Histoire Récente de la Sécurité des Contenus Multimédia Un Focus sur la Dissimulation d'Information

Patrick BAS[1] Gwenaël DOËRR[2] Teddy FURON[3] William PUECH[4]

[1]Centre de Recherche en Informatique, Signal et Automatique de Lille
CNRS, Université Lille, Centrale Lille, Avenue Henri Poincaré, 59655 Villeneuve d'Ascq, France

[2]Synamedia Technologies France 12A rue du Pâtis Tatelin, 35700 Rennes, France

[3]Institut de Recherche en Informatique et Systèmes Aléatoires
Université de Rennes, INRIA, CNRS, 263 avenue du Général Leclerc, 35402 Rennes, France

[4]Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier
Université de Montpellier, CNRS, 860 rue de Saint Priest, 34095 Montpellier, France

Résumé – Le tatouage numérique et la stéganographie sont les deux faces de la dissimulation d'information dans les contenus multimédia. Dans cet article, nous passons en revue les avancées techniques de ces deux domaines et nous indiquons comment ces technologies se sont installées dans notre vie de tous les jours.

Abstract – Digital watermarking and steganography are the two sides of the same information hiding in multimedia content. In this article, we will review the major technical advances of both domains, and we will highlight how these technologies have now become part of our everyday life.

1 Introduction

Ross Anderson organise en Juin 1996 à Cambridge University le premier atelier consacré spécifiquement à la dissimulation d'information [1]. Cet événement marque la naissance d'une longue série connue sous le nom d'*Information Hiding Workshops*. Une terminologie y est établie [16]. La dissimulation d'information, qui est l'art de cacher un message dans un contenu hôte, se scinde en deux applications que sont le *tatouage numérique* et la *stéganographie*. Pour la première, cacher veut dire enfouir au plus profond, lier définitivement message et contenu ; pour la deuxième, cacher veut dire sans laisser aucune trace statistiquement détectable. Cet événement est avant tout la rencontre de deux communautés qui s'ignorent.

Les français.e.s se distinguent très tôt : on parle de *tatouage numérique*, terme insistant sur le côté robuste, non effaçable de la technique alors que les anglophones adoptent le terme *digital watermarking*, i.e. filigrane numérique, terme qui insiste sur l'imperceptibilité. Le terme tatouage est trouvé à Thomson CSF (maintenant Thales), ce qui reflète que cette technique est née dans l'industrie. La grande époque du tatouage suit la transition numérique de la fin des années 90 et conséquemment une démocratisation notable du piratage de contenus multimédia. Hollywood cherche à imposer le tatouage pour protéger ses nouveaux supports, les DVD. Les meilleures équipes sont alors industrielles : Philips, NEC, Sony, etc.

La stéganographie moderne a un début beaucoup plus discret bien qu'elle soit issue d'une longue tradition. L'ouvrage d'Énée le Tacticien au IV^{e} siècle avant J.C. présente plus d'une vingtaine de moyens relevant de la stéganographie [18]. Cette discrétion est due à ses utilisations principales : l'espionnage et le contournement de la censure. Elle ne fascine que quelques universitaires avant de se faire tristement connaître du grand public suite aux attaques du 11 septembre 2001. Les journaux américains prétendent que le réseau Al-Qaïda utiliserait la stéganographie pour communiquer secrètement. Suite à cet événement de nombreuses agences de sécurité nationales, telles l'armée américaine, soutiendront financièrement les recherches en stéganalyse et par extension en stéganographie.

Cet article retrace l'historique des 25 dernières années de ces deux domaines techniques voisins en indiquant quelques contributions majeures, mais aussi comment ces technologies sont devenues une partie intégrante de notre quotidien. Même si cela semble contre-intuitif, le fait que tout le monde ignore que la dissimulation d'information est omniprésente dans nos vies est une preuve indéniable du succès de ces technologies dont la gloire ne souffre pas la lumière. Les principaux faits d'arme ce de domaine ont été capturés dans la Figure 1.

