Capacity of data-hiding system subject to desynchronization

Stéphane Pateux, Gaëtan Le Guelvouit and Jonathan Delhumeau

IRISA/INRIA-Rennes
Campus universitaire de Beaulieu
35 042 Rennes, FRANCE

ABSTRACT

Data hiding has been mainly studied in the last years. Many applications are targeted such as copy-rights management, meta-data embedding for rich-media applications, ... In all these applications, it is crucial to estimate what is the capacity of data hiding. Many works have then been made to study watermarking performance considering data-hiding as a kind of channel communication. However in all these studies, an assumption is made about the perfect knowledge of all attacks parameters (may be known in advance or later estimated with attacks modeling). More especially a malicious attacker may biased its attack so that parameters estimation may not be perfect (desynchronization in parameters). Furthermore, random geometrical attacks for images such as proposed by Stirmark benchmark (more generally desynchronization attacks) show that perfect synchronization may not also be achievable. These last kind of attacks are actually the most effective and lack of theoretical modeling for capacity estimation. We then propose a new model for taking into account desynchronization phenomenon in data hiding (coupled with degrading attacks - i.e. optimal SAWGN attacks). Further, thanks to the use of game theory, we state bounds on the capacity that may be obtained by data hiding systems when subject to desynchronization.

Keywords: Watermarking, capacity estimation, game theory, desynchronization attacks, optimal attack, optimal embedding

1. INTRODUCTION

Data hiding has been mainly studied in the last years. Many applications are targeted such as copy-rights management, meta-data embedding for rich-media applications, ... In all these applications, it is then crucial to estimate what is the capacity of data hiding. I.e. how many bits can be reliably embedded and later extracted even if the watermarked content has been altered (compression, filtering, cropping, intentional attacks, ...).

To this purpose recent studies have considered a link between data hiding and communication over a noisy channel. First works have then consider an analogy with an additive white gaussian noise (AWGN) channel.[1] Host signal and attacks being considered as a gaussian noise limiting performances. Later, based on works of Costa on channel with side information,[2] and use of game theory, more realistic limits on performances have been stated.[3,4] Game theory being used to model the game between the hider and the attacker. The attacker defines its attack to be the most effective while the hider tunes its system to best resist to this most effective attack. In these studies, it has especially been shown that optimal attacks are of the kind scaling and additive white gaussian noise (SAWGN). Though, as mentioned in,[5] these upper bounds, associated to the use of parallel gaussian channels do not satisfy a Nash equilibrium, they may be closely reached thanks to the use of wide spread spectrum and side informed coding techniques.

However in all these theoretical studies, an assumption is made about the perfect knowledge of all attacks parameters (may be known in advance or later estimated with attacks modeling). More especially a malicious attacker may bias its attack so that parameters estimation may not be perfect (desynchronization in parameters). Furthermore, random geometrical attacks for images such as proposed by Stirmark benchmark[6] (more generally desynchronization attacks) show that perfect synchronization may not also be achievable. These last kind of attacks are actually the most effective and lack of theoretical modeling for capacity estimation.

Further author information: (Send correspondence to S. Pateux) E-mail: Stephane.Pateux@irisa.fr

We then propose a new model for taking into account desynchronization phenomenon in data hiding (coupled with degrading attacks - i.e. optimal SAWGN attacks). Further, thanks to the use of game theory, we state bounds on the capacity that may be obtained by data hiding systems when being subject to desynchronization.

In section 2 we present our notations and the general concept of additive watermarking using Wide Spread Spectrum (WSS) technique. Section 3 presents the desynchronisation phenomenon and proposes a new model to take it into account. Then in section 4 we present the concept of game that will be used to define atteignable capacities and solve it in section 5. Finally sections 6 and 7 present experimental results and draw some conclusions.

2. WIDE SPREAD SPECTRUM WATERMARKING

As in,[5, 7, 8] we propose to use a Wide Spread Spectrum technique to embed a message in a host signal*. As explained in,[5] we retained this technique since it is a global technique for embedding and extraction in order to search for a Nash Equilibrium of our game†.

We consider a host signal x_i, $i \in [1..n]$ that is independent and non stationary Gaussian, i.e $x_i \sim \mathcal{N}(0, \sigma^2_{X_i})$. In this signal we then embed a message that is embedded using a watermark w_i; $i \in [1..n]$ and $w_i \sim \mathcal{N}(0, \sigma^2_{W_i})$. This watermark is embedded thanks to Wide Spread Spectrum technique and is thus a summation of weighted carriers $G_{i,j}$, $(i,j) \in [1..n] \times [1..m]$. We note w_j^{ST} the associated weighting factors. In the case of blind watermarking not using side information, these weighting factors could be the bits/symbols defining the message, or the bits/symbols of the codeword associated to this message.

