An unified approach of asymmetric watermarking schemes

Teddy Furon, Ilaria Venturini, Pierre Duhamel

An unified approach of asymmetric watermarking schemes

T. Furon[a], I. Venturini[b], and P. Duhamel[b]

[a]THOMSON multimedia, Security Lab.,
Cesson-Sevigne, 35511 FRANCE

[b]ENST/CNRS, Traitement Signal et Images,
Paris 75013 FRANCE

ABSTRACT

Asymmetric schemes belong to second generation of watermarking. Whereas their need and advantage are well understood, many doubts have been raised about their robustness and security. Four different asymmetric schemes have been proposed up to now. Whereas they were seemingly relying on completely different concepts, they share the same performances. Exploring in detail these concepts, the authors propose a common formulation of the four different detector processes. This allows to stress common features about security of asymmetric schemes.

Keywords: copy protection, asymmetric watermarking, public key schemes, attacks

1. INTRODUCTION

This introduction explains the notions of *asymmetry* and *public key detection*. To clearly state their need and desired advantages, we first focus on a short state of the art.

1.1. State of the art

At the beginning of the watermarking story, researchers named *private detection schemes* techniques which need the original contents (also called the cover contents) at the detection side. Only trusted person could detect watermarks, hence the term private detection. The term *public detection schemes* (also called *blind schemes*) designs techniques where no knowledge of these original contents are required to detect the presence of the watermarks and to retrieve the embedded messages.

The detection may also need a secret key. It grants the ability of detection to a restricted population, or a restricted class of appliances. It also provides more security. As the pirate has not the key, he does not know how the original content has been modified (or what secret signal has been added to the original content). He may only *blindly attack* using coarse compression, low-pass filtering, cropping, scaling etc etc. In still image applications, A. De Rosa and *al.* managed a technique robust against a JPEG compression of quality factor 5%.[1] In video watermarking, joint teams from IBM / NEC, and Philips / Macrovision / Digimarc achieved to fulfill the strict requirements of the MPAA. Yet, some *blind attacks* are still working. The recent SDMI challenge is, for instance, a perfect illustration for audio watermarking domain.[2] But, the robustness against these *blind attacks* is relentlessly improving. Pirates always succeed to forge contents, but the quality of the forge contents is getting worse and worse.

A solution for pirates may be to discover the secret key. In most techniques, this allows them to produce very high quality forged contents (by subtracting the secret signal). Yet, the impact of the secret key disclosure depends on the application. For copyright protection system, it is likely that this secret key depends on the number ID of the author, the registration number of the content, a time stamp, etc etc. Hence, the secret key is different for each copyrighted content. The disclosure of one secret key allows pirates to forge one content. On the opposite, in the copy protection framework, all *compliant devices* share the same secret key and all the protected contents are watermarked in the same way. The disclosure of this unique secret key pulls down the copy protection system because pirates are able to forge all contents.

Author email:
T. F.: furont@thmulti.com;
I. V.: venturi@tsi.enst.fr;
P. D.: pierre.duhamel@lss.supelec.fr

How to find the secret key? A possibility is to 'steal' it. Pirates may reverse engineer the detection process, thus disclose the secret key. This is a real threat in the copy protection framework. Another possibility is to estimate the secret key with classical signal processing tools. This is possible whenever the watermark signal is not properly sealed to the content. Adding the secret signal in uniform areas of images is not only a bad idea from a perceptual point of view but also a deep security flaw. This helps the pirate to estimate part of this secrete key.[3] The same analysis holds when one substitutes part of sound content in frequency bins where human ear is not perceptible with part of the watermark signal. The same danger holds in image watermarking when the secret signal is not low pass filtered.[4] F. Hartung and *al.* stressed conditions to securely bound secrete signal and covert content from a statistical point of view so that the pirate cannot estimate the secret signal with a Wiener filter.[5] Once again, we stress that these attacks are important in copy protection because pirates not only estimate the secret signal for one piece of content but for every protected image. Geometrically speaking, even if they do not estimate exactly the secret, they find an estimated secret key sufficiently colinear to remove enough watermark energy, as mentioned in [4]. More sophisticated attacks on the detection process are also possible.[6] For instance, assuming the pirate has a detector device he tests as many times he likes (which is the case in the copy protection framework), he finally manages to create a forged content using the detector $O(N)$ times. This security flaw is due to the linearity of the detection process.

More theoritical works from Mittelholzer[7] proves that with this conventional spread spectrum scheme, the only way to achieve a perfectly secure system is to never use the same secret signal. Hence, the pirate cannot get any mutual information about the secret key from watermarked contents. This is the Shannon cryptographic theorem translated into the watermarking issue and the previous requirement is the equivalent of the one-time pad stream cipher principle.

1.2. Asymmetry and public key scheme

The problems listed above stem from the symmetry of classical watermarking schemes. The secret watermark signal added by the embedding process is also known by the detector. If the detector could work without knowing the secret signal, it would be less interesting for pirates to reverse engineer the detector. The goal is that no one (even the detector) knows which secret signal has been added. Moreover, if there is a huge amount of possible secret signals, which match the detection capability, the embedder is able to choose one of these signals. If one secret signal has been disclosed, the embedder uses a different signal. This *renewability* feature is highly appreciated all the more so as no change is required at the detection side. Another point is that the embedder can automatically change the secret signal for each piece of content. If pirates succeed to estimate part of one secret signal (due to a flaw listed in subsection 1.1) via one given piece of watermarked content, they can no more improve this estimation with others watermarked contents. The knowledge they acquire is only useful to forge this piece of content, but this does not threaten the whole copy protection system.

Asymmetry stems from this last analysis. The *asymmetric schemes* should be as robust as symmetric techniques with a detector needing a set of parameters called the *detection key* different from the embedding's secret key. For the reasons described above, they are especially dedicated for the copy protection framework.

The understanding of a *public key watermarking scheme* by the community encloses the following requirements (this list is part of a discussion between S. Katzenbeisser, J. Bloom, J. Eggers and T. Furon via the mailing list[8]):

- An original content is marked with a private key. The presence of a watermark can be checked by using a *public key*. Both embedding and detection algorithms are computationally feasible. The publication of the *public key* does not allow the computation of the corresponding private key (in reasonable time).
- The original content is not required in the detection process.
- The knowledge of the *public key* does not allow watermark removal.
- The knowledge of the *public key* does not allow an attacker to forge a pirated content (a watermarked content modified so that it cannot any more be considered as protected by the detection).