2 Tatouage Numérique

L'objectif du tatouage est de modifier de façon *imperceptible* la représentation d'un contenu (image, vidéo, son, objet 3D) pour transmettre de l'*information* de façon *robuste*. La sémantique de cette information varie en fonction du cas d'usage. Par exemple, dans la lutte contre le piratage audiovisuel, il est courant que le tatouage encode une identité (appareil, abonnement, session) afin de d'offrir le moyen de retrouver la source du piratage. Les algorithmes de tatouage sont évalués par rapport à trois grandes métriques de performance :

1. l'imperceptibilité : un être humain ne doit pas différencier un contenu tatoué de sa version originale ;
2. la robustesse : une machine doit extraire l'information de tatouage même si le contenu tatoué a été modifié (filtrage, recompression, ré-échantilonnage...) ;

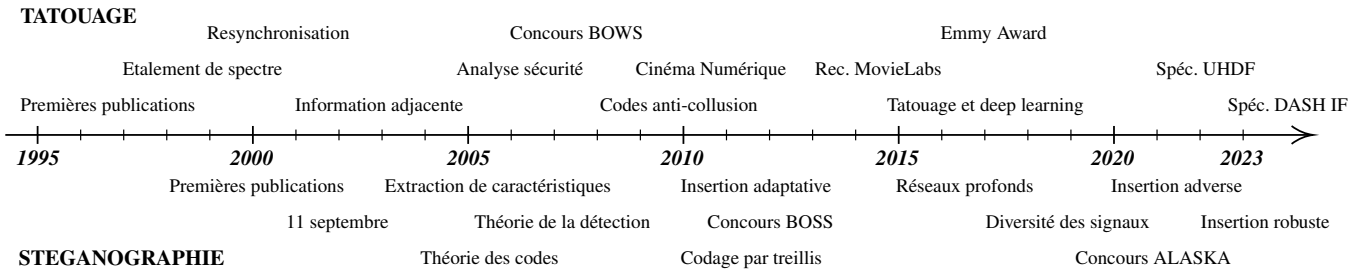

FIGURE 1 : Chronologie des avancées majeures dans le domaine du tatouage et de la stéganographie.

3. le taux d'insertion : la quantité d'information insérée par unité de contenu hôte, e.g. par seconde de musique.

Ces trois métriques s'opposent et l'un des défis est d'offrir le meilleur compromis en fonction de l'application visée. Cela nécessite des connaissances en télécommunication numérique, en sécurité informatique, en traitement du signal et des images, en statistiques ainsi qu'en perception humaine. Bref, le tatouage numérique est une discipline complète.

2.1 L'approche TEMIT

La grande mode du tatouage à la fin des années 90 déclenche un foisonnement mondial de contributions qui, au fil du temps, se consolident autour d'un schéma directeur : TEMIT, acronyme introduit par Kalker pour *Transform, EMbed, Inverse Transform*. Le tatouage se réduit à trois étapes : *i*) appliquer une transformation au signal, *ii*) introduire un biais statistique, et *iii*) appliquer la transformation inverse. L'extraction du tatouage applique la même transformation et vérifie l'absence ou la présence du biais statistique introduit.

C'est au niveau de la transformation du signal que vont se trouver les spécificités liées au type de contenu hôte et à l'application visée. Cette transformation apporte un premier niveau de robustesse. Par exemple, la transformée de Fourier-Mellin offre un espace de tatouage invariant à la rotation, la découpe, et la mise à l'échelle [15]. Cette transformation donne une modélisation fine de la perception humaine, par exemple la DCT ou la transformée en ondelette équipée d'un modèle de Watson. Cette transformation réduit le temps d'insertion, par exemple c'est celle utilisée pour coder le contenu hôte ce qui permet d'ajouter le tatouage sans le décompresser complètement.

