrsim 0$. This is a Locally Most Powerful test near the frontier of the two hypothesis. It is well known that this test has the structure of the Neyman-Pearson decision rule where the tested statistic is replaced by its derivative with respect to $\theta = g^2$ calculated for $\theta = 0$ [38]. In our case, the test becomes:

$$\check{D}_2(\mathbf{r_u}) = \begin{cases} 1 & \text{if} \quad D_2(\mathbf{i}) = \sum_{k=1}^{n} \left(\frac{I_N(f[k])}{\sigma^2_{r_u}} - 1 \right) (|H(f[k])|^2 - 1) > T_2(\omega) \\ 0 & \text{else} \end{cases} \tag{35}$$

• $\check{D}_3$: Take advantage of the side information available at the detection stage. Using in the detector the same perceptual model as the embedding stage provides an estimate of the watermark strength. We are then back to a simple alternative hypothesis, conditioned for each received content by the side information about the estimated embedding strength.

• $\check{D}_4$: Use a generalised Neyman-Pearson test [38]. Another way to determine the strength g_e is to use a maximum likelihood estimator, i.e. $\widehat{g_e}$ is the parameter maximising the log likelihood:

$$\widehat{g_e} = \arg\max_{g>0} \log p(\mathbf{i}|\mathrm{H}_1, g) \tag{36}$$

Then, the classical Neyman-Pearson test is proceeded. This is equivalent to the following decision rule:

$$\check{D}_4(\mathbf{r_u}) = \begin{cases} 1 & \text{if} \quad D_4(\mathbf{i}) = \max_{g>0} D(\mathbf{i}, g) > T_4(\omega) \\ 0 & \text{else} \end{cases} \tag{37}$$

Whereas the first two tests share the same complexity with the previous one for simple hypotheses, the last two decision rules require more computing power. $\check{D}_3$ needs the implementation of the perceptual model used at the embedding stage. $\check{D}_4$ needs a gradient algorithm to find the maximum of the log likelihood ratio.

3) Asymptotic optimality: Having already assumed that N goes to infinity, the four previous tests are compared by their asymptotic performances. Define Γ_θ the Fisher's information with respect to parameter θ

$$\Gamma_\theta = \lim_{N\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left(\frac{\partial}{\partial\theta} S_1(f[k]) \right)^2, \tag{38}$$

and $\Delta_{N,\theta}$ the following random variable:

$$\Delta_{N,\theta} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{I_N(f[k]) - S_1(f[k])}{S_1(f[k])} \frac{\partial}{\partial\theta} S_1(f[k]). \tag{39}$$

$\Delta_{N,\theta}$ is proved to be asymptotically distributed as a Gaussian with variance Γ_θ. Dzhaparidze names the Rao test the following one in the chapter IV of [39]:

$$\check{D}_R(\mathbf{r_u}) = \begin{cases} 1 & \text{if} \quad D_R(\mathbf{i}) = \Gamma_0^{-1/2} \Delta_{N,0} > T_R(\omega) \\ 0 & \text{else} \end{cases} \tag{40}$$

Moreover, he proves this test possesses asymptotically the best average power among all decision rules with a level of false alarm ω. This shows $\check{D}_R$ is asymptotically equivalent to $\check{D}_1$. It is also asymptotically equivalent to the test $\check{D}_4$ as stipulated in [39] because:

$$\frac{1}{n} D_4(\mathbf{i}) - \Delta_{N,0} \Gamma_0^{-1} \Delta_{N,0} \overset{N\to\infty}{\to} 0 \text{ in probability.} \tag{41}$$

Finally, developing expression of Eq. (40), we notice that the decision rule $\check{D}_R$ and $\check{D}_2$ are indeed based on proportional statistics.

The asymptotic properties of the Rao test allow us to conclude that the tests $\check{D}_1$, $\check{D}_2$ and $\check{D}_4$ are equivalent as N goes large. The test $\check{D}_2$ will be preferred, as it requires less computing power. Moreover, it has asymptotically the best average power; hence, it outperforms test $\check{D}_3$.

4) Non Gaussian distributed vectors: In the real world, the extracted vectors are not Gaussian distributed. This pulls down our previous analysis for this simple reason: Eq. (18) no longer holds and the periodograms at Fourier frequencies are not independent anymore. According to the proposition 10.3.2 of [40], the periodograms at frequencies $0 < f_a < f_b < \bar{N}$ have the following properties.

$$\mathrm{var}(I_N(f_a)) = \sigma_{r_o}^4 + \frac{\kappa_4(r_o)}{N} \tag{42}$$

$$\mathrm{cov}(I_N(f_a), I_N(f_b)) = \frac{\kappa_4(r_o)}{N} \tag{43}$$

The fourth cumulant $\kappa_4(r_o)$ is null for the Gaussian distribution, but not necessarily so in the general case: the periodograms are still unbiased estimators of spectrum of non Gaussian time series, but their variance increases. Especially, the periodograms are not independent. Our rationale is still valid if we consider the vector $\mathbf{i} = (I_N(f_1), \ldots, I_N(f_{N_b}))^T$ of fixed length N_b and when $N \to \infty$. It means that periodograms for a fixed number of bins are asymptotically independent. But, the interpretation of the Whittle principal part in III-C.2 is no longer true if we consider the statistic defined in Eq. (21) $\mathbf{i} = (I_N(f[1]), \ldots, I_N(f[N/2-1]))^T$. Yet, we will still study it although we can not justify its use. Changes appear then in the expressions of σ_{H_0} and σ_{H_1}. For instance, let us compare the standard deviation $\sigma_{\mathrm{H}_0,(NG)}$ in the non Gaussian case with $\sigma_{\mathrm{H}_0,(G)}$ in the Gaussian case:

