Applications de la théorie de la décision statistique à l'évaluation de la sécurité de stégosystèmes

Rémi Cogranne, Cathel Zitzmann, Florent Retraint, Lionel Fillatre, Igor V. Nikiforov

Applications de la théorie de la décision statistique à l'évaluation de la sécurité de stégosystèmes

Rémi Cogranne, Cathel Zitzmann, Florent Retraint, Lionel Fillatre et Igor Nikiforov ⋆

ICD - LM2S - Université de technologie de Troyes (UTT) - UMR STMR - CNRS.
12, rue Marie Curie, BP 2060, 10010 Troyes.

Résumé Dans un contexte opérationnel de dissimulation d'informations, il est utile pour un stéganalyste de disposer d'un borne sur les performances que l'on est en droit attendre d'un détecteur. D'un autre côté, il est aussi utile pour les stéganographes de disposer d'une borne sur la longueur du message garantissant une communication sécurisée.
Le but du présent papier est de montrer l'intérêt de la théorie de la décision statistique pour la stéganographie et la stéganalyse. L'approche de Neyman-Pearson est utilisée concevoir le test le plus puissant pour la détection d'informations cachées par correspondance de bit de poids faibles ("*LSB Matching*" ou LSB ± 1). Les performances statistiques de ce test sont analytiquement établies et sont ensuite utilisées pour proposer une méthodologie de mesure de la capacité d'insertion d'un schéma de stéganographie. Enfin, un modèle précis des images naturelles est exploité pour mettre en pratique ce test dont la pertinence est mis en évidence sur la base BOSS [2] constituée de plus de 9000 images.

Keywords: Test statistique d'hypothèses, détection optimale, bornes d'optimalité, mesure de sécurité, modèle local d'image.

1 Introduction

La stéganographie et la stéganalyse sont un jeux du chat et de la souris : d'un côté le stéganographe souhaite cacher des informations au sein d'un média numérique anodin en apparence et, d'un autre côté, le stéganalyste analyse les médias échanger pour tenter de détecter la présence d'informations cachées.
Compte-tenu du nombre important d'outils aisément utilisable, la stéganographie est à la portée de tous. Il est donc crucial pour les forces de sécurité de disposer d'un outil de stéganalyse le plus fiable possible.La principale difficulté est alors de proposer un test dont les probabilités d'erreurs sont analytiquement calculables afin de garantir une probabilité d'erreur prescrite et de comparer la puissance de détection d'un test avec une borne d'optimalité. Du côté du

⋆. Ce travail a été financé par l'Agence Nationale pour la Recherche (ANR) au travers du programme ANR-CSOSG (Projet ANR-07-SECU-004)

stéganographe, la principale difficulté est pouvoir mesurer le nombre d'informations dissimulable dans un média tout en préservant un niveau de sécurité donné. De nombreuses méthodes de détection d'informations cachées ont été proposées dans la littérature, voir notamment [1, 8]. Mais, à l'exception des méthodes proposées dans [5–7], les résultants obtenus ne font généralement pas l'objet d'une étude statistique ; il n'est donc possible d'évaluer leurs performances qu'au travers de simulations numériques. En outre, le seul résultat théorique sur la capacité d'insertion d'un média demeure la loi de la "racine carrée de la capacité stéganographique" [10] qui est difficilement exploitable en pratique.
Le présent article prolonge la méthodologie proposée dans [6, 13] au schéma d'insertion par correspondance de LSB (parfois appelé LSB ± 1 ou "LSB Matching"). Le test le plus puissant est donné et ses performances sont analytiquement établies. La puissance de ce test est utilisée pour mesurer la capacité d'insertion dans média numérique sous contrainte d'un niveau de risque acceptable. Enfin un modèle adapté aux images naturelles est utilisé pour mettre en pratique le test proposé.
Les principales contributions de ce papiers sont : 1) une borne d'optimalité majorant la puissance attendue d'un test , 2) une mesure de la capacité d'insertion d'un média associée à niveau de sécurité, et 3) la conception d'un test pratique évalué sur les 9000 images de la base BOSS [2].