2.2 Etalement de spectre

A la suite des travaux de Cox [7], presque tous les articles utilisent la même technique d'insertion : l'*étalement de spectre*. Cette technique étale le signal de tatouage sur l'ensemble du contenu spatialement et fréquentiellement. L'étalement opère à une puissance faible au sens où chaque coefficient de la représentation du contenu n'est que très peu modifié, ce qui garantit une bonne imperceptibilité. Le signal de tatouage est une modulation de porteuses pseudo-aléatoires impossible à décoder sans connaissance de la clé. Si une attaque tronque le signal (filtrage fréquentiel ou découpe en spatial), il reste suffisamment de signal ailleurs pour décoder le message d'où une grande robustesse. Décoder le tatouage est simple car il suffit de calculer les corrélations avec les porteuses à l'aide de la clé. A cette époque, le tatouage n'est qu'une application de cette technique de communication militaire éprouvée contre le brouillage lors de la 2^{nde} guerre mondiale.

2.3 Information adjacente

C'est au début des années 2000 que le tatouage reçoit ses lettres de noblesse. Jusque là, le taux d'insertion du tatouage reste faible car la source principale de bruit n'est autre que le contenu multimedia hôte. Néanmoins, ce contenu n'est pas vraiment du bruit puisqu'il est connu à l'insertion du tatouage. C'est donc une information adjacente pour l'encodeur, et du bruit inconnu pour le décodeur. Chen & Wornell [4] redécouvrent que Gel'fand & Pinsker [12] puis Costa [6] avaient montré dans les années 80 que la capacité théorique d'un tel schéma est indépendante de la puissance de l'information adjacente et ainsi donné les fondements théoriques du tatouage sans le savoir. Costa fournit une métaphore très parlante : on écrit à quantité d'encre constante sur une page de papier blanc ; en revanche, sur du papier sale, il vaut mieux moduler la pression de la plume en fonction des taches. Ainsi, le signal de tatouage doit aussi s'adapter au contenu hôte pour mieux communiquer.

La déception suit la phase d'euphorie. Ce nouveau concept est très difficile à mettre en oeuvre (lattices imbriquées, modulation en treillis). Si l'étalement de spectre est un 4×4 lent mais passe partout, le schéma de Costa s'avère être une Formule 1 incapable de rouler sur la vérité terrain. Un véhicule juste milieu reste encore à trouver de nos jours.

2.4 La sécurité

L'extrême robustesse du tatouage induit un faux sentiment de sécurité. Ceci a été mis à jour au milieu des années 2000. La robustesse se quantifie par l'augmentation du taux d'erreur à la détection lorsque le contenu tatoué subit des modifications. Mais ces modifications sont des traitements classiques d'édition de contenu, qui n'ont pas été conçus avec l'intention de nuire au tatouage.

La sécurité diffère par l'intention de l'attaquant, mais aussi par sa recherche de l'attaque qui endommage le plus le tatouage. Cela passe par l'estimation de la clé secrète de la technique de tatouage qui donne accès (lire, modifier, effacer) au message caché. Combien de contenus tatoués avec la même technique et la même clé secrète faut-il observer pour estimer finement cette clé ? Cette nouvelle perspective déclenche une vague de contributions scientifiques, lancées par une revisite du travail de Shannon sur la cryptanalyse [3].

Par ailleurs, plutôt que de rester isolés, les pirates peuvent collaborer, pour combiner leurs différentes versions tatouées

d'un même contenu par exemple. Ce type d'attaques nécessite alors d'employer des codes anti-collusion [11] pour éviter d'accuser un innocent. Tardos relance ce sujet en 2008 en proposant un code assez court pour être utilisé en pratique.

2.5 Adoption commerciale

Même si elle subsiste, la production scientifique dans le domaine du tatouage ralentit de façon significative après 2010. Et en même temps, l'adoption de cette technologie sur le marché ne va cesser de s'accélérer au cours de la décennie suivante.

La lutte contre le piratage des films et séries a été historiquement portée par les Studios Holywoodiens, ce qui a favorisé l'introduction du tatouage d'abord en post-production, puis dans le cinéma numérique, et enfin sur le marché de l'hospitalité (hôtels, avions, ...). Cette collaboration technique a d'ailleurs été célébrée par un Emmy Award technique en 2016. Le monde du cinéma a récemment été rejoint par les diffuseurs (broadcast, IPTV, OTT) pour lutter contre la retransmission pirate d'évènements sportifs.