$$\frac{\sigma^2_{\mathrm{H}_0,(NG)}}{\sigma^2_{\mathrm{H}_0,(G)}} = 1 + 2K_4(r_o)G^2 \int_{-1/2}^{1/2} |H(f)|^4 df + o(G^2) \tag{44}$$

where $K_4(r_o)$ is the kurtosis of the random process r_o in the non Gaussian case. This increase is relatively small and does not spoil our previous tests.

A much harder issue is the asymptotic Gaussianity of the statistics used in the four proposed tests when the extracted vectors are not Gaussian. As this implies complex mathematical tools far beyond the scope of this paper, we discard any sketch of proof and invite the reader to see theorems developed in the following references. The major argument is that, under hypothesis H_0, these decision rules are based on linear functionals of periodograms of white noise. This is directly related to the works of G. Fay [41], who proved the asymptotic Gaussianity and consistence of linear and non-linear functionals of white or linear random processes. Yet, under hypothesis H_1, the watermarked signal is not a linear random process. Nevertheless, the filter h has a non-singular frequency response and the watermark strength is so small, that we can assume the random process is almost white. Moreover, the consistency of our method under real conditions was experimentally assessed in [33].

IV. Security assessment

A. Relationship with Cryptography

To introduce our analysis, we would like to establish some comparisons with basic concepts from cryptography.

1) One way function: At the embedding stage, the secret key is the set of the following parameters $\{\mathbf{v}, \mathbf{h}, \varpi\}$, whereas the detection key is composed of $\{|H(f[k])|^2, \varpi^{-1}\}$. Hence, there exits a one way function deriving the detection key from the embedding key. This implies that the detector works yet without knowing what signal has been added to the content. The use of such a one way function has been firstly mentioned by Diffie and Hellman as mandatory to achieve asymmetric crypto-systems [14]. Later on, this concept led to the famous RSA algorithm, where the non-inversion of the one way function is based on the hard mathematical problem of factoring large integers in prime numbers. Somehow, the only thing we did is to find a one way function from the signal processing field and to derive a simple watermarking scheme on this basis.

In the same way, the parameter $\mathbf{v}$ of Eq. (11) can be seen as a *trapdoor* [14]. It is possible to recover the original content C_o from its watermarked version if and only if this trapdoor is known. Yet, remember that the recovery of C_o is not the goal of the pirates. All they need is a content that looks like the distributed content and deludes the detectors. It is possible that such a pirated content could be forged only knowing the detection key. This rises far more concerns as this key is stored in millions of consumer electronic devices.

In other words, whereas asymmetric schemes mean public key crypto-systems in cryptography, this implication is not true in the watermarking field. The detection key is to remain secret. As far as we know, there is no, so far, public key watermarking primitives. This highly desired primitive, if existing, is an asymmetric scheme where it is proved that the knowledge of the detection key can not help the pirates to forge a convenient content. This is not the case in actual asymmetric techniques. We produced in [30] an attack available if the detection key is disclosed. This attack threatens all methods unified in this later paper. Thus, the alternative solution we proposed is defenceless against the *reverse engineering threat* listed in section II-C.

2) Role of the pseudo-random permutation: The role of the permutation is extremely important. Of course, if the extracted vectors are not white processes, its presence is mandatory to whiten them and to make the detection algorithm work. But, this permutation has also a role from a security point of view. It hides in what space the detection algorithm takes place. Without its knowledge, a dishonest user can not predict the impact of his attacks on the detection output. The permutation maintains the pirate in a blindness state, where all he can do is to modify the extracted vector and to hope that this will flip the detection output.

Let us denote $\mathbf{r_p}$ the extracted vector of the pirated content. We assume that the attack can be modelled by the addition of an independent signal $\mathbf{p}$: $\mathbf{r_p} = \mathbf{r_w} + \mathbf{p}$. The crucial question is the impact of the vector $\widetilde{\mathbf{p}}$ on the decision rule. It is likely that this permuted vector is also white, so that it just lowers the watermark to content power ratio to $G = g^2/(\sigma_{r_o}^2 + \sigma_p^2)$. Because the attack must conserve the quality of the content, it decreases slightly the power ratio. J. Stern and J.-P. Tillich investigated the probability that the attack escapes to this likely whitening action of the permutation [28], [42]. They showed that for a given attack vector $\mathbf{p}$ and a given watermarked content $\mathbf{r_w}$, the set of the permutations that will give a successful hack is extremely narrow, i.e. its size decreases exponentially as N goes large. It results in an exponentially small probability of a successful hack from a blind attack.

The use of pseudo-random permutation is a common thing in cryptography in order to hide the space where cipher texts decryption is feasible. The public key crypto-system of McEliece is such an example [37].

