Analyse stéganographique du schéma scalaire de Costa

Gaëtan Le Guelvouit[2], Ahmedou Ould Bouya[1], Joseph Bourgeois[1], Claude Delpha[1], Rémy Boyer[1]

[1]Laboratoire des Signaux et Systèmes (CNRS, Supélec, Université Paris Sud 11)
3, Rue Joliot Curie — 91192 Gif sur Yvette

[2]France Telecom R&D
4, Rue du Clos Courtel — 35512 Cesson Sévigné cedex
Ahmedou.Ouldbouya@lss.supelec.fr, Claude.Delpha@lss.supelec.fr, Remy.Boyer@lss.supelec.fr
gaetan.leguelvouit@orange-ftgroup.com

Résumé – Nous proposons ici une analyse stéganographique du Schéma Scalaire de Costa. D'un point de vue théorique nous montrons ses limites puis nous proposons d'améliorer ce schéma en utilisant la quantification codée en treillis (TCQ). Nous comparons les résultats obtenus d'un point de vue théorique et pratique pour le marquage d'images. Par l'intermédiaire de simulations Monte-Carlo nous prouvons l'efficacité de la solution proposée.

Abstract – We provide in this paper a steganographical analysis of the Scalar Costa Scheme. We then theoretically show the limits of this scheme and we propose to use treillis coded quantization to improve the steganographic quality of the scheme. We compare the theoretical results to the practical ones done for images. Futhermore, Monte-Carlo simulations demonstrate the efficiency of the solution.

Introduction

Dans le classique problème des prisonniers [1], Alice et Bob sont en prison et cherchent à monter une évasion. Ils peuvent se transmettre des documents, mais ceux-ci sont surveillés par un gardien. Si ce dernier trouve que les documents sont suspects, il décidera de couper les échanges entres les deux prisonniers. Alice et Bob doivent donc utiliser la stéganographie pour communiquer secrètement leur plan. Il existe de nombreux schémas de stéganographie s'appuyant sur la modification subtile de documents multimédia [2]. Ils ont en commun le fait de nécessiter une clef secrète, partagée entre Alice et Bob. Même si le gardien trouve le document hôte suspect, il ne peux savoir ce qu'il contient. Mais cette clef doit être échangée entre Alice et Bob. Comment faire s'ils ne peuvent se rencontrer physiquement ? La stéganographie asymétrique cherche à résoudre ce problème.

Une clef publique $\mathbf{k}_{\text{pub}}$ peut être distribuée à tous (même le gardien peut être au courant), et une clef privée $\mathbf{k}_{\text{prv}}$ reste secrète pour le destinataire des messages. Le schéma scalaire de Costa (SCS) [3], habituellement exploité dans le cadre du tatouage robuste, présente des failles de sécurité. Un schéma de stéganographie à clef publique basé sur le SCS a été proposé dans [4]. Cependant, les auteurs indiquent – de manière expérimentale – des contraintes sévères d'utilisation portant sur la puissance d'insertion du tatouage et sur la densité de probabilité du signal d'entrée. Dans ce travail, nous justifions d'un point de vue théorique les résultats obtenus dans [4] et nous proposons une solution garantissant sans contrainte les conditions d'invisibilité statistique nécessaires en stéganographie.

1 Stéganographie à clef publique basée sur le SCS

Cette section fait un rappel sur les travaux précédents cherchant à définir une technique de stéganographie asymétrique de signaux multimédia.