Since we will consider Scaling and Additive White Gaussian Noise (SAWGN) attacks, and that in this context, as shown in,[5, 7, 8] Wiener filtering helps limiting distortion while not perturbing capacity, we perform Wiener filtering at embedding.

Finally, after embedding, we can note the watermarked signal y_i as:

$$y_i = \gamma_i^W \left[x_i + \underbrace{\frac{\sigma_{W_i}}{\sqrt{\sum_{j=1}^{m} (G_{i,j})^2}} \sum_{j=1}^{m} w_j^{ST} G_{i,j}}_{w_i} \right] \tag{1}$$

where $\gamma_i^W = \frac{\sigma^2_{X_i}}{\sigma^2_{X_i} + \sigma^2_{W_i}}$ is the scaling factor associated to the Wiener filtering.

Note that in the case of Watermarking exploiting side information, w_j^{ST} depends on the message to embed and the host signal. Since we will use linear extractors to retrieve an estimation of w_j^{ST}, WSS could be seen as working in a linear sub-space of the signal. In such linear sub-space, original host signal will have a response x_j^{ST}. This response will then be used as proposed in[9] to define w_j^{ST} accordingly.

It has been shown in,[3] that optimal attacks against watermarking system is of the kind SAWGN. We will then consider such attacks. We note y'_i the attacked signal:

$$y'_i = \frac{\gamma_i}{\gamma_i^W} y_i + \delta'_i \tag{2}$$

where γ_i is the scaling factor of the attack, and $\delta_i \sim \mathcal{N}(0, \sigma^2_{\delta'_i})$ is the additive noise.

*This technique is also called Spread Transform in other studies.

†As shown in,[5] usual Parallel Gaussian Channels technique proposed in[3, 4] does not lead to a Nash Equilibrium. In fact Game has been resolved knowing in advance what would be the attacker behavior. But a malicious attacker may change its strategy of attack, and for example decide to attack more severely some channels while not distorting other channels, which turns out to be a more efficient attack. In fact, solution obtained in,[3, 4] is a solution to the informed game. But in the case of watermarking, embedder can't be informed of what will be the final strategy of the attacker...

3. MODELING OF DESYNCHRONISATION PHENOMENON

SAWGN attacks are very effective attacks, but as observed in benchmark tools such as Stirmark,[6] these are not especially the most effective. For example, geometric distortions for images are more effective than noise addition.

In fact, we can consider several desynchronisation phenomenons. A first one is linked to "geometrical" attacks (geometrical distortion for images, phase desynchronisation for audio signals, ...). A second one is linked to the estimation of the attack parameters of the SAWGN channel.

Since we are working with Gaussian signals and WSS technique, we will assume that the extraction can be performed with the use of linear correlators to perform demodulation of the carriers [‡]. Thus after demodulation we can observe a correlation responses c_j of the kind:

$$c_j = \sum_{i=1}^{n} \beta_{i,j} y'_i \tag{3}$$

$$\begin{aligned} c_j = & \left\{ \sum_{i=1}^{n} \gamma_i \beta_{i,j} x_i \right\} \\ & + \left\{ w_j^{ST} \sum_{i=1}^{n} \frac{\sigma_{W_i}}{\sqrt{\sum_{j'=1}^{m} (G_{i,j'})^2}} \gamma_i \beta_{i,j} G_{i,j} \right\} \\ & + \left\{ \sum_{k \neq j} w_k^{ST} \sum_{i=1}^{n} \frac{\sigma_{W_i}}{\sqrt{\sum_{j'=1}^{m} (G_{i,j'})^2}} \gamma_i \beta_{i,j} G_{i,k} \right\} \\ & + \left\{ \sum_{i=1}^{n} \beta_{i,j} \delta_i \right\} \end{aligned} \tag{4}$$

where we can identify four terms:

- $\sum_{i=1}^{n} \gamma_i \beta_{i,j} x_i$ is the side information associated to the host signal,
- $w_j^{ST} \sum_{i=1}^{n} \frac{\sigma_{W_i}}{\sqrt{\sum_{j'=1}^{m} (G_{i,j'})^2}} \gamma_i \beta_{i,j} G_{i,j}$ is the response of the symbol transmitted by the j^{th} carrier,
- $\sum_{k \neq j} w_k^{ST} \sum_{i=1}^{n} \frac{\sigma_{W_i}}{\sqrt{\sum_{j'=1}^{m} (G_{i,j'})^2}} \gamma_i \beta_{i,j} G_{i,k}$ is the Inter Symbol Interference (ISI),
- $\sum_{i=1}^{n} \beta_{i,j} \delta_i$ is the added noise.