B. Malicious Attacks

We can not pretend that the dishonest user will just blindly attack the contents. As described in section II-C, he will certainly try to gain some information about the secret parameters from a huge amount of protected contents and from the detection black box implemented in devices. This section gives credits to the asymmetric methods proving their superiority against these malicious attacks.

1) Attacks on protected contents: Of course, the 'average attack' spotted in section II-C is not anymore valid because of the randomised embedding action of vector $\mathbf{v}$. The aim of the pirate is to gain information about the permutation ϖ. This is the most important secret parameter as it gives access to the 'detection space'.

The only way is to manage a brute force attack where the pirate tries all permutations. His issue is then to know whether he has found a suitable permutation. He can not perform the detection as $\{|H(f[k])|^2\}$, a secret parameter, is missing. A possibility is to build a detector that distinguishes a white noise from a coloured noise, whatever its spectrum's shape. K.Drouiche and G.Fay have recently improved prior art of such a detector proposing the following statistic [41]:

$$\aleph = \log \sigma_{r_w}^2 - \frac{1}{K_N} \sum_{k=1}^{K_N-1} \log(\overline{I_N}(f[k])) \tag{45}$$

where $\overline{I_N}(f[k])$ is the pooled periodogram over p consecutive bins. $p > 4$ and $K_N = N/2p - 1$ (we assumed $N \in 2p\mathbb{N}$). The deflection coefficient of this test, in our framework, is proportional to $G^2\sqrt{N}$. This poor efficiency disables a practical implementation of this *watermarked content only attack*. Nevertheless, this strategy is possible with a *known cover content attack* as the original signal can be removed.

The set of the permutations is, a priori, composed of $N!$ elements. For instance, if $N = 2048$, using Sterling's formula, $N! \sim 2^{19000}$. This outnumbers the particles in the universe ($\sim 2^{270}$). On the other hand, not all permutations are suitable. ϖ must have an extremely good scrambling property to whiten the signals. In the same way, the pirate does not have to find the exact permutation but one sufficiently close to ϖ. To better assess the security level against this attack, we suppose that the pirate knows what pseudo-random permutation generator we used. Its seed is a binary sequence of length $L \ll N$. It is assumed that all these permutations have a sufficient whitening action and that they are independent. Hence, the brute force attack is reduced to a figure of 2^L tries. Imagine that one try (creation of one permutation, interleaving the vector, periodogram calculus and finally the Drouiche test) lasts $1\mu s$. Then, the pirate is likely to find the permutation ϖ in $2^{(L-log_2(10^6*60*60*24*365)-1)} \sim 2^{L-46}$ years. Of course, this is a raw estimation of the security level. The pirates may use several computers in parallel and the Moore's law is not taken into account.

2) Attacks on the detector: The dishonest user is testing the detector feeding it with faked chosen contents. The first difficulty is that he does not have access to the tested statistic but to its binary comparison with the threshold. Hence, he does not always know if its attack has decreased the tested statistic. To overcome this difficulty, T. Kalker proposed to render the detector sensitive, i.e. to feed it with faked contents whose statistics are close to the threshold. This is the reason why this attack is also called the 'sensitivity attack'. Doing this, slight changes in the content are likely to flip the detector output, giving information about the detection key. This strategy is efficient with the DSSS method. An estimation of a secret key of length N is determined within $O(N)$ tries. The pirate removes this estimated signal to the extracted vectors of contents to be forged as in Eq.(9). This not exactly results in the original contents, but in good quality forged contents.

This *oracle attack* does not work so well with this asymmetric method. The first step is to produce vectors whose impact on the detection are known to be negative (i.e. they decrease the tested statistic). Let us create two extracted vectors families $\{\mathbf{e}_{i,j}\}$ and $\{\overline{\mathbf{e}}_{i,j}\}$ so that $e_{i,j}[k] = \delta[k-i] + \delta[k-j]$ and $\overline{e}_{i,j}[k] = \delta[k-i] - \delta[k-j]$. We require that $i \neq j$, thus, each family contains $N(N-1)$ elements. Both families are closed with respect to the permutation $\varpi(.)$:

$$\widetilde{\mathbf{e}_{i,j}} = \mathbf{e}_{\varpi(i),\varpi(j)} \tag{46}$$

$$\widetilde{\overline{\mathbf{e}}_{i,j}} = \overline{\mathbf{e}}_{\varpi(i),\varpi(j)} \tag{47}$$

Their corresponding periodograms are:

$$I_{\widetilde{\mathbf{e}_{i,j}}}(f[k]) = \frac{2}{N}\left(1 + \cos 2\pi(\varpi(i) - \varpi(j))f[k]\right) \tag{48}$$

$$I_{\widetilde{\overline{\mathbf{e}}_{i,j}}}(f[k]) = \frac{2}{N}\left(1 - \cos 2\pi(\varpi(i) - \varpi(j))f[k]\right) \tag{49}$$