1.1 Rappels sur le SCS

Le schéma scalaire de Costa (SCS) est une technique de tatouage basée sur la quantification scalaire des données hôtes. C'est une mise en pratique simple des principes de Costa pour les canaux avec information adjacente [5]. Considérons un message binaire $\mathbf{m}$ et un signal hôte $\mathbf{x}$. Par quantification, le SCS définit deux sous-dictionnaires par échantillon $\mathbf{x}[i]$ à marquer :

$$\begin{aligned} \mathcal{U}_0[i] &= \{k\Delta + \mathbf{d}[i],\, k \in \mathbb{Z}\} \\ \text{et } \mathcal{U}_1[i] &= \left\{k\Delta + \mathbf{d}[i] + \frac{\Delta}{2},\, k \in \mathbb{Z}\right\}. \end{aligned} \tag{1}$$

Le paramètre Δ est le pas de quantification (utilisé pour régler le compromis entre robustesse et distorsion) et $\mathbf{d}[i] \in [-\Delta/2; +\Delta/2]$ est un bruit de *dithering* formant une clef secrète[1]. En fonction du bit $\mathbf{m}[i]$ à insérer, l'un des deux sous-dictionnaires est retenu. Le mot de code $\mathbf{u}^\star[i]$ le plus proche de $\mathbf{x}[i]$ est choisi. L'échantillon marqué est alors donné par

$$\mathbf{y}[i] = \mathbf{x}[i] + \alpha\left(\mathbf{u}^\star[i] - \mathbf{x}[i]\right),$$

[1]Sans ce paramètre, impossible de reconstruire les dictionnaires, et donc de transmettre ou de recevoir des informations.

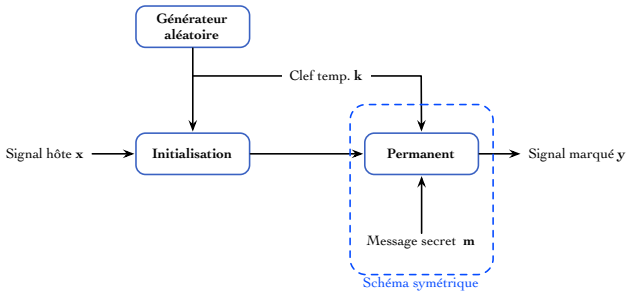


FIG. 1 – Schéma de stéganographie asymétrique. La phase permanente est initialisée avec une clef privée temporaire **k**

où $\alpha \in [0;1]$ est le paramètre de Costa. Au décodage, il est facile de retrouver quel bit a été inséré en vérifiant si $\mathbf{y}[i]$ est plus proche d'un mot de code de $\mathcal{U}_0[i]$ ou de $\mathcal{U}_1[i]$.

1.2 Application à la stéganographie à clef publique

À partir d'idées exposées dans un article d'Anderson et Petitcolas [6], Guillon *et al.* [4] ont proposé une mise en œuvre pratique de la stéganographie à clef publique, en s'appuyant sur la cryptographie asymétrique et sur le SCS. La figure 1 schématise les deux phases de la techniques.

La **phase d'initialisation** consiste à transmettre une clef privée pour la seconde phase. Une clef privée temporaire **k** est générée au hasard. Elle est chiffrée en utilisant un algorithme de chiffrement asymétrique. Le résultat $\mathbf{i} = C(\mathbf{k}, \mathbf{k}_{\text{pub}})$ est inséré dans le signal hôte. La **phase permanente** utilise la clef $\mathbf{k}_{\text{tmp}}$ pour transmettre le message **m**. Elle utilise le SCS pour l'insertion du message dans le signal hôte, avec un bruit de *dithering* secret **d**, généré à partir de la clef **k**.

La phase permanente ne présente pas de difficultés, le SCS avec *dithering* offre de bonnes propriétés de sécurité. Mais la phase d'initialisation nécessite la transmission d'informations publiques sans changement sensible sur le signal hôte (changements statistiques, dégradation de qualité, etc.). À cette fin, Guillon propose d'utiliser le SCS avec un paramètre $\alpha = 0{,}5$. Mais cela n'est valable que pour un signal hôte suivant une loi uniforme. Dans les autres cas, l'insertion d'informations laisse des traces statistiques. Dans leur article, Guillon *et al.* font cette remarque en étudiant expérimentalement les d.d.p. des signaux marqués. La section suivante propose une justification théorique de ces artefacts.