As proposed in[9] we can remove the ISI term by using orthogonal carriers or by considering the ISI information in the side information.

In a watermarking system we then need to know the statistics of the three remaining terms (side information, added signal for watermarking, and noise). These statistics have to be known at the embedding and also at extraction[§]. The realization of the host signal also needs to be known at embedding in the case of side informed schemes. In these case, it is then necessary to know attack parameters prior to extraction (and also at embedding for side informed schemes) in order to estimate these values. Unfortunately these estimations may be biased on a degraded content, or when proposed models do not fit model really applied (e.g. non linear filtering). In the case of geometrical distortions, signal has to be synchronized (with the help of warping techniques). This

[‡]Here as used in channel communication, extraction will be performed thanks to demodulation followed by closest codeword selection (for example with the use of Error Correcting Codes).

[§]Note that in the resulting scheme we propose, it is not generally necessary to know these statistics.

synchronization may also not be perfect due to a limited precision on the deformation parameters search or to a bias in the model considered (e.g. affine vs. homo-graphic transformations).

In the first class of geometrical desynchronization, we can then consider that the observed signal y_i'' is a desynchronized version of y_i'. In this context we can then generally wrote that:

$$y_i'' = c_i y_i' + n_i \tag{5}$$

where c_i is an attenuation factor and n_i is a desynchronization noise. In the context of a 1-D signal, if we use the discrete to continuous formulation to estimate the value of the signal in any point, we have:

$$y'(t) = \sum_i y_i' sinc(t-i) \tag{6}$$

Thus in this case, if signal is desynchronized from Δ, then:

$$\begin{aligned} y_i'' &= y'(i + Delta) \\ &= y_i \underbrace{sinc(\Delta)}_{c_i} + \underbrace{\sum_{j \neq i} y_j sinc(j - i + \Delta)}_{n_i} \end{aligned} \tag{7}$$

In this case if we are facing noisy desynchronizations, then c_i may vary and we will consider the worst case for it considering a maximum value of Δ. That is, if $\Delta < 1$ then $c_i = sinc(\Delta)$ otherwise $c_i = 0$[¶]. We should note also that the noise n_i is a noise due to self interference of the signal. This self interference noise can be considered as a Gaussian noise with an energy $\sigma_{n_i}^2 = (1 - c_i^2)\sigma_{y_i'}^2$ [‖].

In the context of multi-dimensional signal, this model can easily be extended. We then have $c_i = (sinc(\Delta))^d$ and still $\sigma_{n_i}^2 = (1 - c_i^2)\sigma_{y_i'}^2$, where d is the dimension of the signal (1 for audio, 2 for images, ...).

We can observe that this kind of modeling has been already proposed for desynchronization phenomenon on watermarking system in,[10, 11] but not in an aim of estimation of the capacity. Moreover in[11] c_i term was assumed to be 1 which is not a valid hypothesis when considering side informed watermark scheme.

In the second class of desynchronization on the attack parameters estimation, we consider that estimated γ_i is noisy. That is $\gamma_i = \gamma_i^\star + d\gamma_i$, where $d\gamma_i \sim \mathcal{N}(0, \sigma_{d\gamma_i}^2)$. This additional noise on γ_i will then induce a new term in equation 4:

$$\sum_{i=1}^{n} d\gamma_i \beta_{i,j} x_i \tag{8}$$

On the opposite of the side information, this term can't be assumed to be known at the embedding and will lead to consider part of the watermarked signal as a noise. Note that this influence occurs only when considering side informed embedding technique which were assuming that host signal was not interfering.

Finally, in order to take into account all possible situations, we will then consider the following general formulation:

$$y_i = \gamma_i^W \left[x_i + \underbrace{\frac{\sigma_{W_i}}{\sqrt{\sum_{j=1}^{m} (G_{i,j})^2}} \sum_{j=1}^{m} b_j G_{i,j}}_{w_i} \right] \tag{9}$$

$$y_i' = \frac{\gamma_i}{\gamma_i^W} y_i + \delta_i' \tag{10}$$

[¶]Attacker is then free to take any value of desychronization. The best one for him being the one that nullify the contribution of the awaited response.