The pirate feeds the detector with a faked content whose extracted vector is $\mathbf{r} = a\mathbf{e}_{i,j}$. We assume the detection rule is $\check{D}_2$. Then, the sufficient statistic is proportional to one of the coefficients $\{c_v\}$ of the Fourier series of the sequence $\{|H(f[k])|^2\}$:

$$D_2(\mathbf{i}) = \sum_{k=1}^{n} \left(\frac{N I_{\widetilde{\mathbf{e}_{i,j}}}(f[k])a^2}{a^2\sigma_{e_{i,j}}^2} - 1 \right) (|H(f[k])|^2 - 1) \tag{50}$$

$$= \sum_{k=1}^{n} |H(f[k])|^2 \cos 2\pi(\varpi(i) - \varpi(j))f[k] \tag{51}$$

$$= Nc_v \text{ with } v = \varpi(i) - \varpi(j) \tag{52}$$

If $a\mathbf{e}_{i,j}$ (or $a\overline{\mathbf{e}}_{i,j}$) sets the detector output to one, then it means that $c_v > T(\omega)/N$ (respectively $c_v < -T(\omega)/N$). If it's not doing so, then the pirate doesn't know the influence of the couple of indices (i,j) in the detection process. If there is such a

coefficient $c_v > T(\omega)/N$, then there are N couples (i, j) increasing the detection output that the pirate can find. These are all the couples such that $v = \mathrm{mod}(\varpi(i) - \varpi(j), N)$.

Denote N_u the number of Fourier series coefficients such that $|c_v| > T(\omega)/N$. The pirate has NN_u couples of samples' indices to be modified in order to decrease the tested statistic D. The main problem is that the periodogram clearly is a non linear function of the samples $\{r_w[k]\}$. For instance, adding the vector $\mathbf{p} = a\overline{\mathbf{e}}_{i,j}$ to $\mathbf{r_w}$ modifies the tested statistic as follows:

$$D_2(\mathbf{r_p}) \sim D_2(\mathbf{r_w}) - \frac{2a^2}{\sigma^2_{r_w}} c_v + D'_2 \tag{53}$$

where D'_2 is the term for the interference between $\mathbf{r_w}$ and $a\overline{\mathbf{e}}_{i,j}$:

$$D'_2 = \frac{2a}{\sigma^2_{r_w}} \sum_{k=1}^{n} |H(f[k])|^2 \sqrt{I_N(f[k]) I_{\widetilde{\overline{\mathbf{e}}_{i,j}}}(f[k])} \cos(\phi_{\widetilde{r_w}}(f[k]) - \pi(\varpi(i) + \varpi(j)) f[k]) \tag{54}$$

$\phi_{\widetilde{r_w}}(f[k])$ being the angle of the Fourier transform of $\widetilde{\mathbf{r_w}}$ at the frequency $f[k]$. The pirate can not predict this interference term. It clearly depends on the watermarked content he aims to pirate. Its expectation is null, but its variance is proportional to $4a^2/\sigma^2_{r_w}$. This term sometimes helps him in his forgery, sometimes it has the opposite effect. Adding some other vectors $a\overline{\mathbf{e}}_{i,j}$ strengthens his attack, but also the interference term. Finally, the pirate can not predict how many tries are needed to forge a given protected content.

In conclusion, it is recommended to design the template $\{|H(f[k])|^2\}$ so that most of its Fourier series coefficients are lower than $T(\omega)/N$. At least, the pirate needs $O(N^2)$ tries to retrieve all the available information. This is not considered as a good security level in cryptography because the attack takes a polynomial time. We would have clearly preferred an exponential time as in the previous section. But, it is often required in copy protection that the detector takes a decision every 10 seconds. Let us suppose that $N = 2^{14}$ and that the pirate can not speed up the detection process. For DSSS, $O(N)$ tries take almost 2 days. For asymmetric schemes, $O(N^2)$ tries takes 85 years! Once again, this is a raw estimation of the security level. The pirates may use several detectors in parallel. Moreover, once this preliminary step is completed, the attack is a random processing which succeeds after an unknown number of tries, due to the interference term between $\mathbf{r_w}$ and $\mathbf{p}$.

V. Simulations

Our simulations are proceeded on still grayscale images of size 512×512 pixels and 8 bits by pixel. The media space is the spatial domain.

A. Details about the technique

In this subsection, we give the details about the implementation of the DSSS technique invented by A. De Rosa and *al.* [36]. Its robustness is impressive, especially with the optimal version of the detector [43].

$X(.)$ orders in vector $\mathbf{r_o}$ a subset of the magnitude of N discrete Fourier transform coefficients of C_o. These coefficients are extracted between the k-th and the $(k+n)$-th diagonal in the first quadrant and their symmetrical images in the second quadrant [36]. The mixing function modifies the amplitude of the DFT coefficients store in $\mathbf{r_o}$ proportionally to their value:

$$r_w[k] = r_o[k](1 + gw[k]) \ \forall k \in \{0..N-1\} \tag{55}$$

where $g > 0$ fixes the embedding strength. If $w[k] < \frac{-1}{g}$, then $r_w[k]$ is clipped to 0.

The inverse extraction function copies the DFT coefficients of C_o and changes the amplitude of those used at the extraction according to the watermarked vector $\mathbf{r_w}$. It also changes the DFT coefficients of the negatives frequency bins with respect to the symmetry property in order to recover a real array when the IDFT is taken. At last, it quantifies the real array into pixels values in $\{0, \ldots, 255\}$.