1.3 Densité de probabilité du signal marqué avec le SCS

On note Y la variable aléatoire modélisant $\mathbf{y}$, et $u(k,m) = (k + m/2)\Delta$ le $k^{\text{ème}}$ mot de code du quantificateur $\mathcal{U}_m$ correspondant au bit m de la marque à transmettre. Sous hypothèse d'équiprobabilité des bits de tatouage, il vient :

$$P_Y(y) = \frac{1}{2(1-\alpha)} \sum_{k,m} L(u,k,m,\alpha,\Delta) \times P_X\left(\frac{y - \alpha u(k,m)}{1-\alpha}\right), \quad (2)$$

avec

$$L(u,k,m,\alpha,\Delta) = 1_{\left[u(k,m) - \frac{(1-\alpha)\Delta}{2}, u(k,m) + \frac{(1-\alpha)\Delta}{2}\right]}$$

où L est une fonction indicatrice. Pour un k fixé, les deux quantificateurs sont éloignés de $\Delta/2$, et donc les fonctions indicatrices se recouvrent si $(1-\alpha)\Delta/2 > \Delta/4$ (qui est équivalent à $\alpha < 1/2$) ; et dans le cas où $\alpha > 1/2$ les fonctions indicatrices sont disjointes. Cela explique la présence des recouvrements ou des trous dans les statistiques du signal marqué. Pour la valeur $\alpha = 1/2$, il n'y a pas de trous ni de recouvrements mais la continuité aux points de raccordement est assurée si et seulement si $P_Y(u(k,m)/2) = P_Y(u(k,m)/2 + \Delta/4)$. Cette égalité est satisfaite dans le cas où la densité de probabilité est uniforme. Inversement, pour une densité de probabilité gaussienne, il apparaît des discontinuités aux points de raccordement (voir les figure 2(a) à 2(c)). Pour respecter cette contrainte, les auteurs de [4] exploitent un compresseur en amont du tatouage.

Dans le cadre d'une approche stéganographique où l'on souhaite garantir une invisibilité statistique de l'information cachée, les discontinuités observées implique un niveau de sécurité médiocre. Dans la partie suivante, on propose une approche par quantification codée par treillis afin d'augmenter l'invisibilité statistique de notre système.

2 Quantification codée par treillis (TCQ)

Comme le SCS sans *dithering* fait un partionnement régulier du signal hôte, il introduit des artefacts. L'approche que nous proposons est l'utilisation d'un treillis pour construire un partionnement pseudo-aléatoire.

2.1 Principes

Nous considérons un treillis défini par la fonction de transition : $\mathcal{S} \times \{0,1\} \longrightarrow \mathcal{S}$, $t : (s_i, \mathbf{m}[i]) \longmapsto s_{i+1}$, avec $\mathcal{S} = \{0, 1, \ldots, 2^{r-1}\}$ ensemble des états possibles. Contrairement au SCS, le bruit de *dithering* **d** n'est plus aléatoire, mais devient une fonction de l'état courant et du symbole d'entrée : $\mathcal{S} \times \{0,1\} \longrightarrow [-\Delta/2; +\Delta/2]$, $f : (s_i, \mathbf{m}[i]) \longmapsto \mathbf{d}[i]$. Les sous-dictionnaires sont définis par

$$\mathcal{U}_{\mathbf{m}}[i] = \{k\Delta + f(s_i, \mathbf{m}[i]),\ k \in \mathbb{Z}\}$$

et le mot de code le plus proche $\mathbf{u}^\star \in \mathcal{U}_{\mathbf{m}}$ de **s** est calculé par un algorithme de Viterbi avec un *a priori* fort, afin de s'assurer que le mot de code trouvé appartient bien à $\mathcal{U}_{\mathbf{m}}$:

$$\mathbf{u}^\star = \arg \min_{\mathbf{u} \in \mathcal{U}_{\mathbf{m}}} \sum_{i=1}^{n} (\mathbf{s}[i] - \mathbf{u}[i])^2 . \quad (3)$$

Le signal marqué est $\mathbf{x} = \alpha\,(\mathbf{u}^\star - \mathbf{s})$. Les expériences montrent que le meilleur paramètre α est $P/(P+N)$ comme dans le schéma original de Costa. Comme on peut le constater sur les figures 2(d) à 2(f), aucun artefact statistique n'est visible en utilisant cette technique.

