Performances d'un système de tatouage soumis à des désynchronisations

Stéphane PATEUX[1], Gaëtan LE GUELVOUIT[1]

[1]IRISA/INRIA
Campus universitaire de Beaulieu
35042 Rennes Cedex, FRANCE
Stephane.Pateux@irisa.fr, Gaetan.Le_Guelvouit@irisa.fr

Résumé –
Nous proposons dans cet article un nouveau modèle pour prendre en compte dans un système de tatouage des phénomènes de désynchronisations. Basé sur ce modèle et l'utilisation d'un formalisme issu de la théorie des jeux, nous en déduisons alors une borne sur les performances optimales pouvant être obtenus d'un système de tatouage soumis à de telles attaques. Les résultats obtenus illustrent l'intérêt de tatouer dans les basses fréquences du signal, et aussi l'intérêt d'utiliser des techniques de tatouage exploitant l'information adjacente.

Abstract –
In this article we propose a new model for taking into account desynchronizing phenomenum in watermarking system. Further, using game theory formalism, we state upper bounds on performance of a watermarking system subject to desynchronization attacks. Results show the needs of watermarking in the lowest sub-bands of the signal, and thus the crucial role of side informed watermarking techniques.

1 Introduction

La dissimulation d'informations dans un signal hôte (*i.e.* tatouage) est un sujet largement étudié aujourd'hui de par ses domaines d'application potentiels. On peut ainsi citer les problèmes liés à la gestion des droits, ou bien encore d'insertion de méta-données pour des fins de services à valeur ajouté. Dans toutes ces applications, se pose alors le problème de l'estimation de la capacité. I.e. combien de bits peut-on insérer et extraire au final de façon fiable même suite à des manipulations diverses (compression, filtrage, attaques intentionnelles,...)?

Pour cela, il est alors courant d'établir un lien entre le problème du tatouage et celui de communication sur un canal bruité. Les premiers travaux ont ainsi considéré l'analogie avec un canal Gaussien [7] soumis à un bruit additif (AWGN : Additive White Gaussian Noise). L'utilisation de techniques de codage avec information adjacente [2], ainsi que l'utilisation de la théorie des jeux [6, 1] a par la suite permis de définir des limites plus réalistes sur les capacités atteignables (prise en compte d'attaques optimales : SAWGN - Scaling and Additive White Gaussian Noise). Toutefois, comme souligné dans [8], ces limites basées sur l'utilisation de canaux parallèles Gaussiens ne vérifient pas un équilibre de Nash, mais peuvent être toutefois quasiment atteintes via l'utilisation d'étalement de spectre et de codage avec information adjacente.

Cependant dans tous ces travaux théoriques, il est fait l'hypothèse commune d'une connaissance parfaite des paramètres d'attaques. Or l'attaquant peut décider de biaiser son attaque empêchant ainsi l'estimation de ces paramètres. Par ailleurs, le développement des attaques géométriques aléatoires (*i.e.* désynchronisante en localisation) comme illustré par Stirmark [9] montre qu'une parfaite resynchronisation n'est pas toujours possible.

Nous proposons alors dans cet article un nouveau modèle permettant de prendre en compte ces phénomènes de désynchronisation. Par la suite, via la formulation d'un jeu entre un attaquant et un défenseur, nous exprimons quelles sont les performances atteignables par un système de tatouage dans de telles conditions.

2 Modélisation des phénomènes de désynchronisation

Dans un système de tatouage, plusieurs types de désynchronisation peuvent apparaître sur les différentes classes d'attaques. Une première classe est liée aux attaques "géométriques" (déformation géométrique pour des images, déphasage pour des signaux audio, ...). Une seconde classe est quant à elle liée aux attaques par ajout de bruit et filtrage.

Ces désynchronisations viennent du fait que lors de la phase d'extraction, il est nécessaire d'estimer les paramètres de ces attaques (estimation des déformations géométriques par recalage, estimation de la puissance de bruit ajoutée, estimation du filtrage appliqué, ...). Ces estimations sont alors réalisées en aveugle ou non, sur des contenus dégradés, et avec l'utilisation de modèles limités (par exemple déformation géométrique affine locale ou globale, ...). Il en résulte alors une imprécision sur les phénomènes effectivement présents.

Nous allons alors définir les impacts de ces désynchronisations pour des systèmes de tatouage basés sur des détections par corrélation [1].