B. Human perception model

The watermark's invisibility issue is only tackled by the embedding depth g. But, this action is very limited because the watermark signal, once mapped in the media space, is spread all over the image. Uniform areas of the image are very sensitive to watermark addition so that they only support extremely small embedding depth g, whereas edge areas, for instance, support deeper watermark addition. This issue leads to a spatial domain based human perceptual model giving the amount of noise each pixel can support. We selected the human perception model proposed in [43]. This empirical human perception model gives good experimental results. It is based on the computation of the variance of the 9×9 windowed signal and by normalising the obtained arrays with respect to its maximum value. Thus, $\forall (l, c)\ 0 \leq M(l, c) \leq 1$. For better results, we limit the variances to an upper limit and then normalise them:

$$M'(l,c) = \begin{cases} \mathrm{var}\left(C_o(l-u, c-v) \,|\, (u,v) \in \{-4, \ldots, 4\}^2\right) & \text{if} \quad < M'_{max} \\ M'_{max} & \text{else} \end{cases} \tag{56}$$

$M'(l,c)$ is set to zero near the borders of the picture, i.e. if $(l,c) \in \{0,\dots,3\}^2 \cup (\{L-4,\dots,L-1\} \times \{C-4,\dots,C-1\})$, before normalising: $M = M'/M'_{max}$.

Finally, the watermarked content is given by a new inverse extraction function $Y'(.)$:

$$C_w = Y'(\mathbf{r_w}, C_o) = (1-M) \star C_o + M \star Y(\mathbf{r_w}, C_o) \tag{57}$$

where $\star$ is the pixel-wise product. The influence of this masking function is simply taken into account setting g_e as follows:

$$g_e = g \sum_{l=0,c=0}^{L-1,C-1} M(l,c)/LC$$

C. Receiver Operating Characteristic

Our first experiment is to use a large collection of pictures to estimate the distribution of g_e, and then, to find the parameter g_d that maximises the test's power. For $N > 1024$ and $\omega < 0.01$, this happens for extremely small value of this parameter, i.e. $g_d \gtrsim 0$. It means that the decision rule $\check{D}_1$ turns out to be equivalent to the test $\check{D}_2$ (cf. III-D.2), and by the way, to $\check{D}_R$ (cf. III-D.3). This experiment confirms the Dzhaparidze's theorem about the asymptotic optimality of the Rao test.

Our second experiment draws the experimental receiver operating characteristic of the tests $\check{D}_2$, $\check{D}_3$ and $\check{D}_4$ on Fig. 4. We set $N = 2^{14}$, $g = 0.2$, $M'_{max} = 4000$. We use 150 pictures to estimate the distribution of the tested statistics under hypothesis H_0, and we watermark 50 high quality pictures to estimate their distribution under H_1. To increase the number of tries, we use 100 different pseudo-random permutations. Moreover, we also plot the characteristics of the decision rules $\check{D}_2$ and $\check{D}_4$ when the watermark signal is repeated τ times ($\tau = 4$ or 8). This results in new vector lengths $N = 4096$ or $N = 2048$.

The test $\check{D}_3$ is less efficient than the other ones. This reflects the fact that the estimation of parameter g_e at the detection stage is not as easy as it seemed. Especially, if the watermarked content has undergone even a slight compression, this estimation is biased due to the compression noise: such blind attacks, not only tend to decrease the watermark strength, but also, they made g_e overestimated. This pulls down the efficiency of the test conditioning. As stated by R. Blahut in [44], conditioning a test improves its performance in average, but there might be special cases where it spoils its decision. For watermarking techniques, these special cases correspond to common blind attacks, hence it is recommended to give up this test.

The tiling process, of course, increases the performances of the tests. But, we notice that for $N = 2048$, $\check{D}_2$ is clearly less efficient than $\check{D}_4$. These two decision rules are asymptotically equivalent as $N \to \infty$, i.e., in our experimental environment, if $N > 4096$.

The benefit of the tiling processing has been proved only when N is large, as our rationale is based on asymptotic expressions. It clearly means that the gain of the tiling is limited. A huge number of tiles means short vectors whose length is not large enough to enable the statistical test correctly. Fig. 5 draws the power functions P_{de} of the tests $\check{D}_2$ and $\check{D}_4$ for two levels of significance with respect to the number of tiles. For a small number of tiles, the gain of the tiling process is experimentally verified. For bigger numbers of tiles, this gain can not compete with the loss of efficiency due to the non-asymptotical condition: the power function is then decreasing. This clearly shows that, in our experimental environment, the minimum value of N is 2048, which implies a maximum of $\tau = 8$ tiles.

The security criterion is also limiting the number of tiles. A too big number of tiles may enable the pirate to estimate the watermark signal. But, the averaging process in the detector is computed on the permuted coefficients. Hence, the pirate does not know how to average the coefficients to produce a good estimation of the watermark signal.

D. Robustness against a blind attack

We watermark 50 different pictures with 20 different permutations, and compress them with the JPEG algorithm for different quality factor. The power of the test is estimated for two thresholds corresponding to two levels of significance: $\omega = 10^{-3}$ and $\omega = 10^{-4}$. Fig. 6 shows that the decision rules $\check{D}_2$ and $\check{D}_4$ face this blind attack in a similar way. This simulation is far more meaningful than checking the presence of watermarking in one compressed image such as Lena. Even for very low quality factor, the power of the test is not null. There exist contents where we can still detect the presence of the watermark. We could have presented the same experience only with one of these contents pretending the technique is robust to JPEG compression. This would have been of course absolutely not demonstrative. The power function brings far more information about the robustness against the attack.