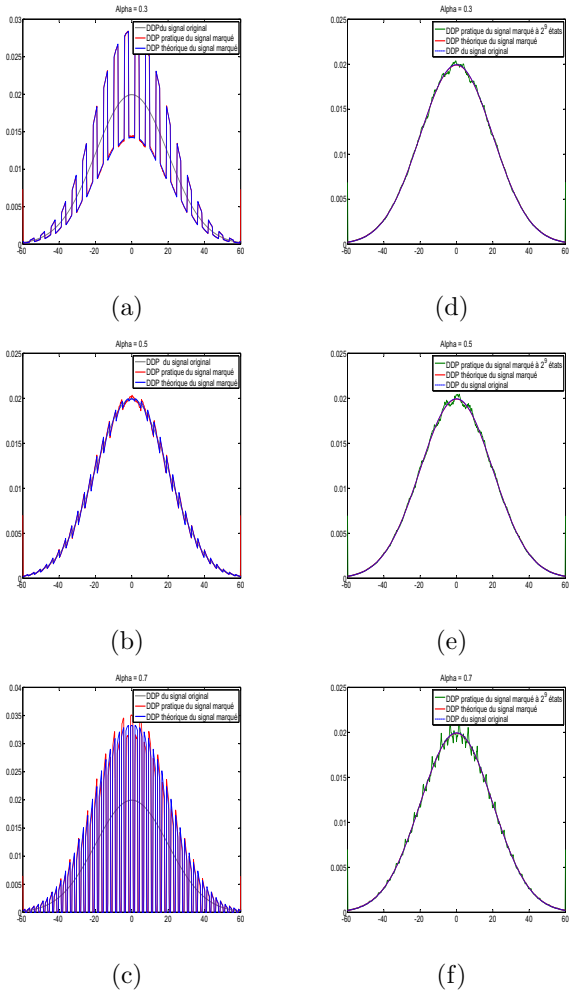


FIG. 2 – Densités de probabilités (d.d.p.) du signal hôte et du signal marqué par le SCS (à gauche) et la TCQ (à droite)

Néanmoins, la technique d'insertion linéaire (c'est-à-dire $\mathbf{w} = \alpha(\mathbf{u}^\star - \mathbf{x})$) laisse un indice pour le gardien en charge de la surveillance des échanges entre Alice et Bob. Posons $d_{\mathbf{x}}$ comme la distance entre un signal hôte non marqué $\mathbf{x}$ et le mot de code le plus proche $\mathbf{u} \in \mathcal{U}$. Nous avons $\mathbb{E}\left[d_{\mathbf{x}}\right] = \Delta^2/48$. Soit $d_{\mathbf{y}}$ la distance entre un signal marqué $\mathbf{y}$ et le mot de code le plus proche. Pour paraître innocent, nous devons avoir $d_{\mathbf{y}} \simeq d_{\mathbf{x}}$, c'est-à-dire utiliser $\alpha = 1/2$. Mais dans la plupart des cas, cette valeur ne permet pas d'atteindre la région de Voronoi associée à $\mathbf{u}^\star$. Il faut donc utiliser un α plus important, ce qui donne $d_{\mathbf{y}} < d_{\mathbf{x}}$ et rend le signal marqué suspect. Nous proposons alors une technique d'insertion itérative pour placer le signal marqué à la frontière de plusieurs mots de code, tout en veillant à rester dans la bonne région.