Le tatouage numérique est aussi largement utilisé en audio, pour faciliter la mesure d'audience des radios et des chaînes TV aux quatre coins du globe par exemple. De même, dans le domaine de la presse, les photos sont tatouées pour rétribuer automatiquement les agences. C'est avec ce type de contenu que des expérimentations ont été tentées pour des scénarios d'usage éloignés de la protection de contenus. Par exemple, aux Etats-Unis, la société Digimarc exploite des technologies de tatouage pour remplacer le code barre des emballages. L'objectif annoncé est alors d'accélérer le passage en caisse dans les supermarchés et de faciliter le recyclage des déchets.

3 Stéganographie et Stéganalyse

Contrairement au tatouage, la stéganographie cherche à dissimuler des données dans un contenu sans qu'un adversaire puisse détecter la présence de cette information. La contrainte de robustesse est ainsi remplacée par une contrainte d'indétectabilité. Contrairement à la cryptographie, la transmission de l'information sensible s'effectue ici par dissimulation et non par chiffrement. Enfin, telle la cryptanalyse vis-à-vis de la cryptographie, la stéganalyse cherche à attaquer les méthodes de stéganographie en détectant la présence d'informations cachées. Les objectifs sont doubles : évaluer la sécurité des méthodes de stéganographie d'une part, détecter de manière opérationnelle l'utilisation de la stéganographie d'autre part.

3.1 Stéganographie et contenus numériques

Avec l'arrivée des contenus numériques au XX$^{\text{e}}$ siècle, la stéganographie moderne est naturellement apparue via la modification des représentations des contenus de façon imperceptible, par exemple en substituant les bits de poids faibles des pixels ou des coefficients DCT d'une image. Ces méthodes appelées méthodes par substitution LSB rajoutent un signal stéganographique $s_i \in \{-1, 0, 1\}$ au contenu. D'autres méthodes plus naïves insèrent des informations dans les méta-données de l'image, ou bien concatènent le fichier à cacher à la suite du contenu. Les premières publications citant le terme "stéganalyse" datent de 1998 [13] ; les méthodes présentées sont rudimentaires, elles se cantonnent à la détection de signatures dans les méta-données ou à l'inspection visuelles d'histogrammes.

3.2 Différentes stratégies en stéganalyse

Il est notable de constater qu'avant le tournant historique du 11 septembre 2001, les quatre principales familles en stéganalyse existaient déjà, à savoir :

1. Les tests statistiques pour la détection de stéganographie par substitution LSB, comme le test du Chi-2 [19] ;
2. les attaques par compatibilité qui trouvent par exemple des modifications de pixels non-compatibles avec une compression JPEG [10] ;
3. l'apprentissage automatique reposant d'une part sur des caractéristiques sensibles à l'insertion, et d'autre part sur l'utilisation d'un classifieur entraîné sur des bases d'images normales ou contenant de l'information cachée [8] ;
4. l'estimation de la taille du message inséré, appelée stéganalyse quantitative [19].

A partir de 2002, les publications en stéganographie et stéganalyse se multiplient considérablement. La stéganographie voit sa sécurité augmenter grâce à la théorie des communications numériques pour, d'un coté réduire le nombre de modifications liées à l'insertion du message (codage par syndromes), de l'autre modifier uniquement les zones les moins détectables de l'image telles les textures ou les contours (codes sur papier mouillé). D'autres résultats liés à la théorie de la détection établissent la loi d'airain de la stéganographie, connue sous le nom de *square root law* : à détectabilité donnée, la taille du message à cacher est proportionnelle à la racine carrée de la taille du contenu hôte [14].

Coté stéganalyse, les méthodes par apprentissage proposent des jeux de caractéristiques dont la dimension ne cesse de croître afin d'augmenter la puissance de détection, mais aussi de faire face aux méthodes d'insertion adaptatives.

3.3 Stimulation des challenges académiques

En 2010, le concours *Break Our Steganographic System* (BOSS) propose aux participants de détecter la présence d'informations cachées dans 500 images parmi 1000 [2]. Cet événement voit la réalisation d'avancées majeures :

- En stéganographie, avec le développement du codage par treillis et syndromes qui donne une insertion adaptative quasi optimale où chaque échantillon est associé à un coût de modification [9] ;
- En stéganalyse, avec l'utilisation conjointe de descripteurs de très grande dimension ($d = 3.10^4$) et de classifieurs par ensemble afin de passer à l'échelle.