[‖]This is simply obtained by assuming that $\sigma_{y_i''}^2 = \sigma_{y_i'}^2$, i.e. that σ_y does not vary too much in a neighborhood.

$$y_i'' = c_i \frac{\gamma_i}{\gamma_i^W} y_i + n_i + \delta_i \tag{11}$$

$$y_i'' = \underbrace{c_i \gamma_i w_i}_{\text{watermark response}} + \underbrace{(c_i \gamma_i x_i + n_i)}_{\text{side information noise}} + \underbrace{\delta_i}_{\text{additive noise}} \tag{12}$$

where :

- x_i is the non iid Gaussian host signal (i.e. $X_i \sim \mathcal{N}(0, \sigma_{X_i}^2)$);
- y_i is the watermarked signal and y_i' is the degraded version after a SAWGN attack;
- y_i'' is the final signal received after a potential geometrical desynchronization;
- n_i is the self interference noise due to the geometrical desynchronization ($n_i \sim \mathcal{N}(0, a_i \gamma_i^2 (\sigma_{X_i}^2 + \sigma_{W_i}^2))$);
- γ_i is the scaling factor applied to coefficient x_i. $\gamma_i^W = \frac{\sigma_{X_i}^2}{\sigma_{X_i}^2 + \sigma_{W_i}^2}$ is the scaling factor associated to the Wiener filtering applied at embedding;
- δ_i and δ_i' are additive white Gaussian noise $\sim \mathcal{N}(0, \sigma_{\delta_i}^2)$.

It should be noted that in the context of non informed watermarking system, n_i term could be omitted. Or equivalently, we can set $a_i = 0$.

4. GAME FORMULATION FOR CAPACITY ESTIMATION

In order to estimate the capacity of a watermarking system subject to desynchronizations, we need to state for a given configuration of attacks (i.e. a set $(\gamma_i, \sigma_{\delta_i}, c_i, a_i)$) the obtainable capacity.

In the same spirit than,[5] we then show that using correlation techniques to demodulate the WSS's carriers, we ends up with an additive gaussian channel whose signal to noise ratio is defined as:

$$\frac{E_b}{N_0} = \sum_{i=1}^{m} \frac{c_i^2 \gamma_i^2 \sigma_{W_i}^2}{a_i \gamma_i^2 (\sigma_{X_i}^2 + \sigma_{W_i}^2) + \sigma_{\delta_i}^2} \tag{13}$$

$\beta_{i,j}$ correlation factors used in equation 3 being defined by:

$$\beta_{i,j} = \frac{c_i \gamma_i \sigma_{W_i}}{a_i \gamma_i^2 (\sigma_{X_i}^2 + \sigma_{W_i}^2) + \sigma_{\delta_i}^2} G_{i,j} \tag{14}$$

It should be noted that equation 13 is a very general formulation. c_i and a_i terms allow to take into account desynchronization phenomenons but not only. For example by setting $a_i = 1$ we can model non informed watermarking system. By taking $a_i > (1 - c_i^2)$ we can take into account non optimal side informed scheme (where part of original signal still influence performances).

Capacity estimation will depend on attack parameters $(\gamma_i, \sigma_{\delta_i}, c_i, a_i)$ and σ_{w_i} that is the distortions that are allowed after embedding and after attacks.

As in,[5] we then consider the following distortions for embedding and after attacks:

$$D_{xy} = \sum_{i=1}^{n} \varphi_i^2 \frac{\sigma_{X_i}^2 \sigma_{W_i}^2}{\sigma_{X_i}^2 + \sigma_{W_i}^2} \tag{15}$$

$$D_{xy'} = \sum_{i=1}^{n} \varphi_i^2 \left(\sigma_{X_i}^2 (1 - \gamma_i)^2 + \gamma_i^2 \sigma_{W_i}^2 + \sigma_{\delta_i}^2 \right) \tag{16}$$

where φ_i is a weighting factor for taking into account perceptual sensibility of the different embedding sites.

We do not directly take into account in attack distortion the impact of the geometrical desynchronisations since this would bias the measure. In fact geometrical distortions generally do not alter the quality of the signal (e.g. signal may be shifted without any perceptual degradations). However in order to limit the amount of geometrical distortion, we play with a_i and c_i parameters. For example using equation 7 we can specify them for an amount of desynchronisation of magnitude Δ.