Thus, a *public key watermarking scheme* is an improvement with respect to an *asymmetric scheme* because it has been proven that the disclosure of the *detection key* strictly confers no advantage for pirates. Hence, this *detection key* is called *public key*. Note that, J. Smith and C. Dodge already introduced these notions in terms of *weak public key* (for asymmetric schemes) and *strong public key* (for public key watermarking).[9]

2. FOUR DIFFERENT SCHEMES

In this section, we introduce the four different asymmetric watermarking schemes which we are aware of. We give the expression of the four different detection processes $D_S(.), D_E(.), D_V(.)$ and $D_F(.)$.

2.1. Notation

To compare these different methods, we need to introduce common notations. This terminology comes from articles [10] and [11].

From an original content $\mathbf{C_o}$ belonging to the "media space" (for instance the pixel's domain for still images) the extraction function $\mathbf{X}(.)$ maps it into a vector in the "watermark space": $\mathbf{r_o} = \mathbf{X}(\mathbf{C_o})$. The "watermark space" is, for instance, a set of DCT coefficients for several fixed frequencies. Extracted vectors are in this article supposed to be central, i.e. $\mu_{\mathbf{r_o}} = E\{\mathbf{r_o}\} = [0, 0, \ldots, 0]^T$. The role of the "mixing function" $\mathbf{f}(.,.)$ is to modify the extracted vector $\mathbf{r_o}$ into a vector $\mathbf{r_w} = \mathbf{f}(\mathbf{r_o}, \mathbf{w})$ which is sufficiently similar to the vector $\mathbf{r_w}$. Usually, this "mixing function" is an addition where $\mathbf{r_w}$ is modulated by a local gain control represented by the vector $\mathbf{g}$: $\mathbf{r_w} = \mathbf{r_o} + \mathbf{w} \star \mathbf{g}$ where $\star$ is the term by term product. The vector $\mathbf{w}$ is normalized so that $\sigma_w = 1$. The embedding strength is controlled via the vector $\mathbf{g}$. For simplicity sake, we restrict $\mathbf{g}$ to be a constant vector $\mathbf{g} = g[1, 1, \ldots, 1]^T$. We define the watermark to original signal power ratio $G = g^2/\sigma^2_{r_o}$. The application of the "inverse extraction" function $\mathbf{X^{-1}}(.)$ concludes the embedding process. It maps back from the "watermark space" to the "media space": $\mathbf{C_w} = \mathbf{X^{-1}}(\mathbf{r_w}, \mathbf{C_o})$. $\mathbf{C_w}$ is the watermarked content. The "inverse extraction" needs the original content because the vector $\mathbf{r_w}$ is not sufficient to fully describe a content. In other words, the "media space" has a bigger dimension than the "watermark space", and, $X(.)$ is not a bijection function. For instance, $X^{-1}(.)$ replaces some of the DCT coefficients of the original image by the ones stored in the watermarked vector $\mathbf{r_w}$, and then, processes the inverse DCT to obtain the watermarked image.

The detection process tests the received content $\mathbf{C_u}$. It extracts the vector $\mathbf{r_u} = \mathbf{X}(\mathbf{C_u})$ and decides whether it belongs to "watermark detection region", noted $\mathcal{R}_w$, which is the set of all points that the detector categorizes as containing the watermark. Indeed, this region is defined via a comparison of the detection function $D(.)$ with a threshold T. Hence, $\mathcal{R}_w = \{\mathbf{r} | D(\mathbf{r}) > T\}$, where T is a positive threshold depending on requirements of the global system, especially the false alarm probability.

Because $\mathbf{r_u}$ is a priori not known by the detection process, $D(\mathbf{r_u})$ has to be considered as a random variable. Its statistical properties, like mean and variance, depend on two hypothesis. $\mathcal{H}_0$ is the hypothesis where $\mathbf{r_u}$ corresponds to a non watermarked content, whereas $\mathcal{H}_1$ is the hypothesis where $\mathbf{r_u}$ corresponds to a watermarked content. We define the detection process efficiency by the ratio e, whose expression is as follows

$$e = \frac{\sqrt{2}(\mu_{D(\mathbf{r_u}|\mathcal{H}_1)} - \mu_{D(\mathbf{r_u}|\mathcal{H}_0)})}{\sqrt{\sigma^2_{D(\mathbf{r_u}|\mathcal{H}_1)} + \sigma^2_{D(\mathbf{r_u}|\mathcal{H}_0)}}} \tag{1}$$

Given a probability of false alarm and the probability density function of the variable $D(\mathbf{r_o})$, one is able to fix the threshold T to achieve such probability of false alarm. Then the power of the test, which is the ability of the detector to identify watermark contents, depends on the ratio e. The greater e is, the bigger is the test's power and the more efficient is the detector.

2.2. The Smith's proposal

J. Smith and C. Dodge proposed a simple asymmetric scheme in [9]. The embedding process hides in the first part of the extracted vector $\mathbf{r_o}$ a random Gaussian sequence $\mathbf{w}$. Its length is $N/2$ (N is even). It is also hidden in the second part of this extracted sequence.

$$r_w[n] = r_o[n] + gw[mod(n, N/2)] \quad \forall n \in \{0, \ldots, N-1\} \tag{2}$$

where $mod(.,.)$ is the modulo function.