Deux autres concepts viennent améliorer la sécurité des méthodes de stéganographie : (i) la prise en compte de l'erreur de quantification obtenue lors de la compression de l'image et (ii) la subdivision de l'image en treillis afin d'effectuer des modifications corrélées entre elles.

Comme pour d'autres domaines en analyse d'images, l'année 2015 voit l'arrivée des réseaux de neurones profonds pour

la stéganalyse [17]. De part leurs performances bien supérieures aux méthodes par extraction de descripteurs construits par l'humain, ils deviennent la référence en 2020 lors du concours Kaggle Alaska-2 [5]. Les réseaux sont tout d'abord spécifiquement construits pour analyser les signaux faibles d'une image, ils sont aussi entraînés sur des très grandes bases comme ImageNet pour encore améliorer leurs performances.

3.4 Défis techniques actuels

Les travaux actuels en stéganalyse se focalisent essentiellement sur les problèmes de robustesse face à des domaines d'apprentissage différents. En effet un détecteur entraîné sur une base d'images provenant d'un appareil A aura généralement des performances médiocres si les images tests sont issues d'un appareil $B \neq A$. Il est ainsi nécessaire de modéliser précisément la distribution des signaux faibles de l'image, généralement issus du bruit du capteur.

Les dernières avancées en stéganographie rebondissent majoritairement sur les progrès en intelligence artificielle. En s'inspirant des exemples adverses, la stéganographie adverse pilote l'insertion du message de manière à contourner la stéganalyse. Les modèles génératifs de textes ou d'images sont aussi maintenant utilisés afin d'insérer des données cachées, leur sécurité et leur capacité sont théoriquement très importantes.

Enfin, sur un plan plus pratique, l'utilisation massive des réseaux sociaux motive actuellement des recherches autour de la stéganographie robuste, l'objectif est d'insérer un message indétectable mais robuste aux traitements effectués lors du dépôt du contenu sur une plateforme.

4 Conclusion et Perspectives

La France est un vivier exceptionnel dans ce domaine, que ce soit au niveau académique ou industriel. Pas moins de huit laboratoires de recherche ont acquis une réputation internationale sur le sujet et contribuent à l'animation du GdR ISIS en France. Avec cinq fournisseurs de tatouage sur son territoire, Rennes fait figure de capitale internationale de cette technologie. Pour ce qui est de la stéganographie, même si le domaine n'a pas atteint la maturité du tatouage, un projet Européen H2020 incluant 11 agences de sécurité a débuté en 2021 et la stéganographie est souvent citée dans la presse spécialisée comme un moyen de communication utilisé par les *botnets*, ce qui ne va pas sans poser des questions éthiques sur la potentielle mauvaise utilisation de ces technologies *discrètes*.

La dissimulation de données reste un domaine de recherche d'actualité. Le deep learning a réveillé une communauté du tatouage endormie depuis 2015 en apprenant des transformées robustes idéales pour un schéma TEMIT, voire même, comme en stéganographie, en apprenant conjointement l'insertion et le décodage. Du côté industriel, plusieurs initiatives visent à standardiser certaines interfaces et ainsi faciliter l'interopérabilité des systèmes de tatouage, en particulier pour les systèmes de tatouage A/B en OTT (UHD Forum, DASH Industry Forum, CTA WAVE Common Access Token).

Par ailleurs, le tatouage et la stéganographie robuste ont été récemment proposés comme solutions aux problèmes de confiance numérique introduit par l'intelligence artificielle. En incluant la dissimulation d'un message dans leurs modèles génératifs, certains propriétaires d'IA expriment le voeu pieux qu'elle ne se fasse pas passer pour un humain ni ne produit des deepfakes... jusqu'à ce que les adversaires soient à nouveaux capables de contourner ces protections. Les résultats en dissimulation de données offrent aussi un terreau fertile pour certaines thématiques de recherche émergentes, telles que l'imagerie à destination de la médecine légale (les sciences forensiques). Retrouver la signature numérique d'un appareil électronique peut en effet s'apparenter à la détection d'un tatouage naturellement introduit par cet appareil. Les techniques de détection de signaux faibles peuvent alors s'avérer très utiles.