Given these distortions, theoretical atteignable performances could be defined thanks to the definition of a game between an attacker and an embedder. Under a given budget of distortion, the attacker tries to minimize the Signal to Noise Ratio of the channel, while the embedder tries to maximize it. That is:

$$\begin{cases} \max_{\{\sigma_{W_i}\}} \min_{\{\gamma_i, \sigma_{\delta_i}\}} \frac{E_b}{N_0} \\ D_{xy} \leq D_{xy}^{\max} \\ D_{xy'} \leq D_{xy'}^{\max} \end{cases} \tag{17}$$

5. GAME RESOLUTION

In order to find the solution of equation 17, we use lagrangian formulations, and proceed as in.[8] We first define the optimal attack strategy ($\min \frac{E_b}{N_0}$), and then search the best strategy for the embedder knowing the worst attack ($\max \min \frac{E_b}{N_0}$).

On the attacker side, we consider minimizing the following functional:

$$J_\lambda = \min_{\{\gamma_i, \sigma_{\delta_i}^2\}} \frac{E_b}{N_0} + \lambda D_{xy'} \tag{18}$$

By setting to 0 the derivatives with respect to the attack parameters, we found the general formulation for them:

$$\begin{cases} \gamma_i & = & \frac{\sigma_{X_i}^2 - \frac{c_i \sigma_{W_i}}{\varphi_i \sqrt{\lambda}}}{(1-a_i)(\sigma_{X_i}^2 + \sigma_{X_i}^2)} \\ \sigma_{\delta_i}^2 & = & \gamma_i(\gamma_i^W - \gamma_i)(\sigma_{X_i}^2 + \sigma_{X_i}^2) \end{cases} \tag{19}$$

However, as in,[8,12] there are in fact three attack domains depending on the respective value between the energy of the watermark and the energy of the host signal.

A first domain $\mathcal{D}_E$ is defined for points where $\sigma_{W_i} \leq \frac{\varphi_i \sqrt{\lambda} \sigma_{X_i}^2}{c_i}$. In this domain, the best strategy of attack is to simply nullify the signal (i.e. setting γ_i and σ_{δ_i} to 0).

A second domain $\mathcal{D}_W$ is defined for points where $\gamma_i > \gamma_i^W$ (using equation 19 for the expression of γ_i). In this domain, the best strategy of attack is just to apply a Wiener filter (in fact nothing since this has already be done at embedding).

A third domain that is located between the two previous one (see figure 1), is defined for points where $0 < \gamma_i < \gamma_i^W$ (using equation 19 for the expression of γ_i). In this domain, the best strategy for attack is to use the parameters defined in equation 19.

It should be noted that these domains depend also on the value of desynchronization parameters (a_i, c_i).

The embedding strategy is then defined knowing the optimal attack. We use once again a lagrangian approach which leads to the following problem:

$$J_\chi = \max_{\{\sigma_{W_i}\}} \frac{E_b}{N_0} + \lambda D_{xy'} - \chi D_{xy} \tag{20}$$

where expression for γ_i and σ_{δ_i} are the one of the optimal attack.

When we are in domain $\mathcal{D}_E$, we found the derivative according to σ_{W_i} to be negative. The solution over σ_{W_i} should then be minimal. This lead to retain the solution on the lower boundary of this domain, that is a point which is also in domain $\mathcal{D}_I$. We can then omit the case where we are in domain $\mathcal{D}_E$, since there is no valid solution there.

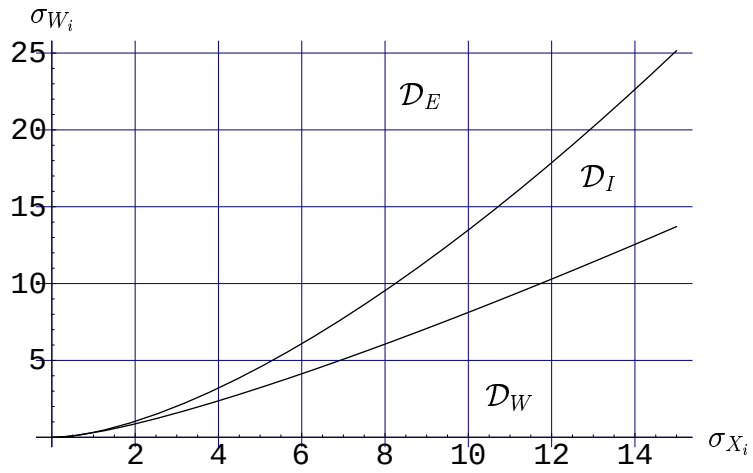


Figure 1. Typical illustration of the domains of attack strategies for the optimal attack.