The detector process creates two vectors $\mathbf{r_u^1}$ and $\mathbf{r_u^2}$ of length $N/2$ each. $\mathbf{r_u^1}$ corresponds to the first part of the extracted vector $\mathbf{r_u}$ whereas $\mathbf{r_u^2}$ corresponds to its second part. The detection process is then:

$$c_S = D_S(\mathbf{r_u}) = \frac{\mathbf{r_u^1}^T.\mathbf{r_u^2}}{N/2} \tag{3}$$

Its statistical expectation $E\{c_S\}$ is equal to $2(g^2E\{\|\mathbf{w}\|^2\} + gE\{\mathbf{w}^T.\mathbf{r_o^1} + \mathbf{w}^T.\mathbf{r_o^2}\} + E\{\mathbf{r_o^1}^T.\mathbf{r_o^2}\})/N$. Under the hypothesis $\mathcal{H}_0$, where the unknown content is not watermarked, g has a null value. So that the authors expect $E\{c_S|\mathcal{H}_0\} = 2E\{\mathbf{r_o^1}^T.\mathbf{r_o^2}\}/N = 0$, assuming that the first part of extracted vector is sufficiently uncorrelated with its second part. Under the hypothesis $\mathcal{H}_1$, where the unknown content has been watermarked, g is likely to have a small value compared to the power of the extracted vector $\sigma^2_{r_o}$, i.e. $G \ll 1$. Assuming $\mathbf{w}$ and $\mathbf{r_o^i}$ are uncorrelated because independent stochastic processes, $E\{c_S|\mathcal{H}_1\} = g^2 + 2gE\{\mathbf{w}^T.\mathbf{r_o^1} + \mathbf{w}^T.\mathbf{r_o^2}\}/N = g^2$. But, in real implementation, the detection output c_S is not a binary value in $\{0, g^2\}$. It is, indeed, a random variable centered on 0 if $\mathcal{H}_0$ is true, or on g^2 if $\mathcal{H}_1$ is true, and it shares its probability distribution function with $2(\mathbf{r_o}^{1^T}.\mathbf{r_o}^2 + 2g\mathbf{w}^T.\mathbf{r_o}^i)/N$. The variance of this variable is $\sigma^2_{c_S} = 2(\sigma^4_{r_o} + 2g^2\sigma^2_{r_o})/N$. This leads to the following efficiency

$$e = \frac{g^2\sqrt{N}}{\sigma^2_{r_o}\sqrt{1 + \frac{g^2}{\sigma^2_{r_o}}}} = \frac{G\sqrt{N}}{\sqrt{1+G}} \propto G\sqrt{N} \tag{4}$$

2.3. The Van Schyndel and Eggers' proposal

We introduce in this subsection an idea first developed by R. Van Schyndel, A. Tirkel and I.D. Svalbe [12], and improved by J. Eggers, J. Su and B. Girod in [13] and [14].

Let the $N \times N$ matrix $\mathbf{A}$ represents a linear application in the "watermark space", which is isomorphic to R^N. The watermark signal represented by the vector $\mathbf{w}$ of length N is chosen so as to be an eigenvector of this application: $\mathbf{Aw} = \lambda_0\mathbf{w}$. The detection function is defined as follows:

$$c = D(\mathbf{r_u}) = \frac{\mathbf{r_u}^T.\mathbf{Ar_u}}{N} \tag{5}$$

Its statistical expectation $E\{c\}$ is equal to $(E\{\mathbf{r_o}^T.\mathbf{Ar_o} + g\mathbf{r_o}^T(\mathbf{A}+\mathbf{A}^T)\mathbf{w} + g^2\lambda_0\|\mathbf{w}\|^2\})/N$. Under the hypothesis $\mathcal{H}_0$, where the unknown content is not watermarked, g has a null value. The authors expect, then, $E\{c|\mathcal{H}_0\} = E\{\mathbf{r_o}^T.\mathbf{Ar_o}\}/N = 0$, assuming that the matrix has a sufficiently decorrelating property. If $\mathbf{r_o}$ is a zero-mean, stationary random process, the matrix has to render $\mathbf{Ar_o}$ uncorrelated with $\mathbf{r_o}$. Under the hypothesis $\mathcal{H}_1$, where the unknown content has been watermarked, assuming $\mathbf{w}$ and $\mathbf{A}^T\mathbf{w}$ are uncorrelated with $\mathbf{r_o}$ because independent stochastic processes, $E\{c|\mathcal{H}_1\} = g^2\lambda_0$. The authors roughly investigate the variance of the variable c: the variance of $(\mathbf{r_o}^T(\mathbf{A}+\mathbf{A}^T)\mathbf{w})/N$ is likely to decrease proportional to $\sigma^2_{r_o}/N$ as N increases, whereas the variance of $(\mathbf{r_o}^T.\mathbf{Ar_o})/N$ is proportional to $\sigma^4_{r_o}/N$. This leads to the following efficiency:

$$e = \frac{\lambda_0 g^2\sqrt{2N}}{\sqrt{2\alpha\sigma^4_{r_o} + 2g^2\beta\sigma^2_{r_o}}} = \frac{\lambda_0 G\sqrt{N}}{\sqrt{\alpha + \beta G}} \propto G\sqrt{N} \tag{6}$$

R. Van Schyndel and *al.* first proposed $\mathbf{A}$ to be the $N \times N$ discrete Fourier transform matrix $\mathbf{A_V}$. J. Eggers and *al.* chose $\mathbf{A}$ to be a pseudo-random permutation matrix $\mathbf{A_E}$, in order to increase the robustness against malicious attacks (cf. subsection 5.2). It means that $\mathbf{A_E}$ is a sparse matrix where there is only one 1 in every column and row, the other values being set to 0. If $\pi(.)$ is the pseudo-random permutation function, $\mathbf{A_E}[i, \pi(i)] = 1$. $\mathbf{A_E}^T$ corresponds to the inverse permutation defined by the function $\pi^{-1}(.)$.

2.4. The Furon's proposal

This is the most complex asymmetric watermarking scheme. It has been firstly proposed in [15]. We presented a more robust version in [16].

The vector $\mathbf{w}$ is a filtered Gaussian central white noise $\mathbf{v}$ with unity variance. It is interleaved by a pseudo-random permutation $\pi(.)$, whose matrix is noted $\mathbf{\Pi}$, before embedding: $\mathbf{r_w} = \mathbf{r_o} + g\mathbf{\Pi}(h \otimes \mathbf{v})$. The filter h is normalized so that $\int |H(f)|^2 df = 1$ and thus $\sigma_w = 1$. From now on, we introduce the tilde notation to note the interleaved version of a vector: $\tilde{\mathbf{r}} = \mathbf{\Pi r}$.

The detection process needs the amplitude of the frequency response of the filter h and the inverse permutation $\pi^{-1}(.)$. A simple hypothesis test decides to which hypothesis the unknown content $\mathbf{C_u}$ is more likely to belong:

• $\mathcal{H}_0$: The extracted signal $\mathbf{r_u}$ is not watermarked. Assuming that the inverse permutation acts like a perfect whitening process, the estimated spectrum $S_0(f)$ of its de-interleaved version $\widetilde{\mathbf{r_u}}$ is flat. $S_0(f) = \sigma^2_{r_u}$ where $\sigma^2_{r_u}$ is the variance of the tested extracted vector $\mathbf{r_u}$.