2 Formalisation du problème de la stéganagalyse

Supposons que le vecteur $\mathbf{C} = (c_1, \ldots, c_N)^T$ représente le média de couverture constitué de N échantillons codés de b bits. L'ensemble des valeurs quantifiées est noté $Z = \{0; \ldots; 2^b - 1\}$. La valeur c_n du n-ième échantillon résulte de la quantification :

$$c_n = Q_\Delta(\theta_n + \xi_n)\,,$$

où la valeur déterministe $\theta_n \in \mathbb{R}$ représente l'espérance mathématique de c_n (*i.e.* le contenu du signal), $\xi_n \sim \mathcal{N}(0, \sigma_n)$ est la réalisation d'une variable aléatoire représentant l'ensemble des bruits d'acquisition et $Q_\Delta(\cdot)$ représente le quantificateur uniforme de pas Δ défini par, en négligeant les phénomènes de saturation, par :

$$\forall k \in \mathbb{Z}\,,\, Q_\Delta(x) = k \Leftrightarrow x \in [\Delta(k - 1/2)\,;\, \Delta(k + 1/2)[\,.$$

La loi de distribution de c_n, notée $P_{\theta_n} = \{p_{\theta_n}[k]\}_{k \in \mathcal{Z}}$, est donc donnée par :

$$\forall k \in \mathbb{Z}\,,\, p_{\theta_n}[k] = \frac{1}{\sigma_n} \int_{\Delta\left(k-\frac{1}{2}\right)}^{\Delta\left(k+\frac{1}{2}\right)} \phi\left(\frac{\Delta x - \theta_n}{\sigma_n}\right) dx = \frac{\Delta}{\sigma_n}\phi\left(\frac{x - \theta_n}{\sigma_n}\right) + o(\frac{\Delta^2}{\sigma_n^2}) \quad (1)$$

où $\phi(u) = \frac{1}{\sqrt{2\pi}} \exp[\frac{-u^2}{2}]$ est la densité de probabilité de la loi normale $\mathcal{N}(0, 1)$.

Afin de modéliser statistiquement l'impact de l'insertion d'informations binaires dans les LSB, les deux assertions suivantes, usuellement admises [6, 7], sont utilisées : 1) les bits du message, 0 ou 1, sont équiprobables et, 2) la probabilité d'insertion dans chacun des échantillons c_n est identique.
Le stéganographe souhaite insérer un message constitué de L bits, *i.e.* avec un

taux d'insertion $R = {}^{L}/_{N}$. Après insertion du message par correspondance de LSB, la loi de probabilité du média $\mathbf{S}$ est donnée par $Q^{\pm}_{\theta_n;R} = \left\{q^{\pm}_{\theta_n;R}[k]\right\}_{k\in\mathcal{Z}}$ avec [9, 12] :

$$q^{\pm}_{\theta_n;R}[k] = \left(1 - \frac{R}{2}\right) p_{\theta_n}[k] + \frac{R}{4}\left(p_{\theta_n}[k+1] + p_{\theta_n}[k-1]\right) . \tag{2}$$

Lors de l'analyse d'un média $\mathbf{Z}$ de nature inconnue, le but de la stéganalyse est de décider entre les deux hypothèses simples suivantes :

$$\mathcal{H}_0 = \{z_n \sim P_{\theta_n}, \forall n = 1\ldots, N\} \text{ vs } \mathcal{H}_1 = \{z_n \sim Q^{\pm}_{\theta_n;R}, \forall n = 1\ldots, N\} . \tag{3}$$

Dans un contexte opérationnel, le but est de trouver un test statistique $\delta : \mathbb{Z}^N \mapsto \{\mathcal{H}_0; \mathcal{H}_1\}$ appartenant à la classe $\mathcal{K}_{\alpha_0}$ des tests dont la probabilité de fausse-alarme est majorée par α_0, voir [11] pour plus d'informations :

$$\mathcal{K}_{\alpha_0} = \{\delta : \mathbb{P}_{\mathcal{H}_0}[\delta(\mathbf{Z}) = \mathcal{H}_1] \leq \alpha_0\}$$

La fonction de puissance $\beta_\delta(\mathbf{Z})$, est la probabilité de détection du message caché $\beta_\delta(\mathbf{Z}) = \mathbb{P}_{\mathcal{H}_1}(\delta(\mathbf{Z}) = \mathcal{H}_1)$. Le but du stéganographe est naturellement de maximiser la puissance $\beta_\delta(\mathbf{Z})$ et cela, si possible, uniformément par rapport à R.