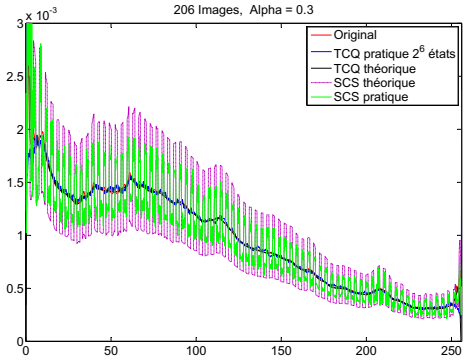


FIG. 3 – Analyse théorique et pratique des deux schémas sur images réelles

2.2 Analyse statistique de la quantification codée par treillis

Afin de justifier de manière théorique l'exploitation de la TCQ du point de vue invisibilité statistique, nous avons déterminé la densité de probabilité du signal marqué selon l'approche déjà vue dans la section 1.3. Nous trouvons

$$P_Y(y) = \frac{1}{\sigma_W\sqrt{12}} \int_{y-\sigma_W\sqrt{3}}^{y+\sigma_W\sqrt{3}} P_X\left(\frac{y - t\alpha\Delta}{1-\alpha}\right) \mathrm{d}t, \qquad (4)$$

où σ_W est l'écart type de la marque ajoutée. Et donc $P_Y(y)$ est la moyenne de la d.d.p. du signal original sur un intervalle centré en y et de largeur $\sigma_W\sqrt{3}$. Du fait du manque de place, la démonstration sera produite dans une communication longue. Après implémentation de l'équation 4 sur un signal de densité de probabilités gaussienne, nous obtenons les résultats présentés par les figures 2(d) à 2(f). On peut remarquer la bonne adéquation de la densité de probabilité obtenue par l'algorithme de TCQ, sa version théorique (voir l'équation 4) et l'originale, ceci même pour des puissances d'insertion élevées (à comparer avec les figures 2(a) à 2(c)). Les avantages de l'approche par TCQ est qu'elle permet d'assurer une invisibilité statistique quelque soit la puissance d'insertion et la densité de probabilité du signal hôte. Les contraintes du schéma proposé dans [4] sont donc levées.

2.3 Application aux images réelles

Nous avons appliqué ces deux schémas (SCS et TCQ) aux images. Pour cette étude, plus de 200 images ont été utilisées. Afin de s'assurer une imperceptibilité visuelle de la marque insérée, nous avons choisi d'utiliser un DWR de 35 dB. Comme le montre la figure 3, dans le cadre des images nous obtenons des d.d.p. similaires entre les courbes théoriques et pratiques. Le modèle théorique proposé peut donc être validé. Nous nous sommes ensuite intéressés à l'étude du niveau d'invisibilité de ces deux schémas.

En s'appuyant sur le calcul de la distance de Kullback-Liebler (D_{KL}), nous avons pu analyser le niveau d'invisibilité de ces schémas en fonction du paramètre α. Nous nous

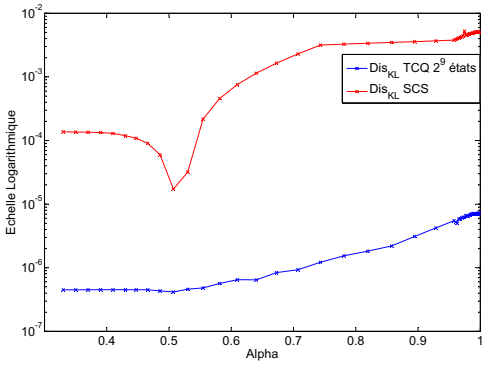


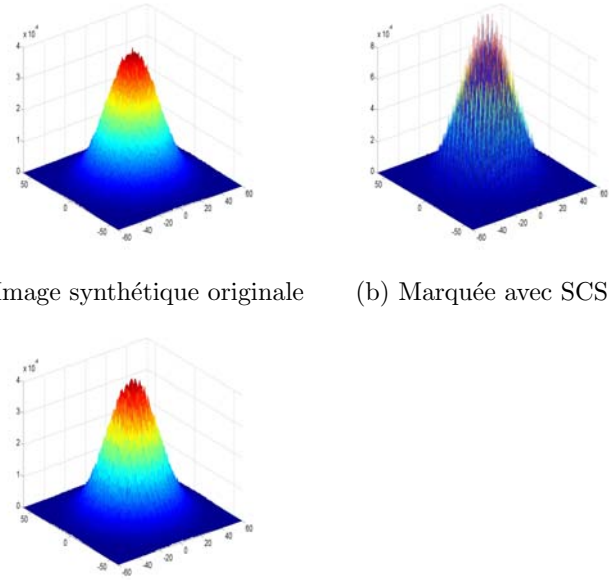
FIG. 4 – Invisibilité à l'ordre 1, distance de Kullback-Leibler D_{KL} en fonction de α