• $\mathcal{H}_1$: The extracted signal $\mathbf{r_u}$ has been watermarked. Assuming $\mathbf{w}$ and $\mathbf{r_o}$ are statistically independent and stationary random vectors, the following relations hold:

$$\begin{aligned} \varphi_{\widetilde{\mathbf{r_u}}}[l] &= E\{\widetilde{\mathbf{r_u}}[m].\widetilde{\mathbf{r_u}}[m+l]\} = \varphi_{\widetilde{\mathbf{r_o}}}[l] + g^2\varphi_{\mathbf{w}}[l] \\ S_1(f) &= \Phi_{\widetilde{\mathbf{r_o}}}(f) + g^2\Phi_{\mathbf{w}}(f) \end{aligned}$$

where $\varphi_{\widetilde{\mathbf{r_o}}}[.]$ is the correlation function of the de-interleaved version of the vector $\mathbf{r_o}$ and $\Phi_{\widetilde{\mathbf{r_o}}}(.)$ its Fourier transform, which is the power spectral density of this vector. Because $\mathbf{w} = h \otimes \mathbf{v}$, $\Phi_{\mathbf{w}}(f) = |H(f)|^2$. Finally, the power spectral density expected if $\mathcal{H}_1$ is true, is:

$$\begin{aligned} S_1(f) &= \sigma^2_{r_o} + g^2|H(f)|^2 \\ &= \sigma^2_{r_u} + g^2(|H(f)|^2 - 1) \end{aligned}$$

because $\mu_{r_u} = \mu_{r_o} = 0$ and $\sigma^2_{r_u} = \sigma^2_{r_o} + g^2$. Thus, under this hypothesis, the estimated spectrum $S_1(f)$ of the extracted central signal r_u is shaped like $|H(f)|^2$.

The detection function is defined as a hypothesis test in spectral analysis based on a maximum likelihood:

$$D_F(\mathbf{r_u}) = U_N(\widetilde{\mathbf{r_u}}, S_0) - U_N(\widetilde{\mathbf{r_u}}, S_1)$$

where $U_N(\mathbf{x}, S_i)$ is the Whittle's principal part of the likelihood that the spectrum of the random vector $\mathbf{x}$ matches the power spectral density $S_i(f)$. Its simplified expression is

$$U_N(\mathbf{x}, S_i) = \sum_{k=1}^{\frac{N}{2}} \log(S_i(f_k)) + I_N(f_k)/S_i(f_k)$$

where $I_N(f)$ is the periodogram of the vector $\mathbf{x}$: $I_N(f) = \left|\sum_{k=0}^{N-1} \mathbf{x}[k].e^{2\pi ikf}\right|^2 \ \forall f \in]-\frac{1}{2}, \frac{1}{2}]$ and f_k is a Shannon frequency $f_k = k/N$. Finally, the detection process is formulated by:

$$c_F = D_F(\mathbf{r_u}) = \sum_{k=1}^{N/2} log(\frac{S_0(f_k)}{S_1(f_k)}) + I_N(f_k)(\frac{1}{S_0(f_k)} - \frac{1}{S_1(f_k)}) \tag{7}$$

Under sufficiently mild conditions, the periodograms values $I_N(f_k)$, which are here random variables, have nice asymptotic properties. As $N \to \infty$, they become mutually independent and identically distributed as a central χ_2 with two degrees of freedom. The following relations turn to be true: $\mu_{I_N(f_k)|\mathcal{H}_i} = \sigma_{I_N(f_k)|\mathcal{H}_i} = S_i(f_k) \qquad i \in \{0, 1\}$. As $N \to \infty$, the Riemann sums converge to their integral forms, i.e.: $\sum_{k=1}^{N/2} log(\frac{S_0(f_k)}{S_1(f_k)}) \to N/2 \int_0^{1/2} log(\frac{S_0(f)}{S_1(f)})df$. After some cumbersome calculus, the efficiency is expressed as follows, where $b(f) = |H(f)|^2 - 1$:

$$e = G\sqrt{N}\frac{\int_0^{1/2} \frac{b^2(f)}{(1+Gb(f))}df}{\sqrt{\int_0^{1/2} \frac{b^2(f)(1+(1+Gb(f))^2)}{(1+Gb(f))^2}df}} \propto G\sqrt{N} \tag{8}$$

3. UNIFIED DESCRIPTION

The motivation for this article came from the fact that, despite these four asymmetric methods were invented independently and a priori are based on completely different ideas, their efficiencies (4), (6) and (8) are proportional to $G\sqrt{N}$. We prove now that these methods are indeed based on the same mathematical formulation at the detection process.

This unified description of the detection function is a quadratic form $Q(.)$ in the "watermark space", so that, $D(\mathbf{r}) = Q(\mathbf{r})/N = (\mathbf{r}^T\mathbf{A}\mathbf{r})/N$. The Van Schyndel's and the Eggers's detection functions $D_V(.)$ and $D_E(.)$ are already formulated as quadratic forms, where $\mathbf{A_V}$ is the discrete Fourier transform matrix and $\mathbf{A_E}$ is a permutation matrix. We have to find the matrices $\mathbf{A_S}$ and $\mathbf{A_F}$ that will allow us to write the Smith's and the Furon's detection functions $D_S(.)$ and $D_F(.)$ as quadratic forms.