Lorsqu'un échantillon y_i d'un signal mono-dimensionnel n'est

1. Comme présenté dans [8], cette hypothèse n'est pas restrictive puisque l'on peut atteindre les limites théoriques de performance d'un système de tatouage via l'utilisation de tests d'hypothèse basés sur des corrélations linéaires.

pas parfaitement localisé, on peut alors considérer que la réponse de ce signal lors d'une phase de corrélation sera équivalente à celle de $y'_i = c \times y_i + n_i$ où $c = sinc(\Delta)$ et n est un bruit de variance[2] $\sigma_n^2 = (1-c^2)\sigma_{Y_i}^2$. Δ représentant l'imprécision de localisation de y_i. Dans le cadre d'attaques désynchronisantes intentionnelles, on peut considérer que pour $\Delta > 1$, $c = 0$ puisqu'il suffit d'induire une désynchronisation inférieure ($\Delta = 1$) annulant complètement la réponse.

En considérant des attaques de type SAWGN, il est nécessaire d'estimer le facteur multiplicatif γ_i appliqué sur le i^{eme} site. Si l'on considère que l'on a une estimation approchée à une précision $d\gamma_i$, alors une interférence non compensable dans une schéma de tatouage exploitant l'information adjacente sera présente. Cette interférence aura une énergie de variance $(d\gamma_i)^2\sigma_{Y_i}^2$, sans pour autant impacter sur la distortion liée à l'attaque.

3 Formulation sous la forme d'un jeu

En considérant les modèles précédents, on peut alors considérer la formulation suivante ([5]) :

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

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

$$y''_i = c_i \frac{\gamma_i}{\gamma_i^W} y_i + n_i + \delta_i \quad (3)$$

où :

- x_i est le signal hôte Gaussien non i.i.d. (soit encore $X_i \sim \mathcal{N}(0, \sigma_{X_i}^2)$) ;
- y_i est le signal tatoué et y'_i le signal attaqué via une attaque de type SAWGN ;
- y''_i est le signal après désynchronisation géométrique ;
- n_i est le bruit d'auto-interférence suite à la désynchronisation géométrique (on a $N_i \sim \mathcal{N}(0, a_i\gamma_i^2(\sigma_{X_i}^2 + \sigma_{W_i}^2))$) ;
- γ_i est le facteur d'échelle appliqué au coefficient x_i. $\gamma_i^W = \frac{\sigma_{X_i}^2}{\sigma_{X_i}^2 + \sigma_{W_i}^2}$ correspond au facteur d'échelle correspondant au filtre de Wiener appliqué à l'insertion (cf. ([5] pour plus de détails) ;
- δ_i (de même que δ'_i) est un bruit additif Gaussien de variance $\sigma_{\delta_i}^2$.

Dans le même esprit que dans [8], on montre alors que l'on a à faire, à l'issue de la phase de corrélation avec les porteuses d'étalement, à un canal Gaussien de rapport signal à bruit :

$$\frac{E_b}{N_0} = \sum_{i=1}^m \frac{c_i^2\gamma_i^2\sigma_{W_i}^2}{a_i\gamma_i^2(\sigma_{X_i}^2 + \sigma_{W_i}^2) + \sigma_{\delta_i}^2} \quad (4)$$

où l'on retrouve les paramètres a_i et c_i liés à une désynchronisation géométrique. Comme vu précedement a_i peut être augmenté dans cette formule afin d'inclure les interférences liées à une imprécision sur les paramètres d'attaques SAWGN, voire en le mettant à 1, de considérer une technique de tatouage n'exploitant pas l'information adjacente.

On définit par ailleurs les distortions d'insertion :

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

(où φ_i est un facteur de pondération afin de proposer une métrique perceptuelle) et d'attaques :

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

Il est à noter que dans ce modèle, on ne prend pas en compte dans $D_{xy'}$ l'impact perceptuel des distortions géométriques (les déformations géométriques sont en effet peu perceptibles généralement). Toutefois les paramètres a_i et c_i quant à eux le prennent en compte lors de leur définition (limitation de Δ et considération des déformations induisant la plus forte diminution du rapport signal à bruit).