E. Attack on the detector

The last figure 7 is the result of the oracle attack described in subsection IV-B.2. It shows the number of couples of indices that set the output of the detector to one. This number is given as a percentage of the N^2 couples. The threshold is higher for lower levels of significance. Hence, the lower is ω, the lower is the percentage. For practical use, ω is so small that the oracle attack does not provide any information leakage of the secret detection key.

VI. Conclusion

this article presents the concept of asymmetry in watermarking and it details one possible method. This method is versatile, as it can be adapted to a large number of watermarking techniques based on DSSS. On the other hand, we only studied its advantages in the copy protection framework. Our rationale was to describe the targeted application, to analyse the possible threats, and then to estimate the complexity of each class of attacks. This defines the security level that the watermarking technique provides to the global copy protection system.

To the best of our knowledge, a *watermarked content only attack* is not possible with this method whereas it is a real threat for DSSS and WCS schemes. A *known cover content attack* requires a brute force attack of size $O(2^L)$, whereas a single pair of watermarked / original content is theoretically enough to disclose the secret key in DSSS techniques. The *oracle attack* needs $O(N^2)$ tries whereas T. Kalker proves $O(N)$ tries are sufficient for DSSS techniques.

The prize to be paid is the larger length of the vectors: The asymmetric detectors need more complexity, more memory and they accumulate a bigger amount of content in order to take a reliable decision. Our future works are the invention of asymmetric schemes with better performances thanks to the side information at the embedding stage.

References

[1] S. Craver and al., "Sdmi challenge information," `http://www.cs.princeton.edu/ sip/sdmi/`.

[2] J. Stern and S. Craver, "Lessons learned from the SDMI," in *Proc. of the Multimedia Signal Processing Workshop*, Cannes, France, October 2001.

[3] A. Kerckhoffs, "La cryptographie militaire," *Journal des sciences militaires*, vol. 9, pp. 5–38, janvier 1883.

[4] S. Singh, *The code book*, Fourth Estate Limited, 1999, Histoire des codes secrets, publié chez JC Lattès.

[5] C. Cachin, "An information-theoretic model for steganography," in *Proc. of the second Int. Workshop on Information Hiding*, D. Aucsmith, Ed., Portland, Oregon, U.S.A., April 1998, vol. 1525 of *Lecture Notes in Computer Science*, pp. 306–318, Springer Verlag.

[6] S. Craver, N. Memon, B.-L. Yeo, and M.M. Yeung, "Resolving rightful ownership with invisible watermarking techniques: limitations, attacks, and implications," *IEEE Journal of selected areas in communications*, vol. 16, no. 4, pp. 573–87, May 1998, Special issue on copyright and privacy protection.

[7] M. Kutter, S. Voloshynovskiy, and A. Herrigel, "Watermark copy attack," in *Security and Watermarking of Multimedia Contents II*, P.W. Wong and E. Delp, Eds., San Jose, Cal., USA, January 2000, vol. 3971, SPIE Proceedings.

[8] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, December 1997.

[9] I. Cox, M. Miller, and A. McKellips, "Watermarking as communication with side information," *Proc. of the IEEE*, vol. 87(7), pp. 1127–1141, July 1999.

[10] D. Kahn, "Cryptology and the origins of spread spectrum," *IEEE spectrum*, pp. 70–80, September 1984.

[11] J. Proakis, *Digital Communications*, Electrical and computer engineering. McGraw Hill, third edition, 1996.

[12] I. Cox and J.-P. Linnartz, "Some general methods for tampering with watermarks," *IEEE Journal on selected areas in communications*, vol. 16, no. 4, pp. 587–93, May 1998, Special issue on copyright and privacy protection.

[13] J.A. Bloom, I.J. Cox, T. Kalker, J.-P. Linnartz, M.L. Miller, and C.B.S. Traw, "Copy protection for DVD video," *Proc of the IEEE*, vol. 87, no. 7, pp. 1267–1276, July 1999, Special issue on identification and protection of multimedia information.

[14] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Trans. on information theory*, vol. 22, no. 6, pp. 644–54, November 1976.

[15] T. Kalker, "A security risk for publicly available watermark detectors," in *Benelux Information Theory Symposium*, May 1998, Veldhoven, The Netherlands.

[16] A. Patrizio, "DVD privacy: It can be done," `http://www.wired.com/news/technology/ 1,1282,32249,00.html`.

[17] M.H.M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory*, vol. 29, no. 3, May 1983.

[18] B. Chen, *Design and analysis of digital watermarking, information embedding, and data hiding systems*, Ph.D. thesis, Massachusetts Institute of Technology, 2000.

[19] J. Chou, S. Pradhan, and K. Ramchandran, "Turbo coded trellis-based constructions for data embedding: channel coding with side information," in *Proc. of the 35th Conf. on Signals, Systems and Computers*, Asilomar,CA,USA, November 2001.

[20] J. Eggers and B. Girod, *Informed Watermarking*, Kluwer Academic Publishers, 2002.

[21] J. Chou, S. Pradhan, and K. Ramchandran, "A robust blind watermarking scheme based on distributed source coding principles," in *Proc. of ACM multimedia conference*, Los Angeles, CA, USA, October 2000.