3.1. Smith's matrix

Let us note $\mathbf{A_S}$ the following matrix:

$$\mathbf{A_S} = \frac{1}{2}\begin{bmatrix} \mathbf{0} & \mathbf{I}_{N/2} \\ \mathbf{I}_{N/2} & \mathbf{0} \end{bmatrix}$$

$\mathbf{I}_{N/2}$ denotes the $\frac{N}{2} \times \frac{N}{2}$ identity matrix. Then, the detection process can be written as:

$$c_S = D_S(\mathbf{r_u}) = \frac{\mathbf{r_u}^T.\mathbf{A_S}\mathbf{r_u}}{N}$$

3.2. Furon's matrix

The periodogram $I_N(f_k)$ of the permuted extracted vector $\widetilde{\mathbf{r_u}}$ can be expressed by a quadratic form:

$$I_N(f_k) = \widetilde{\mathbf{r_u}}^T.\mathbf{P}(f_k)\widetilde{\mathbf{r_u}}$$

where $\mathbf{P}(f_k)$ is a cosine kernel[17]:

$$P(f_k)[m,n] = \cos(2\pi\frac{k}{N}(n-m))$$

As $S_0(.)$ and $S_1(.)$ depend on the variance σ_{r_u} of the extracted vector and the *estimated* embedding depth $\widehat{g}$, the detection process is thus:

$$c_F = D_F(\mathbf{r_u}) = V(\sigma_{r_u}, \widehat{g}) + \frac{\widetilde{\mathbf{r_u}}^T.\mathbf{P}(\sigma_{r_u}, \widehat{g})\widetilde{\mathbf{r_u}}}{N}$$

with $V(\sigma_{r_u}, \widehat{g}) = \sum_{k=1}^{N/2} log(\frac{S_0(f_k)}{S_1(f_k)})$ and $\mathbf{P}(\sigma_{r_u}, \widehat{g}) = N\sum_{k=1}^{N/2}(\frac{1}{S_0(f_k)} - \frac{1}{S_1(f_k)})\mathbf{P}(f_k)$. The permutation of the extracted vector $\mathbf{r_u}$ is made by a matrix $\mathbf{\Pi}$: $\widetilde{\mathbf{r_u}} = \mathbf{\Pi}\mathbf{r_u}$. Finally, up to a constant, the detection process is also a quadratic function with the matrix $\mathbf{A_F}(\sigma_{r_u}, \widehat{g}) = \mathbf{\Pi}^T\mathbf{P}(\sigma_{r_u}, \widehat{g})\mathbf{\Pi}$:

$$c_F = D_F(\mathbf{r_u}) = V(\sigma_{r_u}, \widehat{g}) + \frac{\mathbf{r_u}^T.\mathbf{A_F}(\sigma_{r_u}, \widehat{g})\mathbf{r_u}}{N} \tag{9}$$

We can rewrite the comparison of c_F with the fixed threshold T in the following manner. $T'(\sigma_{r_u}, \widehat{g})$ appears to be an adaptive threshold varying accordingly to the ratio $\widehat{G} = (\widehat{g}/\sigma_{r_u})^2$.

$$\frac{\mathbf{r_u}^T.\mathbf{A_F}(\sigma_{r_u}, \widehat{g})\mathbf{r_u}}{N} \begin{matrix} >^{H_1} \\ <_{H_0} \end{matrix} T'(\sigma_{r_u}, \widehat{g}) \tag{10}$$

where $T'(\sigma_{r_u}, \widehat{g}) = T - V(\sigma_{r_u}, \widehat{g}))$

3.3. Efficiency in the unified approach

We tackle in this subsection the calculus of the detection efficiency in the framework of the unified approach by a quadratic form. The expectation of the detection output is $E\{c\} = E\{Q(\mathbf{r_o}) + g(Q(\mathbf{r_o}, \mathbf{w}) + Q(\mathbf{w}, \mathbf{r_o})) + g^2 Q(\mathbf{w})\}/N$, where $Q(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{A} \mathbf{y}$ is the bilinear form corresponding to the quadratic form $Q(.)$.

Let us recall some classical results about quadratic forms. When $\mathbf{x}$ is a vector of random variables, the expected value of a quadratic form is

$$E\{\mathbf{x}^T \mathbf{A} \mathbf{x}\} = tr(\mathbf{A} E\{\mathbf{x}\mathbf{x}^T\})$$

We note $\mathbf{R_{xx}}$ the covariance matrix: $\mathbf{R_{xx}} = E\{\mathbf{x}\mathbf{x}^T\} - E\{\mathbf{x}\}E\{\mathbf{x}^T\}$. We assumed that we deal only with central extracted vectors, thus $E\{Q(\mathbf{x})\} = tr(\mathbf{A}\mathbf{R_{xx}})$. This result does not depend on the distribution of $\mathbf{x}$. However, the variance of a quadratic form requires that $\mathbf{x}$ follows a multivariate normal distribution. Calculus lead to $\sigma^2_{Q(\mathbf{x})} = 2tr(\mathbf{A}\mathbf{R_{xx}}\mathbf{A}\mathbf{R_{xx}}) + 4E\{\mathbf{x}\}^T \mathbf{A}\mathbf{R_{xx}}\mathbf{A}E\{\mathbf{x}\} = 2tr(\mathbf{A}\mathbf{R_{xx}}\mathbf{A}\mathbf{R_{xx}})$. The expected value and the variance of a bilinear form, in the case of central vectors in input, are then:

$$\begin{aligned} E\{\mathbf{x}^T \mathbf{A} \mathbf{y}\} &= tr(\mathbf{A}\mathbf{R_{xy}}) \\ \sigma^2_{Q(\mathbf{x},\mathbf{y})} &= tr(\mathbf{A}\mathbf{R_{xy}}^T \mathbf{A}\mathbf{R_{xy}}^T) + tr(\mathbf{A}\mathbf{R_{yy}}\mathbf{A}^T\mathbf{R_{xx}}) \end{aligned}$$

We denote $\mathbf{R} = \mathbf{R_{r_o r_o}}/\sigma^2_{r_o}$ the normalized correlation matrix, and $\mathcal{T}(.) = tr(.)/dim(.)$ the normalized trace of a matrix ($dim(.)$ being the dimension of a square matrix). The efficiency is then

$$e = G\sqrt{N} \frac{\mathcal{T}(\mathbf{A}\mathbf{R_{ww}})}{\sqrt{2\mathcal{T}(\mathbf{ARAR}) + G\mathcal{T}(\mathbf{A}(\mathbf{R}\mathbf{A}\mathbf{R_{ww}} + \mathbf{R_{ww}}\mathbf{A}\mathbf{R})) + G^2\mathcal{T}(\mathbf{A}\mathbf{R_{ww}}\mathbf{A}\mathbf{R_{ww}})}} \propto G\sqrt{N} \tag{11}$$

Eq. (11) is very important. It shows that the efficiency e is proportional to $G\sqrt{N}$, in the first order. The efficiency of classical spread spectrum symmetric schemes is proportional to $\sqrt{GN}$. As $G \ll 1$, it means that asymmetric are less efficient than symmetric schemes up to a factor $\sqrt{G}$. If we want to maintain the same efficiency, we may increase the watermark to original power ratio G. This is not really feasible for quality reasons. Another possibility is to increase the length N of the extracted vector by a factor $1/G$. The conclusion is that asymmetric schemes need a bigger amount of contents to detect watermark presence.