When we are in domain $\mathcal{D}_I$, we found the following solution:

$$\begin{cases} \text{if } \lambda > \chi \text{ or } \sigma_{X_i} < \frac{c_i}{\sqrt{a_i}\varphi_i\sqrt{\chi-\lambda}}, \\ \quad \sigma_{W_i} = \dfrac{\left[\begin{array}{c} A_i \\ + \sqrt{(A_i)^2 + 4\varphi_i^2\lambda\sigma_{X_i}^2 c_i^2} \end{array} \right]}{2\varphi_i\sqrt{\lambda}c_i}, \\ \quad \text{where } A_i = \varphi_i^2(\lambda - \chi(1-a_i))\sigma_{X_i}^2 - c_i^2 \\ \text{otherwise} \\ \quad \sigma_{W_i} = 0, \end{cases} \tag{21}$$

When we are in domain $\mathcal{D}_W$, the derivative with respect to σ_{w_i} is either always positive, either always negative. That is the solution is on one of the boundary of this domain. Considering the boundary with $\mathcal{D}_I$, this case could be treated by the optimization in domain $\mathcal{D}_I$. On the lower boundary, we have the solution where $\sigma_{W_i} = 0$.

Considering these three situations, it turns out that the solution defined in equation 21 is the general formulation of the solution. We can then observe as in[12] that not all coefficients will be necessarily watermarked. In fact if $\sigma_{X_i} > \frac{c_i}{\sqrt{a_i}\varphi_i\sqrt{\chi-\lambda}}$, coefficients should not be watermarked. This phenomenon is due to the interference with the host signal. Either because we are not using side informed technique ($a_i = 1$), or because signal may suffer from geometrical desynchronization ($a_i \neq 0$). Furthermore, this threshold over σ_{X_i} depends on c_i. When considering coefficients issued from frequency decomposition, since highest sub-bands are more sensitive to desynchronization than lowest sub-bands (their c_i will be lower), they might have no point retained to be watermark (especially when $c_i = 0$). This last remark gives in fact a theoretical justification for using low sub-bands for watermarking.

Moreover we can observe that in the extreme case where $a = 1$ (i.e. non informed watermarking system), we fall back on the previously obtained observation[12] that for such watermarking system, we have $\sigma_{W_i} = \frac{\varphi_i\sqrt{\lambda}\sigma_{X_i}^2}{c_i}$ if $\sigma_{X_i} < \frac{c_i}{\varphi_i\sqrt{\chi-\lambda}}$. The factor c_i has appeared here in this formulation in order to take into account for geometrical desynchronization.

6. EXPERIMENTAL RESULTS

We have applied our watermarking scheme on images in order to estimate what is the capacity of images subject to desynchronization and intentional attacks. Results are illustrated here for the Lena 512x512 gray level image.

A Discrete Wavelet Transform (DWT) with various number of decomposition levels has been applied in order to obtain a set of coefficients to watermark.

In order to estimate a_i and c_i coefficients, we have used the $sinc^2(\Delta_i)$ formulation for c_i. The displacement errors Δ_i depend on the level of the sub-band considered. We have a geometrical evolution of Δ accross levels l of the kind $\Delta(l) = \frac{\Delta}{2^l}$. This formulation is very close to the one we can obtain through numerical computation of the correlation of the wavelet synthesis filter (see figure 2).

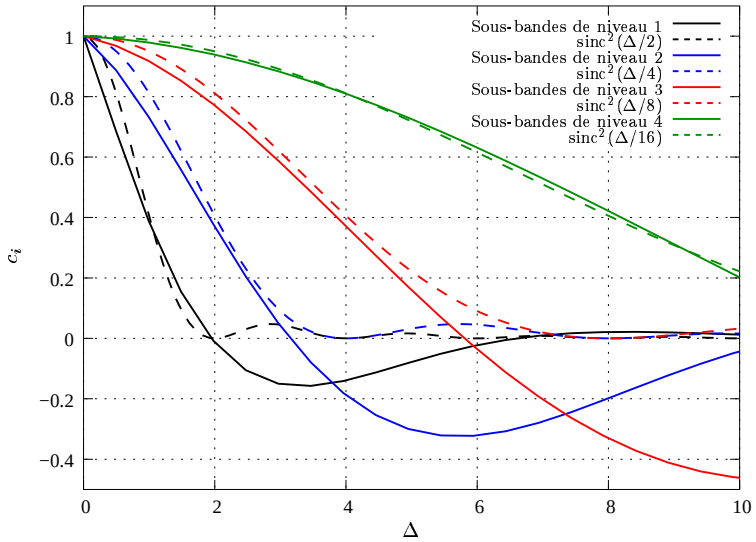


Figure 2. Illustration of the correlation response of wavelet coefficients depending on their decomposition levels.