[22] M. Mansour and A. Tewfik, "Secure detection of public watermarks with fractal decision boundaries," in *XI European Signal Processing Conference, EUSIPCO'02*, Toulouse, France, September 2002.

[23] F. Hartung and B. Girod, "Fast public-key watermarking of compressed video," in *Proc. IEEE Int. Conf. on Image Processing*, October 1997.

[24] T. Furon and P. Duhamel, "An asymmetric public detection watermarking technique," in *Proc. of the third Int. Workshop on Information Hiding*, A. Pfitzmann, Ed., Dresden, Germany, September 1999, pp. 88–100, Springer Verlag.

[25] J.Eggers, J.Su, and B.Girod, "Public key watermarking by eigenvectors of linear transforms," in *Proc. of the European Signal Processing Conference*, Tampere, Finland, September 2000, EUSIPCO.

[26] R. Van Schyndel, A. Tirkel, and I. Svalbe, "Key independent watermark detection," in *Int. Conf. on Multimedia Computing and Systems*, Florence, Italy, June 1999, vol. 1.

[27] J. Smith and C. Dodge, "Developments in steganography," in *Proc. of the third Int. Workshop on Information Hiding*, A. Pfitzmann, Ed., Dresden, Germany, September 1999, pp. 77–87, Springer Verlag.

[28] J. Stern and J.-P. Tillich, "Automatic detection of a watermarked document using a private key," in *4th Int. Work. on Information Hiding*, Ira S. Moskowitz, Ed., Pittsburgh, PA, USA, April 2001, vol. 2137 of *Lecture Notes in Computer Science*, p. electronic version, Springer.

[29] G. Silvestre, N. hurley, G. Hanau, and W. Dowling, "Informed audio watermarking using digtial chaotic signals," in *Proc. of Int. Conf. on Acoustics, Speech and Signal Processing*, Salt-Lake City, USA, May 2001, IEEE.

[30] T. Furon, I. Venturini, and P. Duhamel, "Unified approach of asymmetric watermarking schemes," in *Security and Watermarking of Multimedia Contents III*, P.W. Wong and E. Delp, Eds., San Jose, Cal., USA, 2001, SPIE.

[31] J. Picard and A. Robert, "Neural networks functions for public key watermarking," in *4th Int. Work. on Information Hiding*, Ira S. Moskowitz, Ed., Pittsburgh, PA, USA, April 2001, vol. 2137 of *Lecture Notes in Computer Science*, pp. 142–156, Springer.

[32] L. de C.T. Gomes, M. Mboup, M. Bonnet, and N. Moreau, "Cyclostationnarity-based audio watermarking with private and public hidden data," in *Proc. of the 109th Convention of Audio Engineering Society*, Los Angeles, CA,USA, September 2000.

[33] T. Furon, N. Moreau, and P. Duhamel, "Audio asymmetric watermarking technique," in *Proc. of Int. Conf. on Audio, Speech and Signal Processing*, Istanbul, Turkey, June 2000, IEEE.

[34] M. Swanson, B. Zhu, A. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Signal Processing*, vol. 66, no. 3, pp. 337–355, May 1998.
[35] T. Furon and P. Duhamel, "Robustness of an asymmetric technique," in *Proc. of Int. Conf. on Image Processing*, Vancouver, Canada, September 2000, IEEE.
[36] A. de Rosa, M. Barni, F. Bartolini, V. Cappelini, and A. Piva, "Optimum decoding of non-additive full frame dft watermarks," in *Proc. of the third Int. Workshop on Information Hiding*, A. Pfitzmann, Ed., Dresden, Germany, September 1999, pp. 159–171, Springer Verlag.
[37] A. Menezes, P. Van Oorschot, and S. Vanstone, *Handbook of applied cryptography*, Discrete mathematics and its applications. CRC Press, 1996.
[38] E. Lehmann, *Testing statistical hypothesis*, J. Wiley & Sons, 1986.
[39] K. Dzhaparidze, *Parameter estimation and hypothesis testing in spectral analysis of stationary time series*, Springer Verlag in Stastitics, 1986.
[40] P.J. Brockwell and R.A. Davis, *Time Series: Theory and methods*, Springer Verlag in Stastitics, 1991.
[41] G. Fay, *Théorèmes limite pour les fonctionnelles de périodogramme*, Ph.D. thesis, Ecole Nationale Supérieure des Télécommunications, 2000.
[42] J. Stern, *Contribution à la théorie de la protection de l'information*, Ph.D. thesis, Université de Paris XI, Orsay, Laboratoire de Recherche en Informatique, mars 2001.
[43] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "A new decoder for the optimum recovery of non-additive watermarks," *IEEE Trans. on Image Processing*, vol. 5, pp. 755–66, 2001.
[44] R.E. Blahut, *Principes and practice of information theory*, Addisson-Wesley, 1987.