Références

[1] R. J. ANDERSON, éditeur. *Proc. 1^{st} Intl. Workshop on Inf. Hiding*, volume 1174 de *LNCS*, 1996.

[2] P. BAS, T. FILLER et T. PEVNY : "Break our steganographic system" : The ins and outs of organizing BOSS. *In Proc. Inf. Hiding*, volume 6958 de *LNCS*, pages 59–70, 2011.

[3] F. CAYRE, C. FONTAINE et T. FURON : Watermarking security : Theory and practice. *IEEE Trans. Sig. Process.*, 53(10):3976–3987, 2005.

[4] B. CHEN et G. W. WORNELL : Quantization index modulation : A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Inf. Theory*, 47(4):1423–1443, 2001.

[5] R. COGRANNE, Q. GIBOULOT et P. BAS : Alaska-2 : Challenging academic research on steganalysis with realistic images. *In Proc. IEEE Int. Workshop on Inf. Forensics and Security*, 2020.

[6] M. COSTA : Writing on dirty paper. *IEEE Trans. Inf. Theory*, 29(3):439–441, 1983.

[7] I. J. COX, J. KILIAN, F. T. LEIGHTON et T. SHAMOON : Secure spread spectrum watermarking for multimedia content. *IEEE Trans. Image Process.*, 6(12):1673–1687, 1997.

[8] H. FARID : Detecting steganographic messages in digital images. Rapport technique TR2001-412, Dartmouth College, Hanover, 2001.

[9] T. FILLER, J. JUDAS et J. FRIDRICH : Minimizing additive distortion in steganography using syndrome-trellis codes. *IEEE Trans. Inf. Forensics and Security*, 6(3):920–935, 2011.

[10] J. FRIDRICH, M. GOLJAN et R. DU : Steganalysis based on JPEG compatibility. *In Multimedia Syst. and Apps. IV*, volume 4518 de *Proc. of SPIE*, pages 275–280, 2001.

[11] T. FURON et G. DOËRR : Tracing pirated content on the Internet : Unwinding Ariadne's thread. *IEEE Security & Privacy*, 8(5):69–71, 2010.

[12] S. GEL'FAND et M. S. PINSKER : Coding for channels with random parameters. *Probl. Control and Inf. Theory*, 9(1):19–31, 1980.

[13] N. F. JOHNSON et S. JAJODIA : Steganalysis : The investigation of hidden information. *In Proc. IEEE Inf. Technology Conf., Inf. Env. for the Future*, pages 113–116, 1998.

[14] A. D. KER, T. PEVNÝ, J. KODOVSKÝ et J. FRIDRICH : The square root law of steganographic capacity. *In Proc. ACM Workshop on Multimedia and Security*, pages 107–116, 2008.

[15] C.-Y. LIN, M. WU, J. A. BLOOM, I. J. COX, M. L. MILLER et Y. M. LUI : Rotation, scale, and translation resilient watermarking for images. *IEEE Trans. on Image Process.*, 10(5):767–782, 2001.

[16] B. PFITZMANN : Information hiding terminology - Results of an informal plenary meeting and additional proposals. *In* ANDERSON [1], pages 347–350.

[17] L. PIBRE, J. PASQUET, D. IENCO et M. CHAUMONT : Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover source-mismatch. Rapport technique 1511.04855, arXiv, 2015.

[18] Enée Le TACTICIEN : *Extraits du Traité sur La Défence Des Places*. Mémoires de la Société d'émulation du Doubs, 1870.

[19] A. WESTFELD et A. PFITZMANN : Attacks on steganographic systems : Breaking the steganographic utilities EzStego, Jsteg, Steganos, and S-Tools and some lessons learned. *In Proc. Inf. Hiding*, volume 1768 de *LNCS*, pages 61–76, 2000.