3.4. pdf of the detection ouputs

Under the assumption of central normal distributed extracted vector $\mathbf{r_u}$, the detection output c will have a central χ_2 distribution with N degrees of freedom only if $\mathbf{AR}$ is idempotent, i.e. $\mathbf{ARAR} = \mathbf{AR}$. Else, in the general case, the moment-generating function of this quadratic form can be expressed as $M(t) = \prod_{i=1}^{K} (1 - t2\eta_i)^{-\nu_1/2}$ where the η_i are the K distinct, nonzero eigenvalues of $\mathbf{A}\mathbf{R_{r_u r_u}}$. Each eigenvalue has multiplicity ν_i. Let's define some convenient variables: $d_i = -\frac{1}{2\eta_i}$ and $\gamma = \prod_{i=1}^{K} (2\eta_i)^{-\nu_i/2}$. The probability density function of c is then

$$p_C(c) = \frac{\gamma}{\Gamma(\frac{J}{2})} c^{J/2-1} \Phi_2(\frac{\nu_1}{2}, \ldots, \frac{\nu_K}{2}, \frac{J}{2}, d_1 c, d_2 c, \ldots, d_K c)$$

where $\Phi_2(.)$ is the generalized hypergeometric function (extremely cumbersome to calculate), and J is the total number of nonzero eigenvalues. If all the nonzero eigenvalues have even multiplicities, simplifications occur[17]:

$$p_C(c) = \gamma \sum_{i=1}^{K} \sum_{j=1}^{\nu_i/2} G_{ij} \frac{1}{(j-1)!} c^{j-1} e^{d_i c}$$

3.5. Probability of false alarm

The probability of false alarm is defined with $P_{fa} = 1 - P_{\mathcal{H}_0}(c < T) = 1 - \int_{-\infty}^{T} p_{C|\mathcal{H}_0}(c) dc$. In the special case where the pdf is a central χ_2, the probability of false alarm can be calculated via numerical tables. In the general case, the cumulative function $P_{\mathcal{H}_0}(c < T)$ is very cumbersome. In the case of even multiplicities, its expression is

$$P_{\mathcal{H}_0}(c < T) = \gamma \sum_{i=1}^{K} \sum_{j=1}^{\nu_i/2} G_{ij} \frac{(-d_i^{-j})}{(j-1)!} \gamma(j, -d_i T)$$

where the incomplete Gamma function is defined as $\gamma(a,x) = \int_0^x e^{-t}t^{a-1}dt$.

A classical Chernoff's bound can be derived, using the moment-generating function $M(t)$:

$$P_{fa} \leq e^{-tT}M(t) \quad \forall t > 0$$

Then, $t^\star = argmax(tT - log(M(t)))$ is calculated to estimate the best exponential decreasing rate of this bound. We do not explore in detail here this classical method due to a lack of space.

4. INTERPRETATION

In this section, we give our interpretation in terms of energy distribution of the asymmetric watermarking process.

4.1. Hypothesis $\mathcal{H}_0$

We can find one orthogonal real matrix $\mathbf{O_S}$ such that $\mathbf{O_S}^{-1}\mathbf{A_S}\mathbf{O_S}$ is diagonal with real eigenvalues, because $\mathbf{A_S}$ is symmetric. We can also find one unitary complex matrix $\mathbf{O_E}$ such that $\mathbf{O_E}^{-1}\mathbf{A_E}\mathbf{O_E}$ is diagonal with complex eigenvalues, because $\mathbf{A_E}$ is normal (i.e. $\mathbf{A_E}^H\mathbf{A_E} = \mathbf{A_E}\mathbf{A_E}^H$). The same holds for $\mathbf{A_V}$ thanks to its symmetry. Following step by step the Furon's method (interleaving, Fourier transform, square module, weighting spectrum), we can write $\mathbf{A_F} = \mathbf{\Pi}^T\mathbf{F}^H\mathbf{\Lambda_F}\mathbf{F}\mathbf{\Pi}$, where $\mathbf{F}$ is the complex matrix of the discrete Fourier transform and $\mathbf{\Lambda_F} = diag(\{(1/S_1(f_1) - 1/S_0(f_1)), \ldots, (1/S_1(f_{N-1}) - 1/S_0(f_{N-1}))\})$.

In the Smith, Van Schyndel and Eggers methods, the extracted vectors $\mathbf{r_o}$ are to be white stationary processes. Hence, $\mathbf{R_{r_o r_o}} = \sigma_{r_o}^2\mathbf{I}$. Moreover $tr(\mathbf{A_S}) = tr(\mathbf{A_V}) = tr(\mathbf{A_E}) = 0$. We assumed that an arbitrary random permutation cannot leave some fixed point. It proves that $E\{c|\mathcal{H}_0\} = tr(\mathbf{A}\mathbf{R_{r_o r_o}}) = tr(\sigma_{r_o}\mathbf{A}) = 0$. In the Furon's method, $tr(\mathbf{A_F}\mathbf{R_{r_o r_o}}) = tr(\mathbf{F}^H\mathbf{\Lambda_F}\mathbf{F}R_{\tilde{r}_o\tilde{r}_o})$ We assumed the interleaver behaves like a perfect whitening process, then $\mathbf{R}_{\tilde{\mathbf{r}}_\mathbf{o}\tilde{\mathbf{r}}_\mathbf{o}} = \sigma_{r_o}^2\mathbf{I}$. This leads to $E\{c_F|\mathcal{H}_0\} = \sigma_{r_o}^2 tr(\mathbf{\Lambda_F}) = 0 + \mathcal{O}(g^2)$.