3 Test le plus puissant pour la stéganalyse

Lorsque le taux d'insertion R ainsi que, pour $n \in \{1, \ldots, N\}$, les paramètres θ_n et σ_n sont connus du stéganalyste, le lemme de Neyman-Pearson, voir [11, théorème 3.2.1], nous indique comment construire le test le plus puissant (PP) de la classe $\mathcal{K}_{\alpha_0}$. Pour résoudre le problème (3), le test PP est donné par :

$$\delta^{\pm}_R = \begin{cases} \mathcal{H}_0 \;\text{ si }\; \displaystyle\sum_{i=1}^{N} \ln \Lambda^{\pm}_R(z_n) = \sum_{n=1}^{N} \ln\left(\frac{q^{\pm}_{\theta_n;R}[z_n]}{p_{\theta_n}[z_n]}\right) < \tau_{\alpha_0} \\ \mathcal{H}_1 \;\text{ si }\; \displaystyle\sum_{i=1}^{N} \ln \Lambda^{\pm}_R(z_n) = \sum_{n=1}^{N} \ln\left(\frac{q^{\pm}_{\theta_n;R}[z_n]}{p_{\theta_n}[z_n]}\right) \geq \tau_{\alpha_0} , \end{cases} \tag{4}$$

$$\text{où } \ln \Lambda^{\pm}_R(z_n) = \log\left(\left[\exp\left(\frac{k-\theta_n}{\sigma_n^2}\right) + \exp\left(\frac{\theta_n - k}{\sigma_n^2}\right)\right]\right) - \frac{1}{2\sigma_n^2} - \log(4),$$

et τ_{α_0} est la solution de l'équation $\mathbb{P}_{\mathcal{H}_0}\left[\Lambda_R(\mathbf{Z}) \geq \tau_{\alpha_0}\right] = \alpha_0$.
Caractériser les performances du test $\delta^{\pm}_R$ (4) n'est pas simple. Pour résoudre ce problème, il est proposé d'utiliser le théorème de la limite centrale de Lindeberg [11, théorème 11.2.5] dans le cadre d'une approche asymptotique. Quelques calculs permettent alors de montrer que pour tout $R \in]0; 1]$ le seul de décision τ_{α_0} est asymptotiquement donnée, lorsque $N \to \infty$ par :

$$\tau_{\alpha_0} = \sigma^{\pm}_0 \Phi^{-1}(1 - \alpha_0) + \mu^{\pm}_0 \sqrt{N} , \tag{5}$$

avec $\mu^{\pm}_0$ et $\sigma^{\pm}_0$ (respectivement $\mu^{\pm}_R$ et $\sigma^{\pm}_R$) les deux premiers moments du rapport de vraisemblance $\ln \Lambda^{\pm}_R(z_n)$ sous l'hypothèse nulle $\mathcal{H}_0$ (respectivement sous

l'hypothèse alternative $\mathcal{H}_1$ avec $R \in]0;1]$).
De la même façon, la puissance du test $\delta_R^{\pm} \in \mathcal{K}_{\alpha_0}$ est asymptotiquement :

$$\beta_{\delta_R^{\pm}} = 1 - \Phi\left(\frac{\sigma_0^{\pm}\,\Phi^{-1}(1-\alpha_0) + (\mu_0^{\pm} - \mu_R^{\pm})\sqrt{N}}{\sigma_R^{\pm}}\right). \tag{6}$$

L'expression de la puissance du test PP $\delta_R^{\pm}$ (6) constitue un résultat important. Cela permet dans un cadre "idéalisé" (puisque les paramètres θ_n, σ_n et R sont supposés connus) de calculer la puissance que l'on est en droit d'attendre d'un test pour la détection d'informations cachées. En outre, la fonction de puissance $\beta_{\delta_R^{\pm}}$ (6) et la valeur du seuil de décision τ_{α_0} (5) sont en accord avec la loi de la "racine carrée de la capacité stéganographique" [10].
Enfin, notons qu'une comparaison entre le rapport de vraisemblance $\Lambda_R^{\pm}$ (4) et celui donné dans [6,13] pour le cas de la substitution de LSB permet de comprendre la difficulté supplémentaire posé par le présent problème.

4 Application pour la sécurité d'un stégosystème

Du point de vue du stéganographe, la fonction de puissance du test PP $\delta_R^{\pm}$ (6) permet de mesurer la capacité d'insertion sous contrainte de respect d'un niveau de sécurité. La définition 1, inspirée de [10], formalise cette notion de capacité.