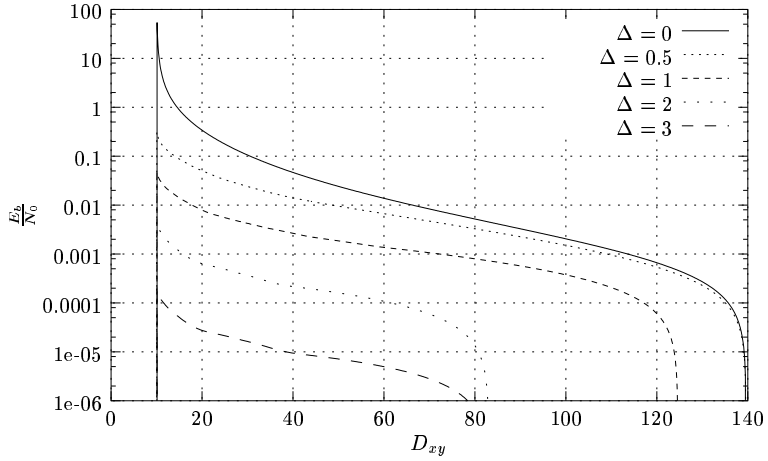
Figure 3.a illustrates estimations of the atteignable Signal to Noise Ratio $\frac{E_b}{N_0}$ using a 3 levels DWT with various desynchronization errors (Δ=0, 0.5, 1, 2 and 3 pixels). Payload may be obtained thanks to the use of formulation of the capacity of a Gaussian channel that is $C = \frac{1}{2}\log_2(1 + \frac{E_b}{N_0})$. Figure 3.b shows the evolution of this SNR when varying the number of levels of decomposition of the DWT, and considering a 2 pixels desynchronization. Just to illustrate the degradation in performance due to geometrical distortion, we have also drawn on this figure the results in the case of no geometrical distortion. Figure 3.c presents a comparison between a watermarking scheme that exploit or not the side information.

From these figures, we can observe the effective degradation of performance due to geometrical attacks. This degradation is even more important when we consider embedding on a few number of DWT levels. For example, on figure 3.a, for 3 DWT levels and an attack of 35 dB, capacity falls down from 56000 bits to less than 5 bits for only 3 pixels desynchronization. These results also show the interest of using side informed watermarking scheme. Indeed to resist to geometrical desynchronization it is necessary to perform embedding in low sub-bands (watermark energy being concentrated in such sub-bands). However since the energy of the host signal is also very important in such sub-bands, without side informed schemes we couldn't reach good level of performance. As an example, by using a 5 levels DWT, it is possible to embed 38 bits while being robust to attacks with PSNR up to 26 dB and geometrical distortions of 2 pixels.

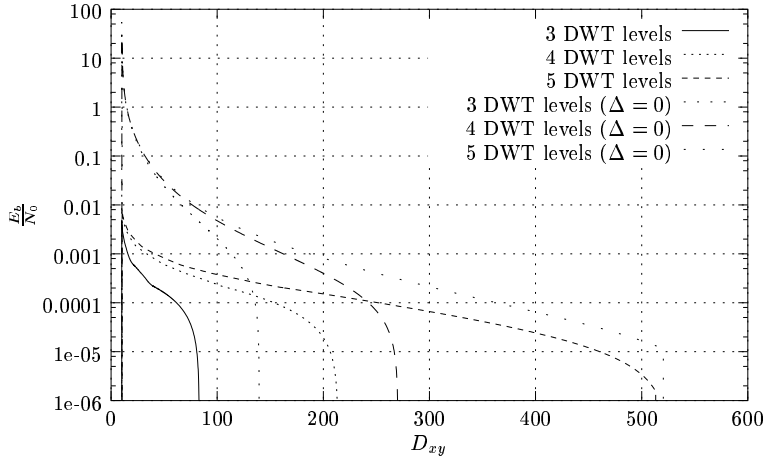
7. CONCLUSION

In this article we have proposed a theoretical modeling of a watermarking system subject to intentional attacks such as SAWGN an geometrical attacks. Thanks to this modeling and the use of game theory, we have stated bounds on the obtainable capacity in such context. Moreover the solution proposed by the resolution of the game leads to a practical implementation of such a watermarking scheme**.