As we found a transform matrix $\mathbf{O}$ such that $\mathbf{A} = \mathbf{O}\mathbf{\Lambda}\mathbf{O^{-1}}$, the following interpretation of the detection process is simply that it transforms the extracted vectors in a new basis where its energy spreads uniformly into each basis bin. The use of a permutation matrix for some techniques (subsections 2.3 and 2.4) or the need of the white process assumption for the other ones (subsections 2.2 and 2.3) helps in creating such whitening application $\mathbf{A}$. In this new space basis, the detection process is indeed the following one:

$$c = \frac{1}{N}\sum_{k=1}^{N}\lambda_k|\rho_o[k]|^2$$

where $|\rho_o[k]|^2$ is the energy in the k^{th} component of this new basis. These energies $\{|\rho_o[k]|^2\}$ in the different eigen-subspaces are supposed to be well balanced, so that $E\{|\rho_o[k]|^2\} = \sum|\rho_o[k]|^2/N = \mu_{|\rho_o|^2} \quad \forall k \in \{1..N\}$. The consequence is that $E\{c|\mathcal{H}_0\} = \frac{1}{N}\sum_{k=1}^{N}\lambda_k E\{|\rho_o[k]|^2\} = \mu_{|\rho_o|^2}\sum_{k=1}^{N}\lambda_k = 0$.

4.2. Hypothesis $\mathcal{H}_1$

The watermark vector $\mathbf{w}$ is an independent random process. Then, in each frequency bin, the watermark energy $|\rho_{gw}[k]|^2$ is added in expectation to the energy $|\rho_o[k]|^2$. The embedding of the watermark is then equivalent to "putting some energy" in some selected subspaces. In the Van Schyndel's, Eggers' and Smith's technique, this addition of energy is focused in one subspace whose corresponding eigenvalue is positive. In the Furon's technique, the energy is added in the subspaces (almost) proportionally to their corresponding eigenvalues. The expectation of the detection output is expressed in (12):

$$E\{c|\mathcal{H}_1\} = \frac{1}{N}\sum_{k=1}^{N}\lambda_k(E\{|\rho_{gw}[k]|^2\} - \mu_{|\rho_{gw}|^2}) \tag{12}$$

Notice here that the detection function is indeed, in expectation, the correlation between the eigenvalue vector $\{\lambda_1, \ldots, \lambda_N\}$ and the 'central' energy vector $\{E\{|\rho_{gw}[1]|^2\} - \mu_{|\rho_{gw}|^2}, \ldots, E\{|\rho_{gw}[N]|^2\} - \mu_{|\rho_{gw}|^2}\}$

5. SECURITY POINT OF VIEW

5.1. Matrix disclosure

This subsection focuses on asymmetric schemes which are not public key schemes (see subsection 1.2). It means the detection process is non public. The issue is whether the pirate can estimate the matrix $\mathbf{A}$ choosing special input vectors and observing the output measures:

$$c = \frac{\mathbf{r}^T.\mathbf{A}\mathbf{r}}{N} = \frac{1}{N}(\sum_{k=1}^{N} A_{kk} r[k]^2 + \sum_{1 \leq m < l \leq N} (A_{lm} + g_{ml}) r[m] r[l])$$

If $\mathbf{r} = \mathbf{e}_i$ for some i, where $\mathbf{e}_i = [0, \ldots, 1, \ldots, 0]^T$, then $Nc = A_{ii}$. If $\mathbf{r} = \mathbf{e}_i + \mathbf{e}_j$, then $(A_{ij} + A_{ji}) = Nc - A_{ii} - A_{jj}$. If $\mathbf{A}$ is symmetric, with $N(N-1)/2$ detections, the pirate can fully estimate $\mathbf{A}$. Else, with $N(N-1)/2$ detections, the pirate has estimated an equivalent matrix $(A + A^T)/2$.

What happens now if the pirate can only access to the comparison to the threshold $c > T$? This can be hardly studied. In classical correlation based symmetric schemes, pirates render the binary detection process sensitive. For instance, the extracted vector in input is half of a known watermarked vector. Now, every small change in this vector is likely to produce a detection flip-flop. Studying the detection behavior according to several small changes helps pirates to forge content and especially to estimate the secret key. The detection process, based on a correlation, is linear: the sum of all these small changes in the extracted vector is likely to result in a forged vector.[6] This is not the case with an asymmetric scheme because a quadratic form is not linear. But, it does not prove that zero knowledge of the matrix $\mathbf{A}$ is provided.

5.2. Malicious attacks

We deal now with a public key watermarking scheme. Knowing the extraction function $\mathbf{X}(.)$ and the matrix $\mathbf{A}$, pirates are designing attacks dedicated to one specific scheme. They are called *malicious attacks* in opposition to the *blind attacks* (see 1.1). The goal of these attacks is not necessary to estimate the watermark signal, but to fool the detection process without losing the content quality.

5.2.1. Watermark signal disclosure

Even if pirates know that the watermark energy is focused in a selected subspace (see 4.2), it is practically impossible to estimate what secret signal has been added if the dimension of this subspace is greater than one. Imagine pirates build an orthonormal basis in the subspace, whatever linear combination of the vectors of this basis (under the constraint that the watermark vector energy is unity) is a possible vector $\mathbf{w}$.

5.2.2. The complementary energy distribution attack

Thanks to the interpretation of subsection 4.2, we found the following strategy. Because it is impossible to disclose the watermark signal, whatever noise $\mathbf{n}$ added to the extracted vector, it is an independent process. Thus, its energy is added in expectation to the energy already present in each subspace. Then, if the energy of this noise is focused on subspaces whose corresponding eigenvalue are negative, this will decrease the detection output. This energy should be spread in this optimal way: $E\{|\rho_n[k]|^2\} + E\{|\rho_{gw}[k]|^2\} = cst \quad \forall k \in \{1, \ldots, N\}$. Because $E\{|\rho_n[k]|^2\} > 0$, the constant is then equal to $max(E\{|\rho_{gw}[k]|^2\})$. Whereas the global power of $\mathbf{gw}$ is equal to g^2, the power of the noise added by pirates is equal to $max(E\{|\rho_{gw}[k]|^2\}) - g^2$. In order to maximize the distortion due to this malicious attack, the maximum of the eigenvalues must be as great as possible.