Definition 1 (capacité d'insertion d'un média face à un adversaire). *Pour tout $(\alpha_0^\star, \beta^\star)$ avec $0 < \alpha_0^\star < \beta^\star < 1$, la capacité d'insertion d'un média* $\mathbf{Z}$ *face à un adversaire disposant du test δ est définie comme le nombre $L_{\max}$ maximal de bits d'informations dissimulable dans* $\mathbf{Z}$ *tout en assurant la sécurité du stéganographe au sens [10] où le test δ vérifie $\alpha_\delta \geq \alpha_0^\star$ ou bien $\beta_\delta \leq \beta^\star$.*

L'expression de la fonction de puissance $\beta_{\delta_R}^{\pm}$ (6) associée au test le puissant $\delta_R^{\pm}$, permet d'établir, en utilisant le fait que $\sigma_0^{\pm} \leq \sigma_R^{\pm} \leq \sigma_1^{\pm}$, la proposition 1.

Proposition 1. *Pour tout $(\alpha_0^\star, \beta^\star)$, $0 < \alpha_0^\star < \beta^\star < 1$, i.e. quel que soit le risque acceptable pour le stéganographe, le réel $L^\star$ majoré par :*

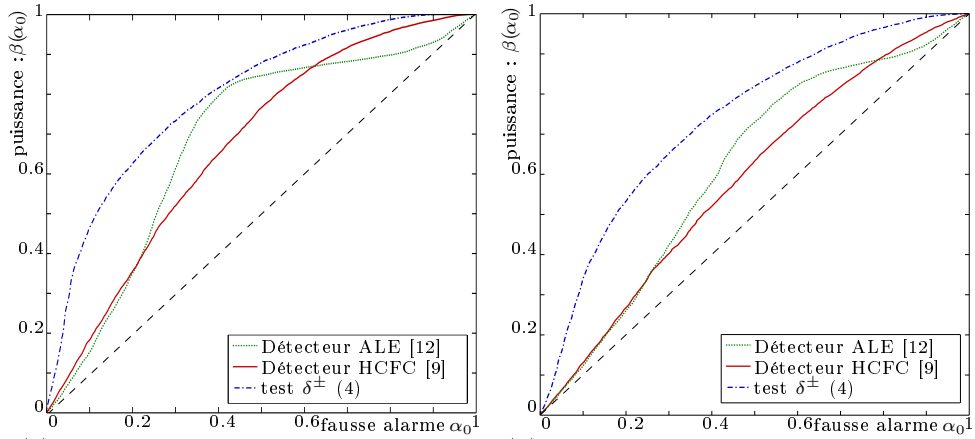
$$L^\star \geq L^{\pm} = \frac{\sigma_1^{\pm}\sqrt{N}}{(\mu_1^{\pm} - \mu_0^{\pm})}\left(\Phi^{-1}(1-\alpha_0^\star) - \Phi^{-1}(1-\beta^\star)\right) \tag{7}$$

assure que la puissance du test $\delta_R^{\pm}$, PP dans la classe $\mathcal{K}_{\alpha_0^\star}$, vérifie $\beta_{\delta_R}^{\pm} = \beta^\star$.

La proposition 1 fournie donc une borne supérieure $L^{\pm}$, assurant que l'insertion d'un nombre $L \leq L^{\pm}$ bits d'informations garantie à un stéganographe que le test le plus puissant de $\mathcal{K}_{\alpha_0^\star}$ (donc tout test) ne peut avoir une puissance supérieure à $\beta^\star$.

5 Résultats numériques

L'approche statistique proposée dans cet article a été appliquée dans [5,6,13] avec l'obtention de résultats satisfaisant qui se comparent très favorablement



(a) Courbes COR obtenues pour $R = 1$. (b) Courbes COR obtenues pour $R = 0.5$.