sommes d'abords intéressés aux statistiques d'ordre 1 (figure 4). En fonction de α, nous vérifions dans le cadre du SCS que la valeur permettant d'obtenir la meilleure invisibilité est pour $(\alpha = 0,5)$. En ce qui concerne la TCQ, nous observons une distance D_{KL} significativement plus petite que celle obtenue dans le cas du SCS quelle que soit la valeur de α. Nous vérifions donc une meilleure invisibilité (donc qualité stéganographique) de ce schéma.

Afin d'évaluer l'impact d'une corrélation inter-pixel, nous avons représenté sur la figure 5 les statistiques conjointes pour deux pixels pour le SCS et la TCQ. La valeur de α est fixée à $\alpha = 0,7$. La distance D_{KL} pour le SCS est $D_{\text{KL-SCS}} = 3,44 \times 10^{-5}$ alors qu'elle est $D_{\text{KL-TCQ}} = 6,13 \times 10^{-7}$ pour la TCQ. Comme cela fut observé à lors de l'étude des statistiques d'ordre 1, ces résultats confirment les résultats précédents : le schéma basé sur la TCQ à une meilleure qualité stéganographique que le SCS.

Conclusion et perspectives

Dans ce travail, nous avons montré théoriquement et de manière pratique les limites du SCS du point de vue de la stéganographie asymétrique. Intuitivement, le SCS sans *dithering* secret réalise un partitionnement régulier du signal hôte, il introduit alors des artefacts dans la densité de probabilité du signal marqué. Partant de ce constat, nous avons proposé une solution basée sur la quantification codée par treillis. Cette technique effectue un partitionnement pseudo-aléatoire, ce qui permet d'avoir un schéma publique plus général (la technique n'est pas dépendante de la distribution du signal à marquer) et plus sécurisé. En appliquant ce résultat aux images, nous avons pu vérifier la pertinence du modèle théorique proposé. En s'intéressant à l'analyse statique de ce schéma, nous avons pu confirmer à l'ordre 1 et 2 les qualités stéganographiques du schéma basé sur la TCQ par rapport au SCS. L'application aux images a permis de confirmer la faisabilité de mise en œuvre de l'approche proposée dans un cas pratique.



(a) Image synthétique originale (b) Marquée avec SCS

(c) Marquée avec TCQ

FIG. 5 – Invisibilité à l'ordre 2

Références

[1] G. J. Simmons. The prisonners' problem and the subliminal channel, in *Advances in Cryptology : Proc. of CRYPTO*, pp. 51–67, 1984.

[2] S. Kazenbeisser et F. A. P. Petitcolas. Information hiding techniques for steganograpy and digital watermarking, Artech House, Dec. 1999.

[3] J. J. Eggers, R. Baüml, R. Tzchoppe et B. Girod. Scalar Costa scheme for information embedding, IEEE Trans. on Signal Processing, Avr. 2003.

[4] P. Guillon, T. Furon et P. Duhamel. Applied public-key steganography, in *Proc. SPIE Electronic Imaging*, San Jose, CA, 2002.

[5] M. H. M. Costa. Writing on dirty paper, IEEE. Trans. on Information Theory, 29(3) : 439–441, Mai 1983.

[6] R. J. Anderson et F. A. P. Petitcolas. On the limits of steganography, IEEE Journal of Selected Areas in Communication, vol. 16, no. 4, pp. 474–481, 1998.