5.2.3. The statistical attack

The strategy of this attack is also to render the expected distribution of the energy in the subspaces $E\{|\rho_u[k]|^2\}$ like the distribution expected by the detection process under the hypothesis $\mathcal{H}_0$. Pirates have all the knowledge to create a new distribution $\rho_p[k]|^2$ as follows:

$$|\rho_p[k]|^2 = |\rho_u[k]|^2 - E\{|\rho_u[k]|^2|\mathcal{H}_1\} + E\{|\rho_u[k]|^2|\mathcal{H}_0\}$$

Then, the expectation of this new distribution is $E\{|\rho_p[k]|^2\} = E\{|\rho_u[k]|^2|\mathcal{H}_0\}$. The problem for pirates, now, is to create an extracted vector $\mathbf{r_p}$ satisfying such energy distribution and leading to a good quality forged content. A simple way is to use the matrix decomposition $\mathbf{A} = \mathbf{O\Lambda O^{-1}}$ as in Eq. (13).

$$\mathbf{r_p} = \mathbf{O} \begin{pmatrix} \frac{|\rho_p[1]|}{|\rho_u[1]|} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{|\rho_p[N]|}{|\rho_u[N]|} \end{pmatrix} \mathbf{O}^{-1}\mathbf{r_u} \tag{13}$$

6. GENERALIZATION AND FUTURE WORK

This section proposes a generalization of the detection process. But we do not have any space to analyze the expected features of such generalized detector. In classical symmetric schemes, detectors are defined via a correlation like in Eq. (14). The vector $\mathbf{a}$ is the secret key and it is the watermark signal added to extracted vectors.

$$D(\mathbf{r}) = \frac{\mathbf{r}^T\mathbf{a}}{N} = \frac{1}{N}\sum_i a_i r_i \tag{14}$$

In asymmetric schemes, detectors are defined via a quadratic form.

$$D(\mathbf{r}) = \frac{\mathbf{r}^T\mathbf{A}\mathbf{r}}{N} = \frac{1}{N}\sum_{i,j} a_{i,j} r_i r_j \tag{15}$$

We define the following natural generalized detection formulation: a n^{th} order detection function is defined by Eq. (16).

$$D(\mathbf{r}) = \frac{\sum_{i_1,\ldots,i_n} a_{i_1,\ldots,i_n} r_{i_1}\ldots r_{i_n}}{N} \tag{16}$$

Our future work will be the study of n^{th} order detection function. We will especially see whether such detection process suffers from the attacks mentioned in section 5.

As far as we know*, the most accomplished *public key watermarking technique* is presented in [18]. With this technique, the detector can retrieve public information by the public key and public and private information by the private key. This detection process is indeed a 4^{th} order detector.

7. CONCLUSION

A common mathematical formulation for asymmetric watermarking schemes was studied in this article. This helps in stating that asymmetric schemes are less efficient that symmetric ones. Longer extracted vectors compensate this shortcoming. The statistical behaviour of the detection output is more complicated than for symmetric schemes.We give an interpretation in term of energy distribution in the eigenspaces of the quadratic form. This helps us to find strategies of attacks available for all these schemes. This concludes that this unified approached is not suitable for public key schemes. More studies are required to fully estimate its security level as an asymmetric scheme.

REFERENCES

1. A. de Rosa, M. Barni, F. Bartolini, V. Cappelini, and A. Piva, "Optimum decoding of non-additive full frame dft watermarks," in *Third International Workshop on Information Hiding*, A. Pfitzmann, ed., pp. 159–171, Springer Verlag, Sept. 1999.
2. S. Craver and al., "Sdmi challenge information," in *http://www.cs.princeton.edu/sip/sdmi/*,
3. J. Fridrich, A.C. Baldoza, and R.J. Simard, "Robust digital watermark based on key-dependent basis functions," in *Second International Workshop on Information Hiding*, D. Aucsmith, ed., pp. 143–157, Springer Verlag, 1998.
4. M. Holliman, N. Memon, and M.M. Yeung, "Watermark estimation through local pixel estimation," in *Security and Watermarking of Multimedia Contents*, P.W. Wong and E. Delp, eds., pp. 134–146, SPIE, 1999.

*We discover this article recently, so that we had no time to analyze in this paper.

5. F. Hartung, J.K. Su, and B. Girod, "Spread spectrum watermarking: malicious attacks and counterattacks," in *Security and Watermarking of Multimedia Contents*, P.W. Wong and E. Delp, eds., pp. 147–158, SPIE, 1999.
6. T. Kalker, "A security risk for publicly available watermark detectors," in *Benelux Information Theory Symposium*, May.
7. T. Mittelholzer, "An infomation-theoritic approach to steganography and watermarking," in *Third International Workshop on Information Hiding*, A. Pfitzmann, ed., pp. 1–17, Springer Verlag, Sept. 1999.
8. messages 35 and upper from the mailing list, "a question about some conception," in *http://www.watermarkingworld.org/WMMLArchive/0008/msg00035.html*,
9. J. Smith and C. Dodge, "Developments in steganography," in *Third International Workshop on Information Hiding*, A. Pfitzmann, ed., pp. 77–87, Springer Verlag, Sept. 1999.
10. I. Cox, M. Miller, and A. McKellips, "Watermarking as communication with side information," *Proc. of the IEEE* **87(7)**, pp. 1127–1141, 1999.
11. M. Miller, I. Cox, and J. Bloom, "Informed embedding: exploiting image and detector information during watermark insertion," in *ICIP'2000*, Sept. 2000.
12. R. Van Schyndel, A. Tirkel, and I. Svalbe, *Key independent watermark detection*, vol. 1, Florence, Italy, June 1999.
13. J.Eggers, J.Su, and B.Girod, "Public key watermarking by eigenvectors of linear transforms," in *EUSIPCO*, (Tampere, Finland), Sept. 2000.
14. J. Eggers, J. Su, and B. Girod, "Asymmetric watermarking schemes," in *Sicherheit in Mediendaten*, (Tampere, Finland), Sept. 2000.
15. T. Furon and P. Duhamel, "An asymmetric public detection watermarking technique," in *Third International Workshop on Information Hiding*, A. Pfitzmann, ed., pp. 88–100, Springer Verlag, Sept. 1999.
16. T. Furon and P. Duhamel, "Robustness of an asymmetric technique," in *ICIP'2000*, Sept. 2000.
17. P.E. Johnson and D.G. Long, "The probability density of spectral estimates based on modified periodogram averages," *Proc. of the IEEE Transactions on Signal* **47(5)**, pp. 1255–1261, 1999.
18. L. de C.T. Gomes, M. Mboup, M. Bonnet, and N. Moreau, "Cyclostationnarity-based audio watermarking with private and public hidden data," *AES* , Sept.