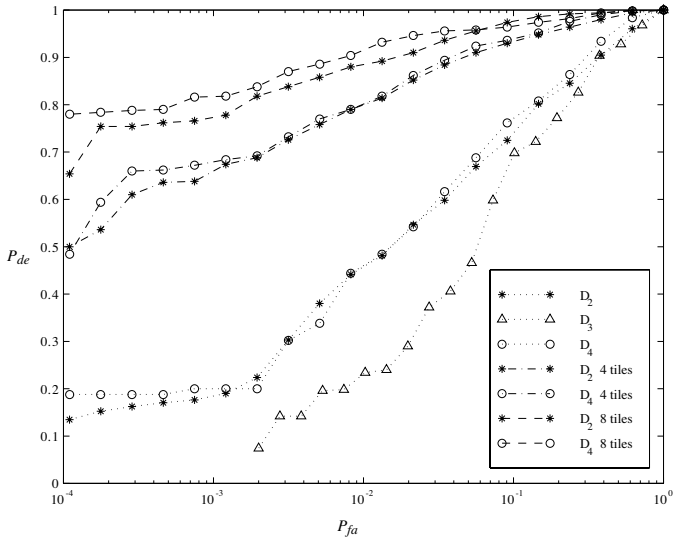

Fig. 4. Receiver Operating Characteristic for three detection strategies. $\check{D}_3$ is too sensitive to the quality of the contents. It gives poor performances. $\check{D}_2$ and $\check{D}_4$ have the same performances. We also tile the watermark signal in order to increase the power ratio G at the detection side. On the other hand, this shortens the vectors whereas the tests were only assessed asymptotically.

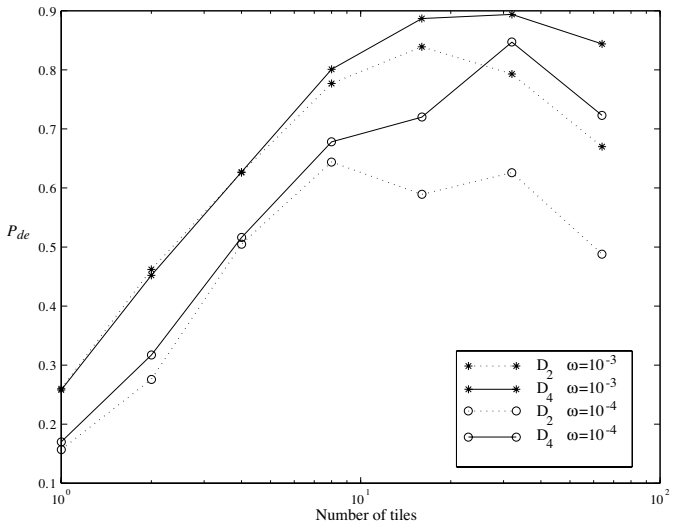

Fig. 5. Influence of the tiling process on power functions. The advantage of tiling is to increase the watermark to content power ratio G resulting in better test powers. Its drawback is to shorten the vectors' length so that the tests are no longer asymptotic. For this watermarking technique, a good trade-off is a number of $\tau = 8$ tiles.

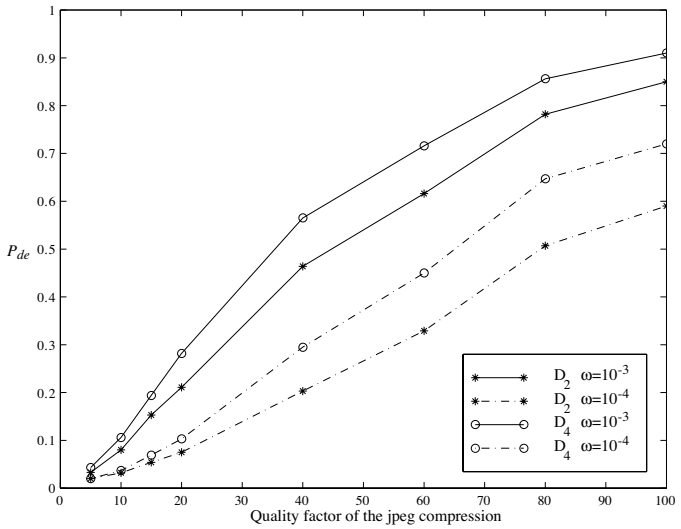

Fig. 6. Robustness against a JPEG compression. This figure gives the probability that watermarked contents, which have supported a JPEG compression attack, are still detected.

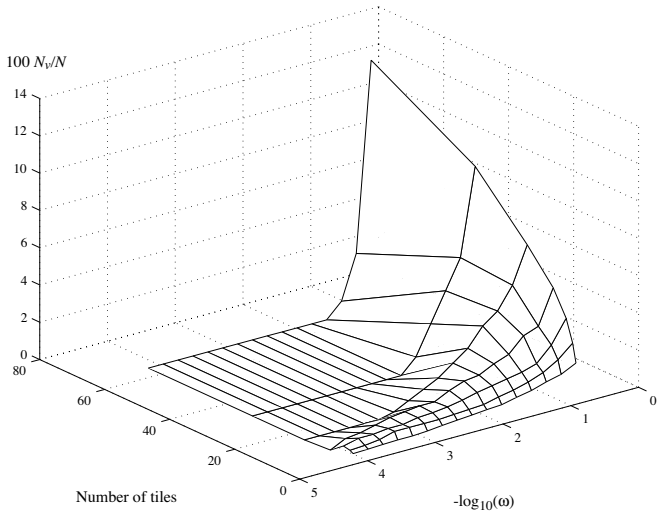

Fig. 7. The oracle attack takes an inventory of couples of indices (i, j) leaking some information about the detection secret. The figure shows the percentage of these couples for different numbers of tiles τ and different significance levels ω.