Par la suite, on évalue les performances théoriques atteignables d'un tel système en considèrant un jeu entre un attaquant et un défenseur, sous des contraintes de distortions d'insertion et d'attaque bornées :

$$\begin{cases} \max_{\{\sigma_{W_i}\}} \min_{\{\gamma_i, \sigma_{\delta_i}\}} \frac{E_b}{N_0} \\ D_{xy} \leq D_{xy}^{\max} \\ D_{xy'} \leq D_{xy'}^{\max} \end{cases} \quad (5)$$

4 Résolution du jeu

Afin de résoudre le jeu, nous procédons comme dans [5] via l'introduction de lagrangiens. Pour l'attaquant, on considère donc la minimisation du lagrangien suivant :

$$J_\lambda = \min_{\{\gamma_i, \sigma_{\delta_i}^2\}} \frac{E_b}{N_0} + \lambda D_{xy'} \quad (6)$$

La résolution amène alors à la définition générale suivante des paramètres d'attaque :

$$\begin{cases} \gamma_i = \frac{\sigma_{X_i}^2 - \frac{c_i\sigma_{W_i}}{\varphi_i\sqrt{\lambda}}}{(1-a_i)(\sigma_{X_i}^2 + \sigma_{X_i}^2)} \\ \sigma_{\delta_i}^2 = \gamma_i(\gamma_i^W - \gamma_i)(\sigma_{X_i}^2 + \sigma_{X_i}^2) \end{cases} \quad (7)$$

Comme dans [5], cette formulation laisse apparaitre 3 domaines d'attaque suivant la valeur des paramètres d'insertion. Un domaine d'effacement $\mathcal{D}_E$ correspondant à l'hypothèse $\sigma_{W_i} \leq \frac{\varphi_i\sqrt{\lambda}\sigma_{X_i}^2}{c_i}$, où l'attaque optimale consiste a purement et simplement annuler le signal ; un domaine de simple filtrage de Wiener $\mathcal{D}_W$, si $\gamma_i > \gamma_i^W$; et le domaine d'attaque intermédiaire $\mathcal{D}_I$, avec les paramètres d'attaque défini par l'équation 7. On peut remarquer sur la définition de ces domaines, qu'ils dépendent des facteurs (a_i, c_i) qui sont liés aux propriété de désynchronisation.

Par la suite, on peut alors chercher, connaissant la pire attaque de l'attaquant, à optimiser les paramètres d'insertion ; ce qui se traduit par la maximisation du lagrangien pour chaque différentes hypothèses d'attaque :

$$J_\chi = \max_{\{\sigma_{W_i}\}} \frac{E_b}{N_0} + \lambda D_{xy'} - \chi D_{xy} \quad (8)$$

2. Cf. modèle proposé dans [3]. Ou bien encore en considérant que le signal continu y' peut se réécrire comme $y'(t) = \sum_i y_i sinc(t-i)$. $c \times y_i$ représente alors la contribution de y_i attendu, n le bruit d'interférence lié aux autres sites voisins.

Dans le domaine $\mathcal{D}_E$, on trouve ainsi que le minimum est atteint pour un σ_{W_i} minimal ; ce qui fait donc sortir de ce domaine et n'est donc pas une solution valide. Dans le domaine $\mathcal{D}_I$, on aboutit sur la définition suivante :

$$\begin{cases} \text{si } \lambda > \chi \text{ ou } \sigma_{X_i} < \frac{c_i}{\sqrt{a_i}\varphi_i\sqrt{\chi-\lambda}}, \\ \quad \sigma_{W_i} = \dfrac{\left[\begin{array}{c} A_i \\ + \sqrt{(A_i)^2 + 4\varphi_i^2\lambda\sigma_{X_i}^2 c_i^2} \end{array}\right]}{2\varphi_i\sqrt{\lambda}c_i}, \\ \quad \text{avec } A_i = \varphi_i^2(\lambda - \chi(1-a_i))\sigma_{X_i}^2 - c_i^2 \\ \text{sinon} \\ \quad \sigma_{W_i} = 0, \end{cases} \qquad (9)$$

Dans le domaine $\mathcal{D}_W$, la solution est soit de maximiser σ_W, soit au contraire de le minimiser à 0.