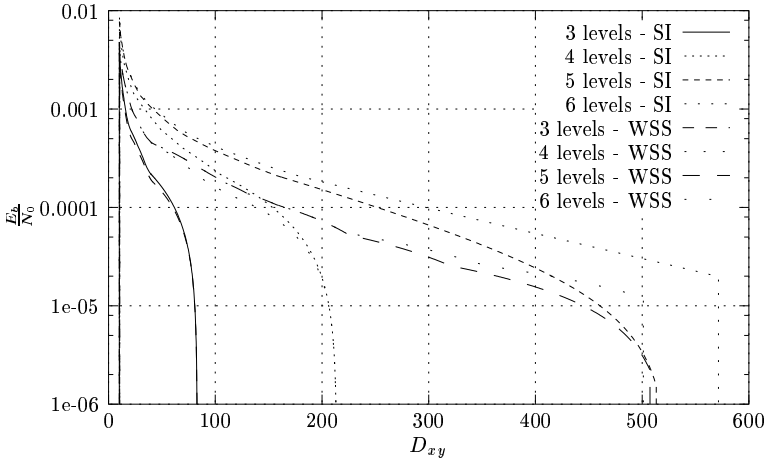
**This is a direct extension of the work proposed in[5] which was only considering SAWGN attacks.



(a) 3 levels DWT



(b) $\Delta = 2$ *pixels*



(c) $\Delta = 2$ *pixels* - watermarking scheme exploiting or not the side information

Figure 3. Capacity estimation for Lena 512x512 gray level image facing geometrical and SAWGN attacks. Distortion is the Mean Square Error. Embedding distortion is set to 10, that is PSNR of 38 dB.

Theoretical results developed here also allow to justify empirical strategies in many efficient scheme. As an example, Lagendijk,[13] in the context of video watermarking, proposed to perform embedding on the mean values of each frame in order to withstand to geometrical distortions (that is using only the lowest sub-band of each frame). Desynchronization models could also be used in other context where a complex transformation is performed prior to embedding (e.g. a non linear transform such as a Fourier Mellin transform). Taking into account model bias as a kind of noise such as n_i which is linked to the energy of the watermarked signal could then allow to state bounds on such system.

REFERENCES

1. P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Information Theory* , Oct. 1999.
2. M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory* **29**, pp. 439–441, May 1983.
3. P. Moulin and M. K. Mihcak, "The data-hiding capacity of image sources," *IEEE Trans. Image Processing* , 2001. preprint.
4. A. S. Cohen and A. Lapidoth, "The gaussian watermarking game," *to appear in IEEE Trans. on Information Theory* , 2002.
5. S. Pateux and G. Le Guelvouit, "Practical watermarking scheme based on wide spread spectrum and game theory," *Signal Processing : Image Communication* , pp. 283–296, Apr. 2003.
6. F. A. P. Petitcolas and R. J. Anderson, "Evaluation of copyright marking systems," in *Proc. Int. Conf. Multimedia Systems*, **1**, pp. 574–579, (Florence, Italy), Jun. 1999.
7. G. L. Guelvouit, S. Pateux, and C. Guillemot, "Information-theoretic resolution of perceptual WSS watermarking of non i.i.d. Gaussian signals," *European Signal Proc. Conference* , Sep. 2002.
8. G. Le Guelvouit, S. Pateux, and C. Guillemot, "Perceptual watermarking of non i.i.d. signals based on wide spread spectrum using side information," in *Proc. Int. Conf. on Image Processing*, **3**, pp. 477–480, (Rochester, USA), Sep. 2002.
9. G. Le Guelvouit and S. Pateux, "Wide spread spectrum watermarking with side information and interference cancellation," in *Proc. SPIE*, (Santa Clara, CA), Jan. 2003.
10. R. Bäuml, J. J. Eggers, and J. Huber, "A channel model for desynchronization attacks on watermarks," in *Proc. SPIE*, **4675**, (San Jose, CA), Jan. 2002.
11. V. Licks, F. Ourique, R. Jordán, and F. Pérez-González, "The effect of the random jitter attack on the bit error rate performance of spatial domain image watemarking," in *Proc. Int. Conf. on Image Processing*, (Barcelona, Spain), Sep. 2003.
12. G. Le Guelvouit, S. Pateux, and C. Guillemot, "Information-theoretic resolution of perceptual wss watermarking of non i.i.d gaussian signals," in *Proc. Eur. Signal Processing Conference*, **1**, pp. 454–457, (Toulouse, France), Sep. 2002.
13. Y. Zhao and R. L. Lagendijk, "Video watermarking scheme resistant to geometric attacks," in *Proc. Int. Conf. on Image Processing*, **2**, pp. 145–149, (Rochester, USA), Sep. 2002.