FIGURE 1: Courbes COR $\beta(\alpha_0)$ présentant la performance des détecteurs adaptés au LSB ± 1 sur la base d'image BOSS [2] avec R=1 (a) et R=0.5 (b).

avec l'état de l'art dans le domaine.
En revanche, la détection d'informations cachées par correspondance de LSB est restée beaucoup moins étudiée. Excepté les méthodes reposant sur l'apprentissage supervisé, méthodes qui ne sont pas applicable dans le contexte opérationnel envisagé, les seules méthodes de détection adaptée à la correspondance de LSB reposent sur l'analyse des modifications de l'histogramme engendrées par l'insertion d'informations. Aussi, il a été choisi d'illustrer la pertinence de la méthodologie en comparant les résultats obtenus avec deux des plus récents détecteurs proposés dans la littérature [9,12], dont les performances ne sont évaluées que empiriquement ; aucune étude statistique n'ayant été menée.
En pratique les paramètres θ_n et σ_n ne sont pas connus du stéganalyste ; pour les estimer le modèle d'image décrit dans [3–5] est utilisé. Ces estimations sont utilisés dans le rapport de vraisemblance $\Lambda_R^{\pm}(z_n)$. Les performances des détecteurs sont illustrées graphiquement, sous la forme de courbes COR, par la figure 1a pour le cas $R = 1$, et par la figure 1b pour le cas $R = 0.5$. Les figures 1 montrent que le test statistique proposé $\delta^{\pm}$ offre une puissance de détection supérieure aux deux détecteurs, pour le cas d'un taux d'insertion $R = 0.5$. La différence de puissance est légèrement moindre pour $R = 1$.

6 Conclusions et perspectives

Dans ce papier la méthodologie proposée dans [5–7,13] est appliquée dans le cas de détection d'informations cachées par correspondance de LSB (LSB ± 1). Le test le plus puissant est dans un premier temps conçu dans un cadre "idéalisé" où les paramètres décrivant un média numérique sont connus. Les performances

du test sont analytiquement explicitées ce qui permet, d'une part, de proposer une borne sur la puissance que l'on est en droit d'attendre et, d'autre part, de mesurer la capacité d'insertion d'un média garantissant un niveau de sécurité prédéfini. Un modèle précis des images naturelles est utilisé pour mettre en pratique le test dont les performances se comparent très favorablement avec l'état de l'art dans le domaine.

Références

1. Böhme, R. : Advanced Statistical Steganalysis. Springer Publishing Company, Incorporated, 1st edn. (2010)
2. BOSS contest : Break Our Steganographic System (2010), http ://www.agents.cz/boss/
3. Cogranne, R., Retraint, F., Fillatre, L., Zitzmann, C. : Détection optimale d'information cachées indépendante du contenu des images (18-19 janvier 2011)
4. Cogranne, R., Zitzmann, C., Fillatre, L., Nikiforov, I., Retraint, F., Cornu, P. : A cover image model for reliable steganalysis. In : Information Hiding. pp. 178 – 192. LNCS vol.6958, Springer (18-20 May 2011)
5. Cogranne, R., Zitzmann, C., Fillatre, L., Nikiforov, I., Retraint, F., Cornu, P. : Reliable detection of hidden information based on a non-linear local model. In : Statistical Signal Processing, Proc. of IEEE Workshop on. pp. 493 – 496 (28-30 June 2011)
6. Cogranne, R., Zitzmann, C., Fillatre, L., Retraint, F., Nikiforov, I., Cornu, P. : Statistical decision by using quantized observations. In : IEEE International Symposium on Information Theory. pp. 1135 – 1139 (August 2011)
7. Dabeer, O., Sullivan, K., Madhow, U., Chandrasekaran, S., Manjunath, B. : Detection of hiding in the least significant bit. Signal Processing, IEEE Transactions on 52(10), 3046 – 3058 (oct 2004),
8. Fridrich, J. : Steganography in Digital Media : Principles, Algorithms, and Applications. Cambridge University Press, 1st edition edn. (2009)
9. Ker, A. : Steganalysis of lsb matching in grayscale images. Signal Processing Letters, IEEE 12(6), 441 – 444 (june 2005),
10. Ker, A.D. : A capacity result for batch steganography. Signal Processing Letters 14(8), 525–528 (2007)
11. Lehman, E., Romano, J. : Testing Statistical Hypotheses, Second Edition. Springer, 3rd edn. (2005)
12. Zhang, J., Cox, I., Doerr, G. : Steganalysis for lsb matching in images with high-frequency noise. In : Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on. pp. 385 –388 (oct 2007),
13. Zitzmann, C., Cogranne, R., Retraint, F., Nikiforov, I., Fillatre, L., Cornu, P. : Statistical decision methods in hidden information detection. In : Information Hiding. pp. 163 – 177. LNCS vol.6958, Springer (18-20 May 2011)