En considérant ces différents cas de figure, on peut alors observer que la solution générale est définie par celle exprimée par l'équation 9. Sur cette équation, on voit apparaitre un phénomène de non marquage de certains sites ($\sigma_{X_i} < \frac{c_i}{\sqrt{a_i}\varphi_i\sqrt{\chi-\lambda}}$), comme observé dans le cadre de l'optimisation d'un schéma de tatouage par étalement de spectre [4]. Toutefois, ici le seuil dépend également du facteur c_i. On peut ainsi observer que ce seuil est particulièrement bas (voir nul) pour les sous-bandes haute fréquence du signal à tatouer. Comme on aurait pu le préssentir intuitivement, ces sous-bandes sont particulièrement sensible à toute désynchronisation, et il est donc souvent préférable de ne pas les tatouer.

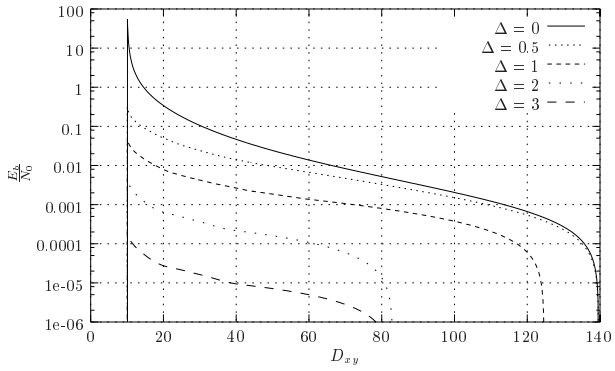
Par ailleurs, on peut remarquer, que le cas de figure $a = 1$ correspond au cas du tatouage par étalement de spectre classique (avec auto-interférence de la marque). On retrouve alors le résultat déjà obtenu dans ce cadre [4] : $\sigma_{W_i} \simeq \frac{\varphi_i\sqrt{\chi-\lambda}}{c_i}$, où le terme c_i est apparu pour s'adapter aux phénomènes de désynchronisation.

5 Résultats expérimentaux

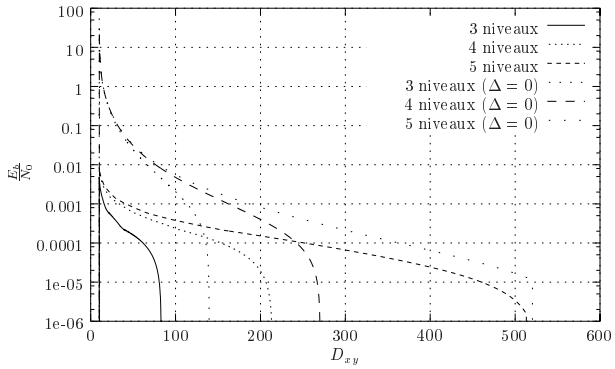
Nous présentons dans ce résumé des résultats liés à l'estimation de la capacité pour l'image Lena soumise à des attaques géométriques et SAWGN intentionnelles. La figure 1.a présente des estimations du rapport signal à bruit atteignable en considérant un tatouage suite à une décomposition ondelette sur 3 niveaux et soumise à différents niveaux d'imprécision de localisation [3]. La figure 1.b représente quant à elle l'évolution de ce rapport signal à bruit en faisant varier le nombre de niveaux de la décomposition utilisée, en considérant une imprécision de localisation de 2 pixels. Pour illustrer la perte de performance liée aux attaques géométriques, on a reporté également sur cette dernière figure les mesures en absence d'attaques géométriques. La figure 1.c présente une comparaison entre les mesures de perfomances atteignables en exploitant ou non l'information adjacente au moment de l'insertion.

Sur ces figures, on peut observer la nette perte de performance liée aux attaques géométriques. Cette perte est d'autant plus prononcée que l'on considère une insertion sur peu de niveaux d'ondelettes (cf figure 1.a, pour 3 niveaux d'ondelettes et une attaque à 35 dB, la capacité chute de plus de 56000 bits

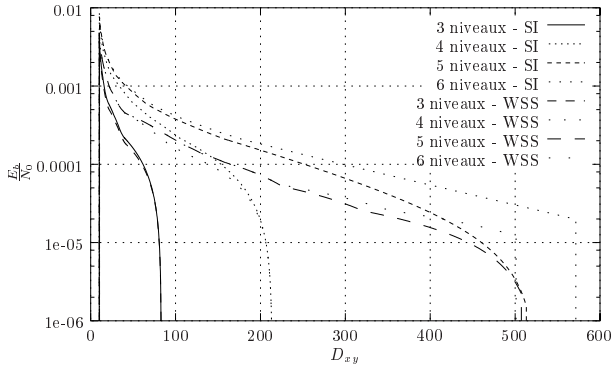
3. La capacité peut être estimée par la suite via la formule de la capacité d'un canal Gaussien, soit $C = \frac{1}{2}\log_2(1 + \frac{E_b}{N_0})$.



(a) 3 niveaux d'ondelettes



(b) $\Delta = 2\ pixels$



(c) $\Delta = 2\ pixels$ - approche avec et sans exploitation de l'information adjacente

FIG. 1: Estimation de la capacité de l'image Lena soumises à des attaques géométriques et SAWGN. Distortion de type Erreur Quadratique Moyenne. Distortion d'insertion : 10, soit un PSNR de 38 dB.

à moins de 5 bits pour $\Delta = 3$ pixels). Ces résultats montrent l'intérêt d'insérer l'information dans des bandes de fréquences relativement basses afin de résister aux attaques géométriques (l'énergie du filigrane se répartissant principalement sur ces bandes). Ainsi en utilisant une insertion sur 5 niveaux d'ondelettes, on peut espérer insérer 38 bits tout en résistant à des attaques engendrant une distortion de 26 dB et une imprécision de localisation de 2 pixels.

6 Conclusion

Dans cet article, nous présentons une modélisation du problème de tatouage soumis à des attaques intentionnelles (dégradation et désynchronisations géométriques). L'utilisation de ce modèle dans le cadre de la modélisation d'un jeu entre attaquant et défenseur permet alors d'estimer les capacités réalisables par un système de tatouage. Par ailleurs, outre l'intérêt théorique, la résolution de ce jeu amène à un schéma pratique de tatouage (comme présenté dans [8] pour le cadre d'attaques SAWGN).

En outre, les résultats obtenus permettent également de justifier les choix stratégiques, définis souvent empiriquement, dans des schémas de tatouage performants. Ainsi, dans le cadre du tatouage vidéo, l'utilisation d'un tatouage sur la valeur moyenne des images proposée dans [10] afin de résister aux attaques géométriques spatiales trouve ici une justification théorique claire (il faut aller tatouer dans les basses fréquences).

Références

[1] Cohen (A. S.) et Lapidoth (A.). – The gaussian watermarking game. *to appear in IEEE Trans. on Information Theory*, 2002.

[2] Costa (M. H. M.). – Writing on dirty paper. *IEEE Trans. on Information Theory*, vol. 29, n3, May 1983, pp. 439–441.

[3] Eggers (J. J.), Bäuml (R.) et Girod (B.). – Digital watermarking facing attacks by amplitude scaling and additive white noise. *In : 4th Int. ITG Conf. on Source and Channel Coding*.

[4] Le Guelvouit (G.), Pateux (S.) et Guillemot (C.). – Information-theoretic resolution of perceptual wss watermarking of non i.i.d gaussian signals. *In : Proc. Eur. Signal Processing Conference*, pp. 454–457. – Toulouse, France, Sep. 2002.

[5] Le Guelvouit (G.), Pateux (S.) et Guillemot (C.). – Perceptual watermarking of non i.i.d. signals based on wide spread spectrum using side information. *In : Proc. Int. Conf. on Image Processing*, pp. 477–480. – Rochester, USA, Sep. 2002.

[6] Moulin (P.) et Mihcak (M. K.). – The data-hiding capacity of image sources. *IEEE Trans. Image Processing*, May. 2000. – submitted.

[7] Moulin (P.) et O'Sullivan (J. A.). – Information-theoretic analysis of information hiding. *IEEE Trans. Information Theory*, Oct. 1999.

[8] Pateux (S.) et Le Guelvouit (G.). – Practical watermarking scheme based on wide spread spectrum and game theory. *To appear in IEEE Trans. on Image Communication*, 2003.

[9] Petitcolas (F. A. P.) et Anderson (R. J.). – Evaluation of copyright marking systems. *In : Proc. Int. Conf. Multimedia Systems*, pp. 574–579. – Florence, Italy, Jun. 1999.

[10] Zhao (Y.) et Lagendijk (R. L.). – Video watermarking scheme resistant to geometric attacks. *In : Proc. Int. Conf. on Image Processing*, pp. 145–149. – Rochester, USA, Sep. 2002.